

**UNIVERSIDADE FEDERAL DE OURO PRETO  
INSTITUTO DE CIÊNCIAS EXATAS E APLICADAS  
DEPARTAMENTO DE ENGENHARIA DE PRODUÇÃO  
CURSO DE GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO**

**SETOR SUCROENERGÉTICO BRASILEIRO: UM ESTUDO ATRAVÉS DE  
TÉCNICAS ESTATÍSTICAS MULTIVARIADAS**

**ANNA CAROLINA SILVA SANTOS**

**TRABALHO DE CONCLUSÃO DE CURSO**

João Monlevade / Minas Gerais

Fevereiro 2018

**ANNA CAROLINA SILVA SANTOS**

**SETOR SUCROENERGÉTICO BRASILEIRO: UM ESTUDO ATRAVÉS DE  
TÉCNICAS ESTATÍSTICAS MULTIVARIADAS**

Trabalho de Conclusão de Curso apresentado ao curso de Engenharia de Produção do Departamento de Engenharia de Produção da Universidade Federal de Ouro Preto, como parte das exigências para obtenção de grau de Bacharel em Engenharia de Produção.

*Orientação: Prof<sup>a</sup> Rafael Lucas Machado Pinto*

**TRABALHO DE CONCLUSÃO DE CURSO**

João Monlevade / Minas Gerais

Fevereiro 2018

S237s

Santos, Anna Carolina Silva.

Setor sucroenergético brasileiro [manuscrito]: um estudo através de técnicas estatísticas multivariadas / Anna Carolina Silva Santos. - 2018.

50f.: il.: color; tabs; Figuras.

Orientador: Prof. MSc. Rafael Lucas Machado Pinto.

Monografia (Graduação). Universidade Federal de Ouro Preto. Instituto de Ciências Exatas e Aplicadas. Departamento de Engenharia de Produção.

1. Bicombustíveis . 2. Álcool. 3. Cana-de-açúcar. 4. Análise multivariada. I. Pinto, Rafael Lucas Machado. II. Universidade Federal de Ouro Preto. III. Título.

CDU: 620.92

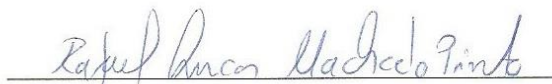
## ATA DE DEFESA

Aos dias 05 do mês de Fevereiro de 2018, às 18:30 horas, na sala B102 deste instituto, foi realizada a defesa do Trabalho de Conclusão de Curso pela aluna Anna Carolina Silva Santos, sendo a comissão examinadora constituída pelos professores: Prof<sup>o</sup> Rafael Lucas Machado Pinto e pelas convidadas Prof<sup>a</sup>. Luciana Paula Reis e Prof<sup>a</sup>. Carla Danielle Araújo Costa. A aluna apresentou o trabalho intitulado: “Setor Sucoenergético Brasileiro: um estudo através de técnicas estatísticas multivariadas”. A comissão examinadora deliberou, pela:

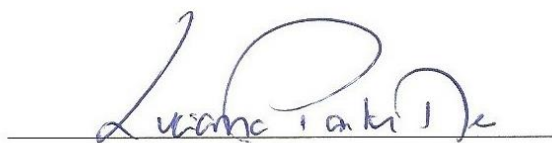
- Aprovação  
 Aprovação com Ressalva - Prazo concedido para as correções: \_\_\_\_\_  
 Reprovação com Ressalva - Prazo para marcação da nova banca: \_\_\_\_\_  
 Reprovação

da aluna, com a nota 9,8. Na forma regulamentar e seguindo as determinações da resolução COEP12/2015 foi lavrada a presente ata que é assinada pelos membros da comissão examinadora e pela aluna.

João Monlevade, 05 de Fevereiro de 2018.




Rafael Lucas Machado Pinto



Luciana Paula Reis



Carla Danielle Araújo Costa

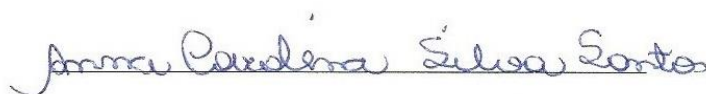


Anna Carolina Silva Santos

## TERMO DE RESPONSABILIDADE

O texto do trabalho de conclusão de curso intitulado “**Setor Sucroenergético Brasileiro: um estudo através de técnicas estatísticas multivariadas**” é de minha inteira responsabilidade. Declaro que não há utilização indevida de texto, material fotográfico ou qualquer outro material pertencente a terceiros sem o devido referenciamento ou consentimento dos referidos autores.

João Monlevade, 05 de Fevereiro de 2018



Anna Carolina Silva Santos

## **AGRADECIMENTO**

Agradeço a Deus por ter me dado a oportunidade de crescimento, ter me abençoado em todos os momentos, pela força e coragem para continuar buscando meus objetivos e enfrentando todos os obstáculos.

Agradeço imensamente aos meus pais, por não medirem esforços para me ajudar a construir um futuro melhor. Por serem o meu porto seguro, pela compreensão e por compartilharem todos os momentos bons e difíceis.

Aos meus irmãos que sempre acreditaram em meu potencial e sempre estiveram ao meu lado.

A minha família que sempre torceu pela minha vitória e por colaborarem, cada um com sua maneira.

Ao meu companheiro, que sempre esteve disposto a escutar, apoiar e por sempre acreditar em mim.

Ao meu orientador, Prof. Rafael Lucas Machado Pinto, pela sua dedicação, profissionalismo e comprometimento, o qual também foi fonte de motivação.

A todos os funcionários e principalmente aos professores por compartilharem seus conhecimentos, experiências e por serem parte desse crescimento.

Aos amigos agradeço pelos conselhos, amizade e por acreditarem em minha capacidade.

## RESUMO

O setor sucroenergético é um setor agroindustrial que o Brasil possui pleno domínio tecnológico, tanto no campo industrial como agrícola, apresentando, portanto, grande importância econômica. O Programa Nacional do Alcool e o surgimento dos veículos flex contribuíram para a ascensão do consumo de etanol, impulsionando diversas melhorias para o setor e progressos em questões socioambientais. O presente trabalho buscou identificar o tipo de relação existente entre oferta de etanol e alguns fatores socioeconômicos do setor, bem como analisar a relação entre as variáveis selecionadas. Para isso, baseou-se na análise de correlação e de regressão, onde foram utilizados dados estatísticos anuais do setor durante o período de 2006 a 2015. Como resultado, correlações significativas foram identificadas, como exemplo, uma forte correlação positiva entre produções de cana e etanol, Frota total de veículos e Frota de veículos bicomcombustíveis. Utilizando a análise *cluster* buscou-se obter grupos formados por estados, que ao final resultaram em 6, os quais apresentaram maiores níveis de similaridade entre si. Tais agrupamentos foram feitos com base em variáveis socioeconômicas relacionadas ao setor e aos estados escolhidos. Por fim, este estudo se justifica por preencher uma lacuna em relação a poucos estudos realizados nesta área e por este setor ser representativo para o desenvolvimento do país.

**Palavras chave:** Bicomcombustíveis, etanol, correlação, regressão, *cluster*.

## ABSTRACT

The sugar-energy sector is an agroindustrial sector that Brazil has full technological dominance, in the industrial as well as the agricultural field, presenting, great economic importance. The National Alcohol Program and the emergence of flex-fuel vehicles have contributed to the rise in ethanol consumption, boosting several improvements to the sector and progress in socio-environmental issues. The present work sought to identify the type of relation between ethanol supply and some socioeconomic factors of the sector, as well as to analyze the relationship between the selected variables. It was based on the correlation and regression analysis, where annual statistical data were used of the sector during the period from 2006 to 2015. As a result, significant correlations were identified, for example, a strong positive correlation between sugarcane and ethanol productions, Total fleet of vehicles and Fleet of bi-fuel vehicles. Using the cluster analysis, it was sought to obtain groups formed by states, which in the end resulted in 6, which presented higher levels of similarity between them. These groups were made based on socioeconomic variables related to the sector and the chosen states. Finally, this study is justified by filling a gap in relation to few studies carried out in this area and because this sector is representative for the development of the country.

**Keywords:** Biofuels, ethanol, correlation, regression, cluster.



## LISTA DE FIGURAS

Figura 1- Processo de produção de etanol .....	16
Figura 2- Produção de Etanol Total.....	18
Figura 3- Técnica Hierárquica Aglomerativa .....	23
Figura 4- Classificação da pesquisa científica em Engenharia de Produção.....	25
Figura 5 - Produção de etanol X Produção de cana.....	30
Figura 6 - Produção de etanol X Frota de automóveis à gasolina .....	32
Figura 7 - Frota de Veículos no Brasil .....	33
Figura 8 - Dendrograma dos Clusters formados.....	39

## LISTA DE TABELAS

Tabela 1- Interpretação dos valores de coeficiente de correlação .....	19
Tabela 2- Correlação entre as variáveis.....	28
Tabela 3- ANOVA para a regressão entre produção total de etanol e produção de cana-de-açúcar.....	29
Tabela 4- Estatísticas de Regressão para o modelo formado entre produção total de etanol e produção de cana-de-açúcar .....	30
Tabela 5- Testes t para os coeficientes da regressão entre produção total de etanol e produção de cana-de-açúcar .....	30
Tabela 6- ANOVA para a regressão entre produção total de etanol e frota de automóveis à gasolina.....	31
Tabela 7- Estatísticas de Regressão para o modelo formado entre produção total de etanol e frota de veículos à gasolina .....	31
Tabela 8- Testes t para os coeficientes da regressão entre produção total de etanol e frota de veículos à gasolina.....	32
Tabela 9- Análise Cluster no setor Sucroenergético brasileiro .....	36
Tabela 10- Explicação dos Clusters .....	37
Tabela 11- Distâncias entre os centros dos centroides .....	38
Tabela 12- Estados e variáveis da análise Cluster.....	40
Tabela 13 - Ordem de Classificação entre Estados .....	41

## SUMÁRIO

<b>1. INTRODUÇÃO</b>	11
<b>2. PROBLEMA DE PESQUISA</b>	14
2.1 Justificativa	14
2.2 Objetivos	15
2.2.1 Objetivo Geral	15
2.2.2 Objetivos Específicos	15
2.3 Estrutura do Trabalho	15
<b>3. REFERENCIAL TEÓRICO</b>	16
3.1 Etanol	16
3.1.1 Processo de produção de etanol	16
3.1.2 Etanol Anidro X Etanol Hidratado	17
3.1.3 Sustentabilidade	17
3.1.4 Oferta e Demanda do Etanol	18
3.2 Análise de Correlação	19
3.3 Regressão Linear Simples	21
3.4 Análise de Agrupamentos ( <i>Cluster</i> )	22
<b>4 METODOLOGIA DE PESQUISA</b>	25
4.1 Classificação Metodológica	25
4.2 Fonte de Dados	26
<b>5. ANÁLISES</b>	28
5.1 Análise de Correlação e Análise de Regressão Linear Simples	28
5.1 Análise de Agrupamentos ( <i>Cluster</i> )	35
<b>6. CONSIDERAÇÕES FINAIS</b>	45
<b>REFERÊNCIAS</b>	47

## 1. INTRODUÇÃO

De acordo com Nastari (2012), o setor sucroenergético, que está ligado a atividades como a produção de açúcar, bioeletricidade e bioetanol. Este é um setor agroindustrial que o Brasil possui pleno domínio tecnológico, tanto do campo industrial como agrícola, e o qual tem grande importância na economia nacional, uma vez que as demandas pelos produtos por ele gerados têm apresentado crescimento nos mercados interno e externo.

Inicialmente este setor era conhecido por sucroalcooleiro, porém, o fato de se estar produzindo outros produtos, além de etanol e açúcar, a partir da cana-de-açúcar o mesmo passou a ser chamado de sucroenergético. O desenvolvimento de novas tecnologias possibilitou a geração de energia elétrica (bioenergia) através do bagaço da cana, o qual era apenas queimado, também levou à produção de plástico e de bio-hidrocarboneto, utilizado como combustível de aviões (CAMPOS, 2014).

Tal setor tem garantido bons resultados no aspecto sustentabilidade, sejam eles relacionados à redução na emissão de gases do efeito estufa (GEE), utilização do sistema de rotação de culturas, a não queima da palha e a diminuição na utilização de fertilizantes, além de ser considerada uma importante fonte de biomassa, ou seja, energia renovável (UNICA). No aspecto social, ele tem apresentado melhorias nas condições de trabalho, assim como tem impulsionado o desenvolvimento das regiões em que se encontram suas atividades. Mesmo passando por más fases nos últimos tempos, algumas relacionadas a regulamentações governamentais para o setor, o mesmo vê grandes chances de crescimento para os próximos anos (CAETANO, PALHARES, 2016).

“A introdução da produção de cana de açúcar no Brasil data do século XIV, início do período colonial. Em meados do século XVII, o Brasil tornou-se o maior produtor de cana-de-cana do mundo, na época destinado ao abastecimento da Europa” (RODRIGUES; ORTIZ, 2006, pp. 5-6). Por volta de 1970, surgiu o Programa Nacional do Álcool (Proálcool), resultante de uma crise do petróleo. O programa permitiu o surgimento de várias melhorias, como a adaptação e inovação de tecnologias nas usinas e destilarias, a criação de sementes adaptadas e também um desenvolvimento da área automobilística (RODRIGUES; ORTIZ, 2006).

Ao final da década de 80, o Proálcool teve fim. Enquanto o financiamento público ao programa diminuía a partir de 1987, os preços do açúcar obtiveram aumentos no mercado

internacional, e como consequência houve uma desmotivação à expansão e renovação dos canaviais para produzir etanol, isto levou os produtores a desviar a matéria-prima para a produção de açúcar, os quais visavam a exportação. Tais fatos culminaram, ao final de 1989, em um desabastecimento do biocombustível, gerando assim grandes filas nos postos por todas as cidades brasileiras (RODRIGUES; ORTIZ, 2006).

Segundo Konishi *et al.* (2014) “o início do séc. XXI é marcado pelo retorno do etanol à posição de destaque como combustível, não somente pela nova tecnologia, mas também como fonte de energia alternativa ao combustível fóssil”. Desta forma, um grande impulsionador da produção de etanol foi a introdução de veículos *flex* (MELO; SAMPAIO, 2014). Boff (2010, apud MELO, SAMPAIO; 2014) diz que “a entrada dessa nova tecnologia causou uma mudança na relação entre a demanda de combustíveis etanol e gasolina”, pois isso significava a substituição do combustível fóssil por um combustível renovável. Segundo Rodrigues (2011), tal veículo “lançado em 2003 e que hoje responde por cerca de 90% das vendas, é o ponto alto da história do sucesso do etanol brasileiro nesta década”.

Atualmente, o setor se encontra em dois cenários. No positivo, houve aumento dos índices de produtividade, novos projetos de implantação de usinas, bem como desenvolvimentos de novas tecnologias no ramo. Já o negativo mostra que crises e falta de incentivo estatal no setor vem causando insatisfações aos empresários do agronegócio, os quais culpam o governo pelo retrocesso no programa de etanol (CAMPOS, 2015).

Diante de tal cenário, o presente estudo tem como objetivo identificar qual o tipo de relação existente entre a oferta de etanol e alguns fatores socioeconômicos do setor, bem como a relação entre esses fatores.

Sabendo-se que há uma certa carência de estudos neste sentido, o trabalho, por meio da Análise de Correlação e Regressão Linear Simples, pode contribuir para melhor compreender este importante setor energético do país e direcionar ações em relação ao planejamento energético nacional relativo a suprimentos e fontes de energia. No mesmo sentido, a Análise *Cluster* contribuirá de forma a se pensar na geração de políticas e práticas direcionadas para grupos de estados formados de acordo com as semelhanças encontradas a partir das variáveis socioeconômicas utilizadas.

A utilização de técnicas de análise Multivariada neste estudo irá contribuir de forma a tornar o processo de análise menos complexo, e que, posteriormente, permitirá que os

resultados possam ser explicados de acordo com as informações coletadas sobre o setor, ou seja, contextualizá-las. Desta forma, este estudo não seria viável sem o auxílio de tais ferramentas, visto que o número de variáveis a serem estudadas é considerável.

## 2. PROBLEMA DE PESQUISA

O presente trabalho busca encontrar e avaliar quais fatores mais influenciam na oferta de etanol no Brasil por meio da análise de correlação e regressão linear simples. Além disso, busca identificar quais estados apresentam maiores similaridades a partir de variáveis socioeconômicas levantadas do setor sucroenergético. Para tanto, será aplicada a análise cluster (ou análise de agrupamentos) para formar grupos de estados semelhantes sob a ótica do setor, o que poderia facilitar e viabilizar o desenvolvimento de políticas e práticas semelhantes para estados pertencentes aos mesmos grupos.

### 2.1 Justificativa

Tal estudo tem como propósito aumentar o nível de conhecimento sobre o setor sucroenergético brasileiro, o qual é considerado muito importante para a economia nacional e tem sido fonte de maiores estudos nos últimos anos. Neste contexto, a utilização de ferramentas estatísticas é vista como uma grande possibilidade de encontrar resultados relevantes e promover a geração de análises que poderão auxiliar em melhores estudos, tomadas de decisões e no desenvolvimento do setor.

Optou-se por utilizar a análise de correlação, análise de regressão linear simples e análise cluster para uma melhor compreensão do setor, de forma a entender como as variáveis socioeconômicas do mesmo estão relacionadas, bem como a similaridade entre estados onde a atividade de produção de etanol ocorre. Em ambas situações, pretende-se realizar o levantamento de variáveis socioeconômicas referentes ao setor estudado, utilizando para a primeira análise dados como: frota de veículos *flex*, preço do etanol hidratado, preço da gasolina e outros. Para a segunda, dados relacionados à produção, como a quantidade de etanol produzida, área de cana-de-açúcar plantada, área colhida, bem como densidade demográfica, tamanho da população por estado, serão os itens a serem abordados.

A escolha de se realizar o estudo, com o auxílio da estatística, se deu pelo fato da mesma, segundo Correa (2003, p. 8), fornecer “métodos para coleta, organização, descrição, análise e interpretação de dados”, assim como apresentar resultados que facilitam a compreensão e auxiliam na tomada de decisões por parte do pesquisador/gestor (IGNÁCIO, 2010).

## 2.2 Objetivos

### 2.2.1 Objetivo Geral

O presente trabalho tem como objetivo geral aplicar Análises Estatísticas Multivariadas para melhor compreender o setor sucroenergético brasileiro, realizando para isso análises sobre a oferta de etanol e sobre a forma como os principais estados produtores de tal combustível se assemelham.

### 2.2.2 Objetivos Específicos

- Aplicar a Análise de Correlação para entender o comportamento entre as variáveis;
- Aplicar a Regressão Linear Simples para entender quais os fatores que mais influenciam sobre a oferta nacional de etanol;
- Aplicar a Análise de Agrupamentos (*cluster*) para obter grupos compostos pelos principais estados produtores de etanol, de acordo com os níveis de similaridade entre eles;
- Explicar os resultados obtidos através de informações coletadas sobre o setor, ou seja, contextualizar o comportamento a respeito das variáveis analisadas neste estudo.

## 2.3 Estrutura do Trabalho

O trabalho apresentado está dividido em 6 seções. A primeira é composta pela introdução. A segunda pela contextualização do problema, justificativa, objetivos e a forma como se dá a estruturação do trabalho. A terceira aborda o referencial teórico relacionado ao tema do trabalho, no qual os leitores poderão conhecer mais sobre o assunto. A quarta seção apresenta a metodologia utilizada, assim como as fontes dos dados utilizados. A quinta aborda as análises propostas. Por fim, a sexta seção aborda as considerações finais do estudo realizado.



### 3. REFERENCIAL TEÓRICO

Para a realização do estudo proposto é fundamental que se tenha o conhecimento sobre o principal produto abordado, o etanol, assim como entender as análises de correlação, regressão linear simples e de agrupamentos que serão utilizados ao decorrer do trabalho.

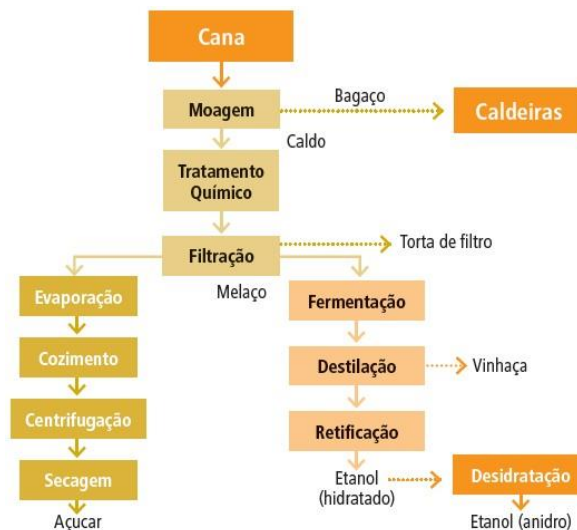
#### 3.1 Etanol

##### 3.1.1 Processo de produção de etanol

São conhecidas várias formas de se obter o etanol, sendo a hidratação do etileno e redução do acetaldeído duas delas, porém a forma mais utilizada é a fermentação. Através desse processo, é possível obter os dois tipos de álcoois, anidro e hidratado (NOVACANA, s.d.).

A Figura 1 descreve esse processo indicando todas as etapas necessárias

Figura 1- Processo de produção de etanol



Fonte: Adaptado de Machado, 2014

Primeiramente tem-se a cana-de-açúcar, a qual passa por três processos, moagem, tratamento químico e filtração, até se chegar ao melaço. Tal produto passa por mais três

processos, fermentação, destilação e retificação, obtendo-se o etanol hidratado (utilizado como combustível).

### **3.1.2 Etanol Anidro X Etanol Hidratado**

Como se sabe há diferenças entre os tipos de etanois produzidos, a mais importante diz respeito à proporção de água presente em suas composições. A NovaCana, empresa considerada o maior veículo de comunicação do setor sucroenergético do Brasil, informa que o etanol hidratado, mais conhecido como combustível vendido em postos, possui entre 95,1% a 96% de etanol e o restante de água. Por sua vez, o etanol anidro, que é misturado à gasolina, possui 99,6% de etanol, no mínimo (NOVACANA, s.d.).

Outra diferença encontrada está nas etapas de produção. Como se pode visualizar na Figura 1, após a produção do etanol hidratado, o mesmo passa pelo processo de desidratação chegando assim ao anidro (NOVACANA, s.d.).

### **3.1.3 Sustentabilidade**

Quando se trata de sustentabilidade e combustíveis, um aspecto muito abordado é a emissão de gases de efeito estufa (GEE). Segundo estudos, o etanol brasileiro é capaz de reduzir tal emissão em aproximadamente 90%, quando comparado à gasolina. Tal fato foi confirmado quando a Agência de Proteção Ambiental Americana (EPA), em 2010, colocou o etanol de cana-de-açúcar como capaz de reduzir emissões de GEE de 61% a 91% e assim o classificando como combustível avançado (UNICA, s.d.).

Outro fator a ser considerado é a forma de utilização do solo. Atualmente, a incorporação da cultura da cana-de-açúcar, matéria-prima do etanol, têm ocorrido em áreas com terrenos mais pobres (pastagens). A utilização do sistema de rotação de cultivo da cana com outras culturas em áreas de pastagem tem contribuído significativamente com processo de incorporação de Nitrogênio e matéria orgânica nos solos, o que contribui para uma menor utilização de fertilizantes e, conseqüentemente, em uma redução de emissão de gases de efeito estufa, visto que para a produção do mesmo são necessários combustíveis fósseis (NASTARI, 2012).

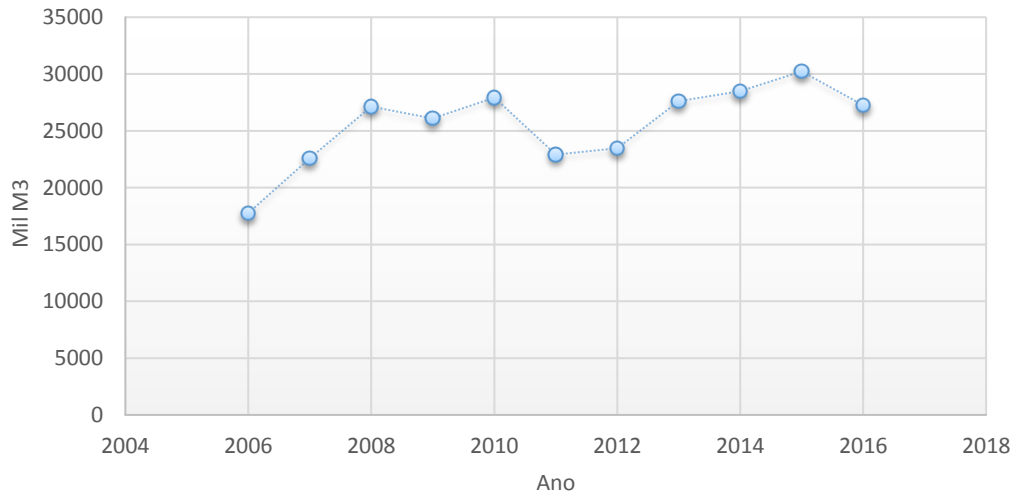
Outra prática relaciona-se à diminuição da queima da palha, que é fundamental para a conservação e proteção do solo e suas propriedades, tal fato está ocorrendo graças à mecanização e modernização das formas de colheita, visto que essas práticas contribuem muito para a prevenção da erosão, principal problema encontrado (NOVACANA, s.d.).

### 3.1.4 Oferta e Demanda do Etanol

Em 2015, a produção nacional de etanol obteve um acréscimo de 6,0% atingindo um montante de 30.249 mil m<sup>3</sup>, destes 61,8% refere-se ao etanol hidratado, correspondendo a 18.684,6 mil m<sup>3</sup> (acréscimo de 14,7% na produção deste em relação a 2014). Já a produção de etanol anidro, registrou uma queda de 5,4%, totalizando 11.564,6 mil m<sup>3</sup> (BALANÇO ENERGÉTICO NACIONAL, 2016).

A Figura 2 ilustra o cenário em que se deu a oferta de etanol total até o ano de 2016. A partir dela é possível identificar que, após uma considerável queda da oferta nos anos de 2011 e 2012, houveram aumentos significativos nos três anos seguintes.

Figura 2- Produção de Etanol Total



Fonte: Adaptado do Anuário Estatístico Brasileiro de Petróleo, Gás Natural e Biocombustível, 2016.

Em relação à demanda, pode-se constatar que o consumo de combustíveis em 2015 obteve um aumento de 0,34% em relação ao ano anterior, sendo o etanol hidratado carburante responsável por 33,3% do total consumido (UNICA, 2016).

### 3.2 Análise de Correlação

No estudo do relacionamento entre duas variáveis, a definição de correlação é de suma importância para o entendimento das variações existentes. A análise de correlação avalia o grau de relacionamento linear entre duas variáveis. De acordo com Triola (1999), ao se considerar duas variáveis, o Coeficiente de Correlação Linear amostral  $r$  mede o grau de relacionamento linear entre os dados emparelhados de  $x$  e  $y$  em uma amostra, e o seu valor pode ser calculado pela equação 1.

$$r = \frac{n \cdot \sum(x \cdot y) - (\sum x) \cdot (\sum y)}{\sqrt{n \cdot (\sum x^2) - (\sum x)^2} \cdot \sqrt{n \cdot (\sum y^2) - (\sum y)^2}} \quad (1)$$

De acordo com Triola (1999) o Coeficiente de Correlação Linear amostral  $r$  deve estar sempre compreendido entre -1 e +1, sendo que valores próximos de -1 e +1 demonstram maior correlação e valores próximos de zero pode-se concluir pela ausência de correlação entre as duas variáveis  $x$  e  $y$  analisadas.

A Tabela 1 apresenta uma forma de como é possível interpretar uma correlação dado o valor numérico.

Tabela 1- Interpretação dos valores de coeficiente de correlação

Valor de $r$ (+ ou -)	Interpretação
0,00 a 0,19	Correlação bem fraca
0,20 a 0,39	Correlação fraca
0,40 a 0,69	Correlação moderada
0,70 a 0,89	Correlação forte
0,90 a 1,00	Correlação muito forte

Fonte: Silvia E. Shimakura, 2006

De acordo com Larson e Farber (2010), pode-se utilizar um teste de hipóteses para determinar se o coeficiente de correlação amostral  $r$  fornece evidência suficiente para concluir que o coeficiente de correlação  $\rho$  de uma população é significativo. A tomada de decisão envolve testar a hipótese  $H_0: \rho=0$  contra  $H_1: \rho \neq 0$ ; sendo  $\rho$  o coeficiente de correlação da população. Rejeita-se  $H_0$  caso o valor absoluto de  $r$  for maior que o valor crítico da tabela *t student*. Ao rejeitar  $H_0$ , conclui-se que existe correlação linear significativa. Em caso contrário, não há evidências suficientes para apoiar a existência de correlação significativa. A

equação 2 apresenta a estatística do teste padronizada para verificação da significância da correlação  $\rho$  de uma população.

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} \quad (2)$$

Onde  $n$  é o tamanho da amostra.

Uma outra forma de verificação da significância da correlação  $\rho$  de uma população é por meio da análise do P-valor, que será utilizado neste estudo. O P-valor é a probabilidade de que a estatística do teste tenha valor extremo em relação ao valor observado (estatística do teste) quando a hipótese nula  $H_0$  é verdadeira (PORTAL ACTION, 2017). Caso o P-valor for menor que o nível de significância  $\alpha$  considerado, então o valor calculado para a estatística do teste está na região crítica e, portanto, rejeita-se a hipótese nula  $H_0$ . Quanto menor for o p-valor, menos provável é hipótese de aceitar  $H_0$  e menor é probabilidade de se cometer um erro do tipo I (rejeitar  $H_0$ , dado que  $H_0$  é verdadeira).

Mannarelli Filho (2005) desenvolveu um estudo da análise da Correlação Estatística e das Análises de Regressão Linear da expansão açucareira da Região Oeste do Estado de São Paulo. O autor utilizou um método exploratório como forma de melhor entender a maneira como está ocorrendo a expansão da indústria açucareira na região estudada, e correlacionando a evolução da produção em relação ao Estado de São Paulo, a nível Brasil e, em especial, os volumes mundiais de açúcar. Tal trabalho possui uma interface com este estudo pois aborda a temática da análise de correlação no setor sucroalcooleiro. O autor analisou um banco de dados referente ao período de 1990 a 2003 e concluiu que a produção brasileira de açúcar, no período abordado, teve aumento sustentado e continuado e que o fator preço, que seria o fator determinante para este aumento na produção, na verdade não representa importância no aumento da produção de açúcar; ao contrário, tendo em vista que um ajustamento linear de regressão entre estas variáveis apresentou um coeficiente angular negativo. Conclui, também, que houve um forte ajuste de correlação positivo, com coeficiente de explicação de 89,93%, entre a produção brasileira de açúcar com a região açucareira Oeste do Estado de São Paulo, de modo a inferir que a produção desta região está integrada economicamente com a produção brasileira de açúcar.

Konishi et al. (2014) também realizou um estudo de análise de correlação entre produção de automóveis biocombustíveis, preços e produção nacional de combustíveis fósseis

e não-fósseis e produção de cana-de-açúcar. O mesmo utilizou dados estatísticos durante o período de 2003 a 2012 e teve como base metodológica a análise de correlação e regressão multivariada. Os resultados encontrados mostraram a existência de uma forte correlação entre produção de cana-de-açúcar e etanol, associada à produção de automóveis com o surgimento dos veículos *flex*. Encontrou-se também um coeficiente de explicação de 82,4% entre a produção de automóveis biocombustíveis e as variáveis escolhidas.

### 3.3 Regressão Linear Simples

A Regressão Linear Simples é explicada por Ribeiro e Caten (s.d.) como uma forma de estimar um modelo, cujo valor da variável dependente  $Y$  é predito pela variável independente  $X$ . Segundo os autores, chama-se Regressão Linear Simples pelo fato de haver apenas uma variável independente e se considerar uma tendência linear entre  $Y$  e  $X$ .

A equação 3 é chamada equação de Regressão Linear Simples:

$$Y = \beta_0 + \beta_1 X + \varepsilon \quad (3)$$

em que  $\beta_0$  representa a interseção,  $\beta_1$  a inclinação da reta e  $\varepsilon$  o erro aleatório.

Segundo o Minitab (2017) esta regressão é responsável por avaliar a relação linear entre duas variáveis, sendo uma a resposta ( $y$ ) e a outra preditora ( $x$ ). Através da relação entre elas é possível se chegar a um valor mais exato da resposta ( $y$ ) levando-se em consideração o valor da preditora ( $x$ ). Esta regressão torna possível encontrar a “linha reta” que melhor se ajusta aos dados e esta pode ser utilizada para avaliar o comportamento da variável resposta quando há mudanças na variável preditora.

Para Formiga (2013), considera-se Regressão Linear Simples “quando o modelo estabelece que a saída depende apenas de um atributo de entrada. É uma função linear de uma única variável”. Dentro deste contexto, também é importante utilizar algumas técnicas estatísticas para verificar a validade do modelo de regressão gerado e os coeficientes obtidos. Para tanto, podem ser utilizados a análise de variância (ANOVA) e os testes  $t$ , respectivamente. Assim como visto para verificação da significância da correlação  $\rho$  de uma população, a análise do  $p$ -valor também pode ser empregada para avaliação da validade da regressão e também dos coeficientes do modelo. A interpretação ocorre da mesma forma: quando  $P$ -valor for inferior ao nível de significância considerado, considera-se o modelo de

regressão linear simples válido e os coeficientes que o compõem como significativos (GUIMARÃES, s.d.).

### 3.4 Análise de Agrupamentos (*Cluster*)

A análise de agrupamentos tem como objetivo desenvolver subgrupos de indivíduos ou objetos, ou seja, classificar uma amostra em um número reduzido de grupos, com base na similaridade entre as entidades. Os grupos não são pré-definidos, utilizando-se assim a técnica para encontrá-los. Tal técnica envolve basicamente três passos: (1) medir, de alguma forma, a similaridade entre as entidades e assim determinar o número de grupos; (2) processo de agrupamento e (3) estabelecer o perfil das variáveis para determinar sua composição (HAIR et al., 2005).

Para Frei (2006) a análise de agrupamentos é colocada como técnicas estatísticas multivariadas, com conotação exploratória, onde grupos de objetos são formados de acordo com algum tipo de critério determinado. A partir dela obtém-se uma matriz de dados que será manuseada por um algoritmo e, que ao final, resultará na formação de grupos homogêneos. Este autor afirma que formação dos grupos é feita de acordo com o nível de similaridade entre as variáveis, e esta similaridade é obtida por meio de coeficientes os quais dependem da escala de mensuração escolhida.

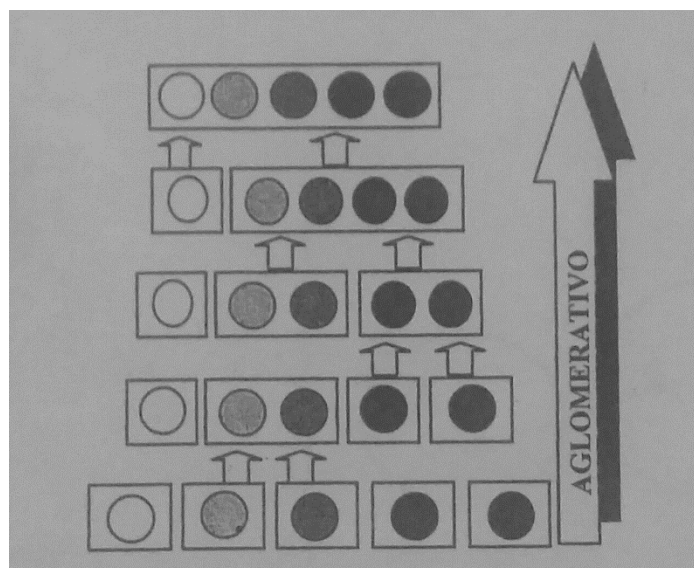
Sabe-se que a similaridade mede o nível de semelhança entre dois objetos, desta forma, quanto maior for o valor deste, maior será a semelhança dos objetos estudados, e a dissimilaridade indica o nível de diferença entre os mesmos, ou seja, quanto maior o seu valor maior a distância entre os objetos. Uma das formas de se medir a distância entre dois elementos amostrais é por meio do cálculo da distância Euclidiana. Esta medida, que foi utilizada neste estudo, compreende uma maneira de mensurar a dissimilaridade entre dois elementos amostrais. Ela corresponde à distância geométrica entre dois pontos de coordenada  $(X_{i1}, \dots, X_{ip})$  e  $(X_{j1}, \dots, X_{jp})$ , e seu cálculo é dado pela Equação 4.

$$d_{ij} = \sqrt{\sum_{f=1}^p (x_{if} - x_{jf})^2} \quad (4)$$

Onde:  $d_{ij}$  é a distância euclidiana e  $X_{if}$  é a representação da  $f$ -ésima medida do  $i$ -ésimo objeto (em que  $i = 1, \dots, n$  e  $f = 1, \dots, p$ ), sabendo que  $n$  são objetos e  $p$  são variáveis contínuas (FREI, 2006).

Outro fator importante à análise *cluster* está relacionado à técnica para a construção de conglomerados, ou grupos. Essa técnica é classificada de duas formas: hierárquicas e não hierárquicas, sendo a primeira dividida em aglomerativa e divisiva (MINGOTI, 2005). Para a realização do estudo optou-se por utilizar a técnica hierárquica aglomerativa, que, segundo a autora, inicialmente cada dado é considerado um *cluster*, em que se tem a partição com a menor dispersão interna. Ao decorrer dos estágios, os grupos que apresentam maiores similaridades vão se unindo até que ao final obtém-se um único grupo composto por todos os demais. A Figura 3 representa a técnica hierárquica aglomerativa.

Figura 3- Técnica Hierárquica Aglomerativa



Fonte: adaptado de Mingoti (2005)

Também é importante que se escolha o critério de parada, ou seja, o critério que irá determinar o melhor número de grupos a serem formados. Existem alguns critérios como: análise do comportamento do nível de fusão (distância), análise do comportamento do nível de similaridade, análise da soma de quadrados entre grupos: coeficiente  $R^2$ , estatística pseudo  $F$  e outros (MINGOTI, 2005). Os critérios utilizados no estudo foram o de análise de comportamento do nível de fusão, que implica em encontrar o maior salto entre as distâncias dos grupos formados. Este salto irá informar o melhor número de grupos a serem estudados. O outro critério é o de análise de comportamento do nível de similaridade, que é semelhante



ao anterior, porém procura-se onde ocorreu um decréscimo acentuado na similaridade dos grupos (MINGOTI, 2005).

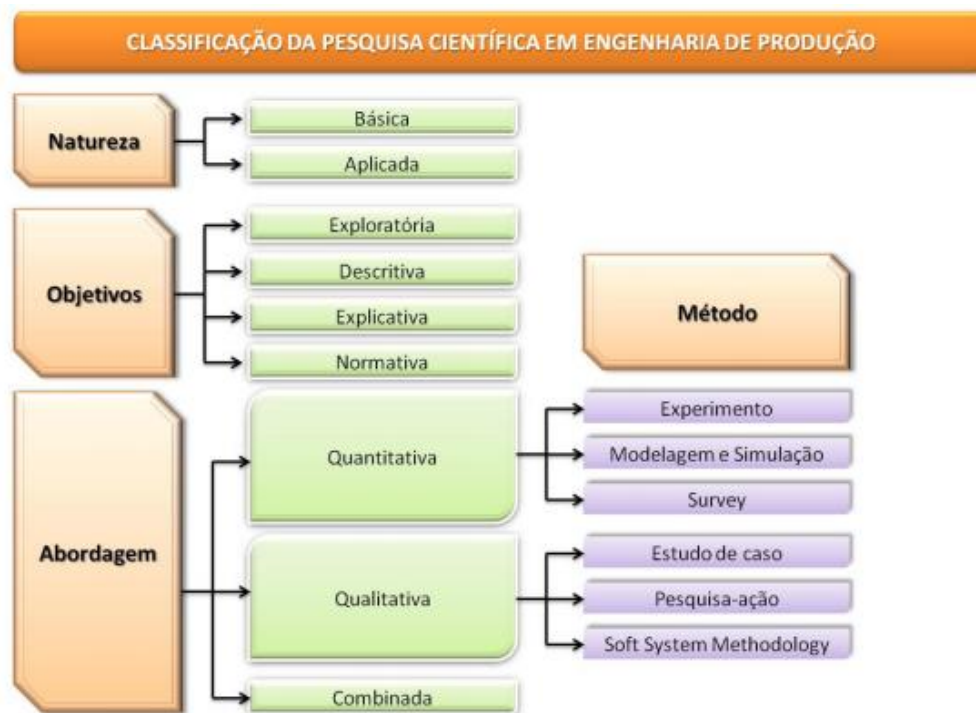
## 4 METODOLOGIA DE PESQUISA

O presente estudo tem como propósito analisar um setor muito importante para a economia brasileira. Para que este possa ser realizado de forma coerente, é importante que se escolha uma metodologia eficaz e que sirva de base para o mesmo. A seguir, será abordada a classificação metodológica e apresentadas as fontes dos dados que serão utilizados posteriormente.

### 4.1 Classificação Metodológica

De acordo com Turrioni e Mello (2012), a pesquisa científica em Engenharia de Produção pode ser classificada da forma apresentada na Figura 4.

Figura 4- Classificação da pesquisa científica em Engenharia de Produção



Fonte: Turrioni e Mello (2012)

O trabalho apresenta-se como uma pesquisa de natureza aplicada, em que os resultados a serem encontrados poderão auxiliar no entendimento de situações reais e desta forma, propor alternativas de melhorias para o setor abordado, como afirma Appolinário (2006 apud

TURRIONI; MELLO, 2012), a pesquisa aplicada é gerada com objetivos comerciais e direcionadas ao mercado. Quanto aos seus objetivos, ela possui aspectos tanto descritivos como explicativos, pois estabelece relações entre variáveis e auxiliam na identificação dos fatores que contribuem para a ocorrência de algum evento. Possui uma abordagem quantitativa, que utiliza técnicas estatísticas para realizar análises e buscar explicações, e por fim, possui métodos que se assemelham mais à modelagem e simulação, uma vez que se obtém um modelo em que a resposta se altera de acordo com modificações propostas ao mesmo (TURRIONI; MELLO, 2012).

#### **4.2 Fonte de Dados**

Para a realização das análises de Correlação e Regressão Linear Simples, utilizou-se banco de dados de algumas organizações que abordam o tema como fonte de dados para entrada. Valores relacionados à produção total de etanol, produção de etanol hidratado e oferta (produção de cana-de-açúcar) / PIB são oriundos do Balanço Energético Nacional (ano base 2015). Já a frota de automóveis movidos à gasolina, à etanol, frota de veículos bicompostíveis e frota total de automóveis estão disponíveis no Relatório da Frota Circulante do Sindicato Nacional da Indústria de Componentes para Veículos Automotivos (SINDIPEÇAS) e Associação Brasileira da Indústria de Autopeças (ABIPEÇAS). Os preços médios da Gasolina e Etanol Hidratado no Brasil são encontrados no Anuário Estatístico Brasileiro do Petróleo, Gás Natural e Biocombustíveis (2016) e por fim, a produção de cana-de-açúcar total localiza-se na União da Indústria De Cana-de-Açúcar (UNICA). Todos os dados apresentados estão em um período de 2006 - 2015.

Por sua vez, a análise de agrupamentos, utilizou da produção de etanol hidratado, vendas de etanol hidratado pelas distribuidoras, preço do etanol hidratado ao consumidor, retirados do Anuário Estatístico Brasileiro do Petróleo, Gás Natural e Biocombustíveis (2016), frota de automóveis e utilitários, densidade demográfica, estimativa da população para 2016, cana-de-açúcar - quantidade produzida, cana-de-açúcar- valor da produção, cana-de-açúcar área plantada e área colhida e rendimento médio, advindos do site do IBGE. Os valores referem-se ao ano de 2015, exceto a densidade demográfica, referente a 2010, e a estimativa da população para 2016, os dois não possuem valores concretos para 2015 devido à realização do último censo no país ter ocorrido em 2010.

Estes bancos de dados foram estruturados utilizando-se planilhas do Microsoft Excel para, posteriormente, passarem por análises estatísticas. Utilizou-se o Minitab 17 para realizar a Análise de Correlação, Regressão Linear Simples e Análise *Cluster* entre as variáveis analisadas. É válido ressaltar que uma das limitações do estudo é que apenas um grupo de variáveis foi utilizado para a realização das análises.

## 5. ANÁLISES

### 5.1 Análise de Correlação e Análise de Regressão Linear Simples

Utilizando-se o Minitab 17, os dados coletados foram analisados por meio de Análise de Correlação. A Tabela 2 corresponde à matriz de correlação entre as variáveis e também apresenta os P-valores correspondentes.

Tabela 2- Correlação entre as variáveis

Variáveis	Produção total etanol	Produção total cana	Frota veículos gasolina	Frota veículos etanol	Frota veículos bicomb.	Frota total de veículos	Oferta (produção cana/PIB)	Preço médio gasolina	Preço médio etanol
Prod. total etanol	<b>1</b>								
Prod. total cana	<b>0,929</b>	<b>1</b>							
Frota veíc. Gasolina	<b>-0,834</b>	<b>-0,904</b>	<b>1</b>						
Frota veíc. Etanol	<b>-0,613</b>	<b>-0,809</b>	<b>0,887</b>	<b>1</b>					
Frota veíc. Bicomb.	<b>0,697</b>	<b>0,853</b>	<b>-0,954</b>	<b>-0,975</b>	<b>1</b>				
Frota total de veículos	<b>0,672</b>	<b>0,838</b>	<b>-0,941</b>	<b>-0,978</b>	<b>0,999</b>	<b>1</b>			
Oferta (prod. cana/PIB)	<b>0,720</b>	<b>0,755</b>	<b>-0,739</b>	<b>-0,661</b>	<b>0,708</b>	<b>0,696</b>	<b>1</b>		
Preço médio gasolina	<b>0,554</b>	<b>0,627</b>	<b>-0,867</b>	<b>-0,828</b>	<b>0,868</b>	<b>0,863</b>	<b>0,643</b>	<b>1</b>	
Preço médio etanol	<b>0,380</b>	<b>0,569</b>	<b>-0,793</b>	<b>-0,916</b>	<b>0,900</b>	<b>0,908</b>	<b>0,578</b>	<b>0,905</b>	<b>1</b>

Fonte: Elaborado pela autora

Através da Tabela 2 é possível observar que a produção total de etanol e produção total de cana-de-açúcar possui uma correlação muito forte e positiva, indicando que quanto maior for a produção de cana maior será a oferta de etanol ( $r = 0,929$  e P-valor = 0,000). Já a frota de veículos à gasolina possui uma correlação forte ( $r = -0,834$  e P-valor = 0,003), porém inversa com a produção total de etanol, uma explicação possível para isso seria o fato de que o aumento da utilização de veículos que utilizam apenas a gasolina poderia reduzir o nível de utilização do etanol, ou seja, reduzindo o consumo deste combustível a produção e oferta do

mesmo seriam prejudicadas. É válido ressaltar que nos casos em que o P-valor das variáveis é menor que o nível de significância 0,05, significa que o coeficiente de correlação amostral  $r$  é significativo para explicar o coeficiente de correlação populacional  $\rho$ . Ou seja, nestes casos o coeficiente de correlação amostral é válido para explicar o grau de relacionamento entre as variáveis. Lembrando que esta é uma forma de teste de hipótese estatístico, conforme descrito na revisão bibliográfica.

Complementar à Análise de Correlação, foram realizadas as Regressões Lineares Simples para produção total de etanol e produção de cana-de-açúcar, e entre a produção total de etanol e a frota de veículos à gasolina. Escolheram-se tais variáveis por serem as mais significativas para explicar o comportamento da variável resposta produção de etanol. A Tabela 3 ilustra a análise de variância (ANOVA) para o modelo de regressão entre produção total de etanol e produção de cana-de-açúcar. Por meio desta tabela, é possível validar o modelo de regressão, uma vez que o p-valor correspondente é menor que o nível de significância  $\alpha$  considerado (0,05), e também afirmar que produção de cana é representativa para o modelo, também por possuir um p-valor é menor que 0,05.

Tabela 3- ANOVA para a regressão entre produção total de etanol e produção de cana-de-açúcar

Fonte	Df	Adj SS	Adj MS	F-Valor	P-Valor
Regressão	1	107386475	107386475	50,78	0,000
prod. cana (10 <sup>3</sup> m <sup>3</sup> )	1	107386475	107386475	50,78	0,000
Erro	8	16918212	2114776		
Total	9	124304686			

Fonte: a autora

Através da Tabela 4 pode-se verificar o percentual de variabilidade explicada pelo modelo, sendo ele,  $R^2$ , 86,39%, ou seja, a produção de cana de açúcar é capaz de explicar mais de 86% da produção de etanol total. Já o  $R^2_{\text{ajustado}}$  possui um valor de 84,69%, ou seja, obteve-se uma diminuição de 1,7% do percentual de variabilidade explicada pelo modelo e o  $R^2_{\text{pred}}$  mostra que o poder de predição do modelo é de 81,12%.

Tabela 4- Estatísticas de Regressão para o modelo formado entre produção total de etanol e produção de cana-de-açúcar

Estatísticas de Regressão	
$R^2$	86,39%
$R^2_{ajustado}$	84,69%
$R^2_{pred}$	81,12%

Fonte: Elaborado pela autora

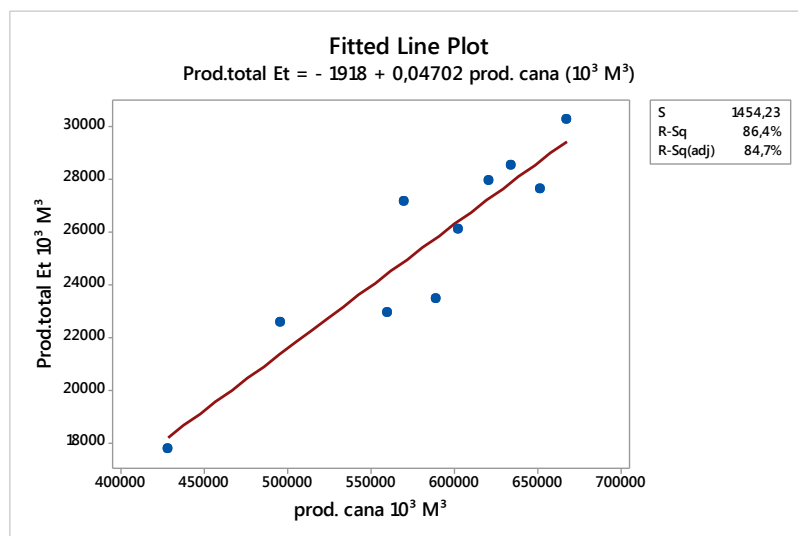
Já a Tabela 5 representa os testes t para verificação dos coeficientes da equação de regressão obtida. Nota-se que os p-valores, tanto para os coeficientes da constante quanto a variável produção de cana-de-açúcar, são iguais a 0,000, o que indica que ambos são significativos. Já a Figura 5 ilustra a equação do modelo de Regressão Linear Simples ajustado. A cada  $1 \times 10^3 \text{ M}^3$  adicionado de produção de cana-de-açúcar, a produção de etanol aumenta em  $0,04702 \times 10^3 \text{ m}^3$ . Ela também mostra que os pontos estão próximos de formarem uma reta e confirma o que foi encontrado pela equação de regressão, que à medida que a produção de cana aumenta a produção de etanol também aumenta.

Tabela 5- Testes t para os coeficientes da regressão entre produção total de etanol e produção de cana-de-açúcar

Termos	Coef.	SE Coef	T-Valor	P-Valor	VIF
Constant	-1918	3865	-0,50	0,633	
prod. cana ( $10^3 \text{ m}^3$ )	0,04702	0,0066	7,13	0,000	1,000

Fonte: a autora

Figura 5 - Produção de etanol X Produção de cana



Fonte: Elaborado pela autora

A Tabela 6 indica a ANOVA para o modelo de Regressão Linear Simples entre produção total de etanol e frota de automóveis à gasolina. A partir dela pode-se verificar a validação do modelo de regressão, em que novamente o p-valor correspondente é menor que o nível de significância  $\alpha$  utilizado (0,05) e confirma que a frota de veículos à gasolina é representativa para o modelo.

Tabela 6- ANOVA para a regressão entre produção total de etanol e frota de automóveis à gasolina

Fonte	Df	Adj SS	Adj MS	F-Valor	P-Valor
Regressão	1	86414591	86414591	18,25	0,003
Frota Veíc.-Gasol.	1	86414591	86414591	18,25	0,003
Erro	8	37890095	4736262		
Total	9	124304686			

Fonte: Elaborado pela autora

A Tabela 7 indica um percentual de variabilidade explicada ( $R^2$ ) de 69,52%, concluindo assim que a frota de automóveis à gasolina é capaz de explicar cerca de 69% da produção total de etanol. O  $R^2_{\text{ajustado}}$  possui um valor de 65,71%, obtendo-se uma diminuição de 3,81% do percentual de variabilidade explicada pelo modelo, por sua vez, o  $R^2_{\text{pred}}$  mostra que o poder de predição do modelo é de 53,59%. É possível inferir que esta variável apresenta uma capacidade de explicação da variável resposta inferior à variável produção de cana-de-açúcar.

Tabela 7- Estatísticas de Regressão para o modelo formado entre produção total de etanol e frota de veículos à gasolina

Estatísticas de Regressão	
$R^2$	69,52%
$R^2_{\text{ajustado}}$	65,71%
$R^2_{\text{pred}}$	53,59%

Fonte: Elaborado pela autora

Pela Tabela 8, nota-se que os P-valores encontrados possuem os valores de 0,000 e 0,003, indicando que ambos são significativos. A Figura 6 ilustra, por meio da equação de



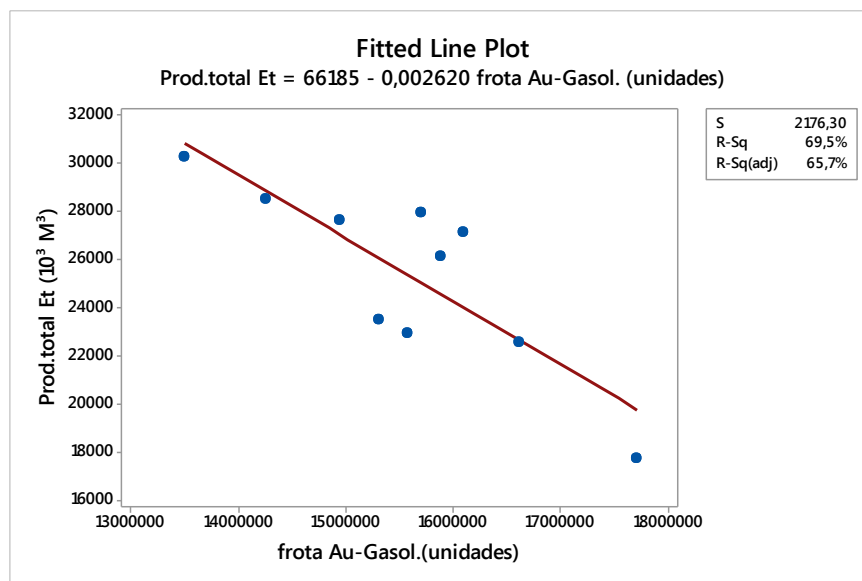
Regressão Linear Simples obtida, para cada 1 unidade de veículo à gasolina produzida, a produção de etanol diminui em  $0,002620 \times 10^3 \text{ m}^3$  por ano. Nota-se, também, que os pontos estão mais dispersos se comparados com os da Figura 6, fato confirmado pelas estatísticas de regressão na Tabela 7, porém com grande tendência a formarem uma reta com relação inversa entre a variável resposta e a preditora, o que significa que quanto maior o número de veículos à gasolina, menor será produção de etanol.

Tabela 8- Testes t para os coeficientes da regressão entre produção total de etanol e frota de veículos à gasolina

Termos	Coef	SE Coef	T-Valor	P-Valor	VIF
Constant	66185	9567	6,92	0,000	
Frota Au-Gasol.	-0,002620	0,000613	-4,27	0,003	1,000

Fonte: Elaborado pela autora

Figura 6 - Produção de etanol X Frota de automóveis à gasolina

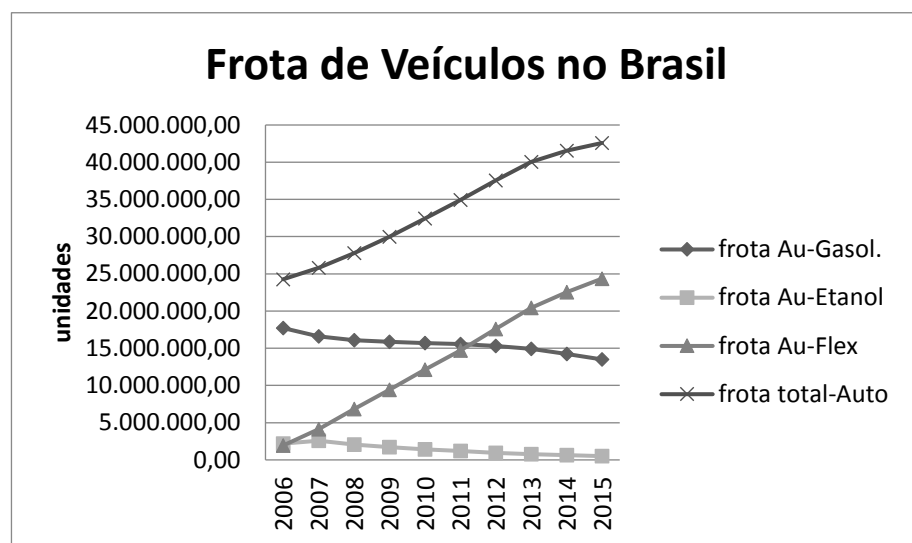


Fonte: Elaborado pela autora

Analisando a correlação entre as demais variáveis, por meio da Tabela 2, o valor de 0,999, que corresponde à relação entre Frota total veículos X Frota veículos bicombustíveis, pode-se dizer que as variáveis possuem uma correlação muito forte e que estão associadas positivamente, ou seja, à medida que uma delas aumenta a outra também aumenta. O mesmo acontece com o valor de 0,900 (Preço médio etanol X Frota veículos bicombustíveis).

Nos dois casos é possível explicar os valores levando em conta a atual situação do país, e o que já vem ocorrendo desde alguns anos anteriores. O Brasil tem apresentado bons números em relação à frota de veículos bicompostíveis, também conhecido como *flex*, o que vem aumentando consideravelmente ao longo dos anos, enquanto que os veículos que utilizam de apenas um tipo de combustível (gasolina ou etanol) vê seus números caindo. Tal afirmação pode ser confirmada ao observar a Figura 7 que mostra a frota de veículos no Brasil de 2006 a 2015.

Figura 7 - Frota de Veículos no Brasil



Fonte: Adaptado do Relatório de Frota Circulante de 2014 e 2016 do SINDIPEÇAS

É válido ressaltar que, como indicado na Tabela 2, quando o preço médio do etanol aumenta, a frota total de veículos também aumenta. A afirmação se mostra em seu primeiro instante em uma certa contradição, mas isto pode ser explicado quando analisa-se que o consumidor, ao possuir um veículo bicompostível, possui também a possibilidade de utilizar a opção que mais lhe favoreça. Este resultado está de acordo com os encontrados em estudo de Emissões e Tecnologia da União da Indústria de Cana-de-açúcar (UNICA), onde se descreve que a principal vantagem dos veículos flex em relação aos movidos à combustão ou à eletricidade é a liberdade de escolha proporcionada ao consumidor na hora do abastecimento e os inúmeros benefícios socioeconômicos e ambientais advindos da utilização do biocombustível anidro e do hidratado (UNICA, 2017).

Em situações contrárias, ainda conforme a Tabela 2, tem-se o valor de -0,975 (Frota veículos bicompostíveis X Frota veículos etanol) e o valor -0,954 (Frota veículos

bicombustíveis X Frota veículos gasolina). Ambos os casos possuem forte correlação, porém são associados negativamente, isso determina que quando um fator aumenta o outro tende a diminuir.

É possível chegar a essa conclusão novamente através do Figura 7, que mostra que ao passar dos anos as unidades de veículos bicombustíveis aumentou em mais de 20 vezes, o que não ocorreu com a frota dos demais. A médio prazo, o carro *flex* continuará figurando no segmento automotivo nacional como a melhor opção de compra para o consumidor e, segundo estudo realizado pela Accenture Research, o principal motivo para isso é o uso em larga escala do etanol em modelos bicombustíveis, que apresentam melhor custo/benefício em relação aos veículos elétricos e híbridos à gasolina (UNICA, 2017).

A crise no setor sucroenergético, iniciada em 2008, foi causada principalmente pela forte recessão na economia dos Estados Unidos, afetando o mundo. Junto a ela, vieram o congelamento no preço do etanol, reflexo do controle sobre o valor da gasolina imposto pelo governo, as expectativas com o petróleo do pré-sal e a escassez do crédito bancário (DASSIE, 2016). Tal cenário se manteve até 2015, a partir de então se vê a possibilidade de melhora nos próximos anos, como por exemplo, no segundo semestre de 2015 começaram a ocorrer mudanças para o etanol e o açúcar, e isso passou a gerar uma receita maior por hectare de cana produzida e há boas perspectivas. No caso do Brasil a meta é que os biocombustíveis, que hoje respondem por 15% da matriz energética do País, passem a representar 18% em 2030 e que, até 2050, sejam produzidos 50 bilhões de litros de etanol por ano, 20 bilhões de litros acima da atual produção (CAETANO; PALHARES, 2016).

O fator que mais contribuiu para aumentar a demanda do etanol nos últimos anos foi o aumento da mistura de etanol anidro na gasolina, passando de 25% para 27%. Vale ressaltar que voltou-se a reaver parcialmente a Contribuição de Intervenção no Domínio Econômico (Cide) no combustível fóssil, saltando de zero para R\$ 0,10/litro. Houve também o aumento no preço da gasolina e no valor do Imposto sobre Circulação de Mercadorias e Serviços do mesmo combustível em vários Estados, resultando no aumento da demanda por etanol hidratado. E para aproveitar o cenário positivo para o setor é preciso ajustar as tecnologias de agricultura de precisão e investir em ganhos de produtividade (CAETANO; PALHARES, 2016).

A produção deste carburante gera quase 3 milhões de empregos (direto e indiretos) e renda direta em 1.000 municípios, além de ser de origem limpa e renovável ajuda a combater

o aquecimento global. O potencial de redução das emissões de gases de efeito estufa com o uso do etanol em carros *flex*, desde que os mesmos chegaram ao país em 2003 até 2016, evitou que 370 milhões de toneladas de CO<sub>2</sub> fossem lançadas na atmosfera, e o combustível ‘verde’ emite até 90% menos o mesmo gás se comparado à gasolina (UNICA, 2017).

### 5.1 Análise de Agrupamentos (*Cluster*)

De acordo com Mingoti (2005, p. 155), “ a análise de agrupamentos tem como objetivo dividir os elementos da amostra de forma que os elementos pertencentes a um mesmo grupo sejam similares entre si com respeito às variáveis que neles foram medida”. Neste sentido, tal análise foi aplicada no contexto com o objetivo de identificar o número de grupos (*cluster*) ideais, ou seja, a quantidade de grupos que é formada de acordo com o critério de parada escolhido. Como citado anteriormente, as variáveis que possibilitaram a realização de tal análise foram: produção de etanol hidratado, vendas de etanol hidratado pelas distribuidoras, preço do etanol hidratado ao consumidor, frota de automóveis e utilitários, densidade demográfica, estimativa da população para 2016, cana-de-açúcar - quantidade produzida, cana-de-açúcar- valor da produção, cana-de-açúcar área plantada e área colhida e rendimento médio.

Os critérios de parada adotados no estudo correspondem ao nível de dissimilaridade, que consiste no momento em que ocorre o maior salto nos níveis de distância entre os clusters e também ao salto da similaridade, ou seja, onde ocorre o maior salto na porcentagem de similaridade entre os grupos, como explicado no tópico 3.3.

Para melhor compreensão, é válido ressaltar que no início da análise cada estado representa um grupo, e à medida que vão ocorrendo os próximos passos os grupos vão se unindo, ou seja, os grupos que possuem maior nível de similaridade, ou menor distância, formam um novo grupo. Através da Tabela 9 pode-se visualizar o seguinte exemplo: no Passo 1, o número de grupos são 18, os *clusters* 6 e 10, formados por Rio Grande do Norte e Sergipe, se uniram e formaram o novo grupo denominado Grupo 6, o nível de Similaridade entre eles foi de 95,6584% (valor mais alto) e a Distância correspondente é de 0,55417 (menor valor). Já no Passo 5, o número de grupos passou a 14, onde ocorreu a união dos grupos 3 e 9, este novo grupo é constituído por 5 estados e seus níveis de Similaridade e Distância correspondem a 93,4162% e 0,84038 respectivamente.

Tabela 9- Análise Cluster no setor Sucroenergético brasileiro

Passo	Número de Grupos	Nível de Similaridade	Nível de Distância	Grupo Unido	Novo Grupo	Nº de obs. no Grupo
1	18	95,6584	0,55417	6 10	6	2
2	17	94,8407	0,65854	3 6	3	3
3	16	94,1802	0,74285	3 4	3	4
4	15	93,6811	0,80656	5 13	5	2
5	14	93,4162	0,84038	3 9	3	5
6	13	92,9625	0,89829	7 8	7	2
7	12	92,7625	0,92381	1 3	1	6
8	11	91,7323	1,05531	5 7	5	4
9	10	91,1546	1,12905	1 5	1	10
10	9	90,2585	1,24343	1 11	1	11
11	8	89,5100	1,33897	12 16	12	2
12	7	89,4608	1,34526	1 2	1	12
13	6	89,3861	1,35479	17 19	17	2
14	5	86,5749	1,71361	17 18	17	3
15	4	86,4516	1,72935	12 17	12	5
16	3	85,9563	1,79257	1 12	1	17
17	2	71,7032	3,61188	1 14	1	18
18	1	24,4263	9,64642	1 15	1	19

Partição Final  
Número de Grupos: 6

Fonte: Elaborado pela autora

A Tabela 9 mostra que o número de *clusters* (grupos) ideais para o estudo em questão seriam seis. Chega-se a esta conclusão pelo fato dos maiores saltos entre os níveis, tanto de similaridade quanto de distância, terem ocorrido quando se têm uma diminuição de números de *cluster* de seis para cinco, ou seja, o algoritmo realiza a parada no passo anterior ao primeiro maior salto dos valores encontrados. Em relação ao nível de similaridade, com seis grupos o valor era de 89,3861 e com cinco caiu para 86,5749, uma redução de 2,8112. Já o nível de distância, que é calculado através da distância Euclidiana e que faz uma comparação de cada par de estados com cada uma das 19 variáveis estudadas, mostra que com seis grupos o valor era de 1,35479 e com cinco esse número passou para 1,71361, um aumento de 0,35882.

A tabela acima também indica que o nível de similaridade entre os dois últimos grupos unidos foi alto (89,3861), indicando que os estados pertencentes a um mesmo grupo são muito similares entre si, para os seis agrupamentos finais formados.

A Tabela 10 resume cada grupo da partição final através das seguintes informações:

- *Número de observações:* número de observações em cada grupo, o primeiro *cluster* contém doze estados, o segundo dois, o terceiro e o quarto são formados por apenas um estado, o *cluster* cinco possui dois e o último tem apenas um.
- *Soma de quadrados dentro do grupo:* mede a variabilidade dentro de cada grupo, quanto menor o valor mais similar são as variáveis componentes do grupo. No caso apresentado, os *clusters* três, quatro e seis possuem valores zero, isso porque cada um é composto por apenas um estado, com isso não se tem variabilidade nos mesmos. Por sua vez, o *cluster* um, composto por doze estados, possui a maior variabilidade apresentando um valor de 15,9308. Tal resultado está diretamente ligado ao fato do *cluster* um ser o mais numeroso em termos de número de estados que o compõe.
- *Distância média do centroide:* indica a distância média das observações ao centroide do grupo. Por ela também pode-se notar que a maior distância está presente no *cluster* um (1,07889), aquele que possui maior número de estados. É válido ressaltar que o centroide se refere ao vetor de média das variáveis para as observações naquele grupo, e é usado como ponto central de um grupo.
- *Distância máxima do centroide:* mostra a distância máxima das observações ao centroide do grupo. Pela Tabela 10, observa-se que o grupo um também possui a maior distância (1,80458), porém os grupos três, quatro e seis possuem valores iguais à zero e os grupos dois e cinco possuem valores próximos, 0,66949 e 0,67739 respectivamente.

Tabela 10- Explicação dos *Clusters*

	Nº de Obs.	Soma dos Quadrados dentro do Grupo	Distância Média do Centoide	Distância Máxima do Centoide
Grupo 1	12	15,9308	1,07889	1,80458
Grupo 2	2	0,8964	0,66949	0,66949
Grupo 3	1	0,0000	0,00000	0,00000
Grupo 4	1	0,0000	0,00000	0,00000
Grupo 5	2	0,9177	0,67739	0,67739
Grupo 6	1	0,0000	0,00000	0,00000

Fonte: Elaborado pela autora

A Tabela 11 mostra as distâncias entre os grupos formados. Por ela conclui-se que os grupos menos distantes, ou seja, mais semelhantes são os *Clusters* 2 e 5. Porém os que apresentam a maior distância são os *Clusters* 1 e 4.

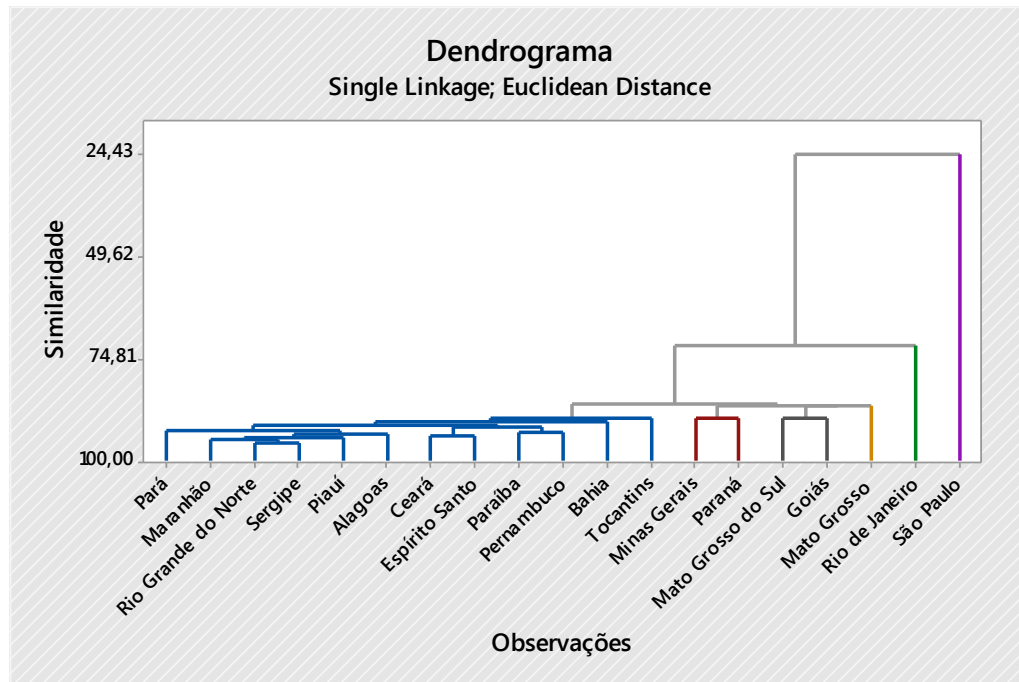
Tabela 11- Distâncias entre os centros dos centroides

	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5	Cluster6
Cluster1	0,0000	2,9773	4,0922	12,3368	2,9083	2,6277
Cluster2	2,9773	0,0000	5,0549	10,0870	1,7750	2,2072
Cluster3	4,0922	5,0549	0,0000	12,0786	5,7127	5,5972
Cluster4	12,3368	10,0870	12,0786	0,0000	10,8099	11,7773
Cluster5	2,9083	1,7750	5,7127	10,8099	0,0000	1,8830
Cluster6	2,6277	2,2072	5,5972	11,7773	1,8830	0,0000

Fonte: elaborado pela autora

A Figura 8 indica a composição de cada *cluster*, através dela é possível identificar que o estado de São Paulo forma um grupo, isso pode ser explicado pelo fato do mesmo apresentar valores discrepantes, em quase todos os fatores estudados, quando comparados aos demais estados, como pode ser visualizado nas Tabelas 12 e 13, que indicam os estados, as variáveis utilizadas para análise, seus respectivos valores e a colocação de cada estado.

Figura 8 - Dendrograma dos Clusters formados



Fonte: Elaborado pela autora

Como se pode observar na Tabela 12, São Paulo possui os maiores valores de produção de etanol hidratado (7.555,87 Mil M<sup>3</sup>), vendas de etanol hidratado pela distribuidora (9.456,06 Mil M<sup>3</sup>), frota de automóveis e utilitários (17.037.195 unidades), quantidade de cana produzida (423.419.511 t) e outros. Ele se destaca em vários dos critérios abordados, como é possível visualizar através da Tabela 13.

No Brasil, grande parte das usinas de açúcar e bioetanol concentram-se na região noroeste do estado de São Paulo, este estado é responsável por aproximadamente 60% da produção de cana, açúcar e etanol do país (NOVACANA, n.d), o que explica os bons números de produção de etanol hidratado e quantidade de cana produzida. Ele foi responsável por produzir, na safra de 2014/15, 53,4% da produção nacional de cana-de-açúcar, 48,4% da produção de etanol e 61,6% da produção de açúcar e também foi responsável por consumir 52,9% de etanol consumido no Brasil em 2015, cerca de 9.456 mil m<sup>3</sup> (UNICA, 2016).



Tabela 12- Estados e variáveis da análise *Cluster*

Estados	PROD. ET. Hid. (Mil M <sup>3</sup> )	VEN. ET. MIO. pela dist. (Mil M <sup>3</sup> )	PREÇO MIO. ao cons. (R\$/L)	FROTA Auto/util. (Un)	DENS. dem. (Hab/Km <sup>2</sup> )	EST. POP. 2016 (Pes.)	CANA QTD Prod. (t)	LANA VAL. prod. (Mil R\$)	CANA ÁREA plant. (ha)	LANA ÁREA colhida (ha)	LANA REND. Méd. (Kg/ha)
PA	11,14	52,97	2,942	535386	17,65	8272724	918297	77613	13301	13301	69040
TO	61,86	56,69	2,636	189056	3,36	1532902	2855536	162161	36935	36287	78693
MA	42,28	57,21	2,735	390669	19,81	6954036	3124141	224314	47685	47685	65516
PI	3,34	38,4	2,727	309063	12,4	3212180	861393	80853	15022	15022	57342
CE	14,6	173,59	2,682	1040163	56,76	8963663	906912	98436	19149	19149	47361
RN	34,11	68,9	2,699	501800	59,99	3474998	3724046	280543	59491	59487	62603
PB	238,35	131,57	2,399	471909	66,7	3999415	6801981	425356	120559	120504	56446
PE	262,16	279,68	2,492	1220271	89,62	9410336	15965218	1199039	315467	309487	51586
AL	221,51	58,71	2,641	317080	112,33	3358963	20714441	1643489	308006	308006	67253
SE	134,76	45,48	2,646	301100	94,36	2265779	3220415	189661	57258	54097	59530
BA	175,07	508,12	2,576	1638873	24,82	15276566	6227728	456300	104863	104709	59477
MG	2098,06	1789,94	2,317	5492880	33,41	20997560	69017764	4074646	917878	910927	75767
ES	91,8	63,51	2,795	890601	76,25	3973697	3320809	124069	76683	76683	43306
RJ	53,8	664,32	2,732	4243995	365,23	16635996	3919509	204141	79388	79388	49372
SP	7555,87	9456,06	2,1	17037195	166,23	44749699	423419511	22642728	5576838	5527560	76602
PR	924,33	1690,43	2,255	4153178	52,4	11242720	47368045	2601666	626375	626198	75644
MS	2066,2	231,7	2,369	662186	6,86	2682386	43924003	2819025	546099	545650	80498
MT	792,81	699,3	2,099	632094	3,36	3305531	20077293	1211122	291100	289091	69450
GO	3713,35	1240,76	2,29	1700158	17,65	6695855	70412725	4885659	911847	911847	77220

Fonte: adaptado de Anuário Estatístico Brasileiro do Petróleo, Gás Natural e Biocombustíveis (ANP 2016) e Instituto Brasileiro de Geografia e Estatística (IBGE)

Tabela 13 - Ordem de Classificação entre Estados

Estados	Prod. Etanol Hidr. (Mil M <sup>3</sup> )	Venda Etanol Hidr. Pela Distrib. (Mil M <sup>3</sup> )	Preço Etanol Hidr. ao cons. (R\$/L)	Frota Autom/ Utilitários (Un)	Densidade Demog. (Hab/Km <sup>2</sup> )	Estimativa População para 2016 (Pessoas)	Cana Qtd produzida (t)	Cana valor da produção (Mil R\$)	Cana área plantada (ha)	Cana área colhida (ha)	Cana rendimento médio (Kg/ha)
PA	18	17	19	12	14	8	17	19	19	19	8
TO	13	16	10	19	19	19	16	15	16	16	2
MA	15	15	17	15	13	9	15	12	15	15	10
PI	19	19	15	17	16	16	19	18	18	18	14
CE	17	10	13	8	9	7	18	17	17	17	18
RN	16	12	14	13	8	13	12	11	13	13	11
PB	8	11	7	14	7	11	9	10	9	9	15
PE	7	8	8	7	5	6	8	8	6	6	16
AL	9	14	11	16	3	14	6	6	7	7	9
SE	11	18	12	18	4	18	14	14	14	14	12
BA	10	7	9	6	12	4	10	9	10	10	13
MG	3	2	5	2	11	2	3	3	2	2	5
ES	12	13	18	9	6	12	13	16	12	12	19
RJ	14	6	16	3	1	3	11	13	11	11	17
SP	1	1	2	1	2	1	1	1	1	1	4
PR	5	3	3	4	10	5	4	5	4	4	6
MS	4	9	6	10	17	17	5	4	5	5	1
MT	6	5	1	11	18	15	7	7	8	8	7
GO	2	4	4	5	15	10	2	2	3	3	3

Fonte: Elaborado pela autora

Outro estado que também constitui um *cluster* é o Rio de Janeiro. Ele produziu cerca de 53,8 Mil M<sup>3</sup> de etanol hidratado em 2015, valor que se encontra entre os mais baixos dos 19 estados estudados, os números relacionados à área de cana plantada e colhida, quantidade de cana produzida também são pequenos. Porém, possuiu a maior densidade demográfica (365, 23 hab/Km<sup>2</sup>), como pode ser identificado na Tabela 12, os valores referentes à frota de automóveis e utilitários (4.243.995 unidades) e estimativa da população para 2016 (16.635.996) deste estado estão entre as três maiores. O que se observa é que o Rio de Janeiro se destacou dos demais pelos valores das variáveis socioeconômicas e não pelas variáveis mais voltadas ao setor sucroenergético, dado que este estado apresenta baixa relevância com relação às variáveis específicas do setor estudado.

Mato Grosso também é considerado um grupo. Ele possui, segundo a Tabela 12, o menor valor do preço do etanol hidratado ao consumidor (2,099 R\$/L) e também a menor densidade demográfica (3,36 Hab./Km<sup>2</sup>), tais fatos também podem ser confirmados através da Tabela 13. Este estado não possui boa representatividade na produção de etanol hidratado, cerca de 792,81 Mil M<sup>3</sup> referente ao ano de 2015, porém os valores de quantidade de cana produzida (20.077.293 t), área plantada (291.100 ha) e área colhida (289.091 ha) estão entre os mais significativos.

Algumas possíveis explicações para os números acima seriam as condições em que o setor se encontrava no período. Segundo o G1 (2015) após os reajustes dos preços da gasolina e do diesel, respectivamente 6% e 4%, se tornou mais vantajoso abastecer os veículos com o etanol. Este chegou a preços muito baixos, em alguns locais a 1,77 R\$/L, a causa apontada para tais valores seria resultado de fatores como a política tributária do governo do estado e uma oferta maior que a demanda. No estado do Mato Grosso, o açúcar tem remunerado melhor os produtores (G1, 2014), este estado também é o principal produtor de milho no país, e recentemente, a viabilidade de produção de etanol a partir deste cereal foi aprovada, e mostra que essa nova medida aumentaria o valor agregado do produto, além de trazer benefícios sociais e ambientais (FOLHA DE SÃO PAULO, 2017). Os fatos anteriores poderiam ser possíveis explicações para o pequeno número de produção de etanol hidratado a partir da cana-de-açúcar em Mato Grosso.

Pela Figura 8, outro grupo (*cluster*) é formado por Mato Grosso do Sul e por Goiás. Os dois estados possuem valores consideráveis de produção de etanol hidratado, 2.066,2 Mil M<sup>3</sup> em MS e 3.713,35 Mil M<sup>3</sup> em GO, os preços do etanol ao consumidor são próximos, os números para a quantidade de cana produzida são significativos, sendo esses valores

43.924.003 t e 70.412.725 t respectivamente, e em ambos os estados os valores da área de cana plantada e área colhida foram quase semelhantes.

Segundo a Unica (2017) Goiás é o segundo maior produtor de cana no Brasil, ele produziu na safra de 2016/17 mais de 67 milhões de toneladas de cana e a partir daí obteve-se mais de 2,10 milhões de toneladas de açúcar e 4,31 bilhões de litros de etanol (anidro e hidratado). Mato Grosso do Sul possui 22 unidades em operação, para produção de etanol e açúcar, e segundo a Biosul (Associação dos Produtores de Bioenergia do Mato Grosso do Sul), tinha-se uma expectativa de crescimento da safra de 2015/16, o que resultaria em crescimentos de 15% no processamento de cana, 22% na produção de açúcar e 18% de etanol (FIEMS, 2015).

Há outro *cluster* formado por dois estados, são eles Minas Gerais e Paraná. Ambos possuem bons números de vendas de etanol pela distribuidora, como se pode observar na Tabela 13. Eles possuem valores altos de frota de automóveis e utilitários, 5.492.880 unidades e 4.153.178 unidades respectivamente, também valores consideráveis são encontrados nas quantidades de cana produzidas, 69.017.764 t em MG e 47.368.045 em PR. Os números de rendimento médio de cana são muito próximos, 75.767 Kg/ha (MG) e 75.644 Kg/ha (PR).

Minas Gerais é o terceiro maior produtor de cana-de-açúcar no país, ele fica atrás somente de São Paulo e Goiás (UNICA, 2013), como pode ser constatado na Tabela 13. Da safra de 2015/16, este estado foi responsável pelo plantio de aproximadamente 9% de um total de 8.995,5 mil hectares de cana-de-açúcar (CONAB, 2015). Em 2013 ele possuía 49 indústrias canaveiras e destas, a maior parte (28) se concentravam na região do Triângulo Mineiro/Alto Paranaíba (CAMPOS, 2014). Por sua vez, o Paraná possui uma área de 640 mil hectares de cultivo de cana, o que corresponde a 9% da área plantada no estado e que estão concentradas em 154 municípios (AGÊNCIA DE NOTÍCIAS DO PARANÁ, 2015).

Ao observar a Tabela 12, é possível identificar que os *clusters* formados por Mato Grosso do Sul e Goiás e Minas Gerais e Paraná possuem algumas variáveis com valores próximos, como o preço do etanol hidratado ao consumidor, rendimento médio de cana e quantidade de cana produzida, porém o que mais os diferem seriam os valores da densidade demográfica e estimativa da população. Minas Gerais e Paraná possuem valores acima de 30 hab/Km<sup>2</sup> para a densidade demográfica, 33,41 hab/Km<sup>2</sup> e 52,4 hab/Km<sup>2</sup> respectivamente, enquanto Mato Grosso do Sul e Goiás não chegam a 20 hab/Km<sup>2</sup>, 6,86 hab/Km<sup>2</sup> e 17,65 hab/Km<sup>2</sup>. Para a estimativa da população para 2016 o segundo grupo também possui os

maiores números, Minas com 20.997.560 pessoas e Paraná com 11.242.720 pessoas, em situações contrárias Mato Grosso do Sul tem 2.682.386 pessoas e Goiás 6.695.855 pessoas.

Por fim, tem-se o último e maior *cluster*, composto por doze estados, como indica a Figura 8. Pela Tabela 12 observa-se que os mesmos possuem os menores valores para a produção de etanol hidratado, a maior parte deles obtiveram valores iguais, ou sofreram pequenas reduções, nas variáveis área de cana plantada e área colhida. Os números de venda de etanol hidratado pela distribuidora também foram menores, apenas a Bahia atingiu um valor pouco acima de 500 Mil M<sup>3</sup>. Paraíba e Pernambuco foram os únicos estados com preços de etanol hidratado ao consumidor abaixo de 2,5 R\$/L, mesmo assim eram valores próximos, 2,399 R\$/L e 2,492 R\$/L respectivamente.

Observa-se que este grupo formado por doze estados é composto principalmente por estados das regiões norte e nordeste. A região nordeste, com maior representatividade na quantidade de estados no grupo, possui uma estrutura produtiva, de acordo com Cunha (2016), de 60 unidades produtoras e ela é responsável por aproximadamente 260.000 a 300.000 empregos diretos. Segundo o autor, a produção média de cana nesta região, na safra de 2010/11 a 2015/16, foi de cerca de 3,7 milhões de toneladas, com uma produção de 1,8 milhões de M<sup>3</sup> de etanol no mesmo período. Os estados que representam a região norte são Tocantins e Pará. Segundo a Unica (2013), os estados do norte do Brasil produziram na safra de 2012/13 cerca de 3.100,4 toneladas de cana-de-açúcar e deste total o Tocantins foi responsável por 60% da produção. Desta forma ele é considerado o maior produtor de cana na região.

## 6. CONSIDERAÇÕES FINAIS

O presente estudo se propôs a utilizar técnicas estatísticas multivariadas para analisar variáveis relativas ao setor sucroenergético brasileiro. Inicialmente, foi realizada uma análise de correlação para avaliar o segmento analisado em relação às variáveis que influenciam a oferta do etanol. O setor, que teve como principais impulsionadores o surgimento do programa Proálcool e a criação dos veículos bicompostíveis, hoje se encontra em um cenário de busca pela recuperação, tendo em vista bons números de crescimento e expectativas de melhorias para os próximos períodos. Relembrando que o etanol hidratado, produto advindo do mesmo, é consumido em quase toda sua totalidade pelo setor de transportes e também é conhecido como combustível ‘verde’, isso porque a sua utilização ajuda na diminuição de emissões de gases de efeito estufa e sua produção respeita quesitos de sustentabilidade, além de gerar renda para as regiões onde se localizam as indústrias.

Utilizando-se a análise de correlação, com o objetivo de avaliar o grau de relacionamento entre os fatores analisados, chegou-se a resultados como uma correlação forte e positiva entre produção de cana-de-açúcar e produção total de etanol ( $r = 0,929$ ), assim como frota total de veículos e frota de veículos bicompostíveis com valor de  $r = 0,999$ , em contrapartida encontrou-se uma forte correlação, porém negativa, entre frota total de veículos e frota de veículos a gasolina ( $r = -0,941$ ) e entre frota de veículos à gasolina e produção total de etanol ( $r = -0,834$ ). As variáveis mais significativas encontradas, por meio do banco de dados utilizado, para explicar o comportamento da produção de etanol, foram produção de cana, frota de veículos à gasolina e bicompostíveis e oferta (prod. Cana) / PIB. Essas variáveis apresentaram p-valores menores que 0,05, o que corroborou para validar estatisticamente que o coeficiente de correlação amostral explica o comportamento do setor estudado. A regressão linear feita confirmou que as variáveis produção de cana-de-açúcar e frota de veículos à gasolina são representativas e que as mesmas são capazes de explicar, respectivamente, 86,32% e 69,52% da produção total de etanol.

Através da análise *cluster* foi possível formar seis grupos compostos pelos principais estados que atuam no setor sucroenergético brasileiro. A forma de definir a composição de cada grupo foi através da análise do nível de similaridade entre as variáveis estudadas. A partir deste estudo chegou-se à conclusão que os estados de São Paulo, Rio de Janeiro e Mato Grosso são considerados três grupos distintos. O primeiro se difere pelos níveis de produção e

investimento no setor, já o segundo se destaca pelas variáveis socioeconômicas e o terceiro por estar mais voltado à produção de açúcar. Mato Grosso do Sul e Goiás constituem um cluster, assim como Minas Gerais e Paraná, já o último grupo é composto principalmente pelos estados do norte e nordeste do país. A partir desta análise é possível pensar em formas de melhorias e desenvolvimento do setor de forma mais simplificada, ou seja, buscar melhorias não apenas para os estados de forma separada, mas sim para os grupos acima formados.

Uma vez que o trabalho alcançou o objetivo proposto, entende-se que a busca pelo desenvolvimento do setor, novas melhorias, ou até mesmo na expansão da tecnologia *flex* para outros tipos de veículos, de maior porte ou que ainda só utilizam combustíveis fósseis, seriam possibilidades a serem avaliadas. O etanol apresenta muitas vantagens de custo/benefício, além de ter um grande foco em questões de sustentabilidade, como a redução de emissão de gases de efeito estufa, preservação do solo e mecanização da colheita que, a longo prazo, irá substituir a queima da palha. Por se tratar de um setor que possui boas expectativas de crescimento e desenvolvimento, é preciso que haja investimentos e incentivos para que o mesmo possa se tornar realidade e trazer ganhos tanto para o país quanto para os consumidores e produtores.

Como proposta para realização de trabalhos futuros, sugere-se ampliar o número de variáveis analisadas neste estudo, aplicar técnicas de regressão múltipla e aprofundar no estudo a respeito de políticas e estratégias específicas voltadas para potencializar cada um dos grupos encontrados por meio da análise cluster, como forma de se minimizar esforços de planejamento e obter melhores resultados globais para o setor sucroenergético nacional.

## REFERÊNCIAS

**Agência de notícias do paraná.** Disponível em:

<<http://www.aen.pr.gov.br/modules/noticias/article.php?storyid=84852>>. Acesso em: Outubro de 2017.

ANP- Agência Nacional do Petróleo, Gás Natural e Biocombustíveis. **Anuários Estatístico Brasileiro do Petróleo, Gás Natural e Biocombustíveis.** Rio de Janeiro: Autores, 2016.

**AutomotiveBusiness.** São Paulo: Redação AutomotiveBusiness, 2016. Disponível em:<

<http://www.automotivebusiness.com.br/noticia/23905/frota-circulante-no-brasil-cresceu-25>>. Acesso em: Janeiro de 2017.

**Balanco Energético Nacional 2016:** Ano base 2015 / Empresa de Pesquisa Energética. – Rio de Janeiro: EPE, 2016.

CAETANO, M.; PALHARES, S. **Perspectivas apontam para virada no setor sucroenergético.** Curitiba: 2016. Disponível em: <<https://www.novacana.com/n/cana/safra/perspectivas-mercado- virada-clima-setor-sucroenergetico-100516/>>. Acesso em: Fevereiro de 2017.

CAMPOS, Natália Lorena. **Políticas de estado no setor sucroenergético.** Revista Geo UERJ. Rio de Janeiro, nº 26, 2015. Disponível em: <<http://www.e-publicacoes.uerj.br/index.php/geouerj/article/view/12696/13404>>. Acesso em: Janeiro de 2017.

CAMPOS, Natália Lorena. **Redes do agronegócio canavieiro:** a territorialização do grupo Tércio Wanderley no Triângulo Mineiro/Alto Paranaíba – MG. Uberlândia: UFU, 2014. Pág. 209. Dissertação (mestrado) - Universidade Federal de Uberlândia, Programa de Pós-Graduação em Geografia, Uberlândia, 2014.

CONAB- Companhia Nacional de Abastecimento. **Acompanhamento da Safra Brasileira de Cana-de-Açúcar.** Brasília: Conab, 2015.

CORREA, Sonia Maria Barros Barbosa. **Probabilidade e Estatística.** 2ª ed. Belo Horizonte: PUC Minas Virtual, 2003.

CUNHA, Renato Augusto Pontes. **Desafios para o abastecimento regional:** Região Nordeste – um piloto para a regionalização do abastecimento. Brasília, 2016. Disponível em:

<<http://www.mme.gov.br/documents/10584/4323103/13-12-2016-+-+Apresenta%C3%A7%C3%A3o+RenovaBio+-+Renato+Cunha.pdf/11e275d5-9ed0-44a9-ac7c-60fd646e88ee>>. Acesso em: Outubro de 2017.

DASSIE, César. **Cana-de-açúcar:** Globo Rural faz balanço da crise do setor. Revista Globo Rural, 2016.

Disponível em: <<http://g1.globo.com/economia/agronegocios/noticia/2016/06/cana-de-acucar-globo-rural-faz-balanco-da-crise-do-setor.html>>. Acesso em: Fevereiro de 2017.

**Empresa de Pesquisa Energética.** Cenários de Oferta de Etanol e Demanda do Ciclo Otto. Rio de Janeiro: Autores, 2016. Disponível em: <[http://www.epe.gov.br/Petroleo/Documents/EPE-DPG-SGB-Bios-NT-02-2016\\_Cen%C3%A1rios%20de%20Oferta%20de%20Etanol%20e%20Demanda%20do%20Ciclo%20Otto.pdf](http://www.epe.gov.br/Petroleo/Documents/EPE-DPG-SGB-Bios-NT-02-2016_Cen%C3%A1rios%20de%20Oferta%20de%20Etanol%20e%20Demanda%20do%20Ciclo%20Otto.pdf)>. Acesso em: Dezembro de 2016.

**Estadão.** Demanda por etanol se iguala à de gasolina A pela 1ª vez desde 2010.Revista Globo Rural, 2015.

Disponível em: <<http://revistagloborural.globo.com/Noticias/Agricultura/Cana/noticia/2015/06/demanda-por-etanol-se-iguala-de-gasolina-pela-1-vez-desde-2010.html>>. Acesso em: Janeiro de 2017.

FIEMS- Federação das Indústrias do Estado de Mato Grosso do Sul. **Setor sucroenergético de MS gerou o maior número de empregos na indústria no semestre.** 2015. Disponível:

<<http://www.fiems.com.br/noticias/setor-sucroenergetico-de-ms-gerou-o-maior-numero-de-empregos-na-industria-no-semester/19086>>. Acesso em: Outubro de 2017.

**Folha de São Paulo.** Usinas de etanol de milho são viáveis em MT, aponta estudo. 2017. Disponível em: <<http://www1.folha.uol.com.br/colunas/vaivem/2017/09/1917776-usinas-de-etanol-de-milho-sao-viaveis-em-mt-aponta-estudo.shtml>>. Acesso em: Outubro de 2017.



FORMIGA, Andrei. **Regressão Linear Simples**. 2013. Disponível em: <<http://andreiformiga.com/post/reglinear/>>. Acesso em: Setembro de 2017

FREI, Fernando. **Introdução à Análise de Agrupamentos**: Teoria e prática. São Paulo: Editora UNESP, 2006. Disponível em: <<https://books.google.com.br/books?id=LXhbj4ccPyMC&printsec=frontcover&hl=pt-BR#v=onepage&q&f=false>>. Acesso em: Setembro de 2017.

GUIMARÃES, Paulo Ricardo B. **Análise de Regressão**. Disponível em: <<https://docs.ufpr.br/~jomarc/regressao.pdf> >. Acesso em: Setembro de 2017.

G1. Açúcar garante melhor remuneração a produtores do oeste de MT. 2014. Disponível em: <<http://g1.globo.com/mato-grosso/agrodebate/noticia/2014/09/acucar-garante-melhor-remuneracao-produtores-do-oeste-de-mt.html>>. Acesso em: Outubro de 2017.

G1. Mato Grosso tem o etanol mais barato do país, aponta pesquisa da ANP. 2015. Disponível em: <<http://g1.globo.com/mato-grosso/noticia/2015/10/mato-grosso-tem-o-etanol-mais-barato-do-pais-aponta-pesquisa-da-anp.html>>. Acesso em: Outubro de 2017.

HAIR, Jr., Joseph F.; ANDERSON, Rolph E.; TATHAM, Ronald L.; BLACK, William C. **Análise Multivariada de Dados**. 5ª ed. Porto Alegre: Bookman, 2005.

IBGE - Instituto Brasileiro de Geografia e Estatística. 2017. Disponível em: <<https://cidades.ibge.gov.br/>>. Acesso: Junho de 2017.

IGNÁCIO, Sérgio Aparecido. **Importância da Estatística para o processo de conhecimento e tomada de decisão**. Revista Paranaense de Desenvolvimento. Curitiba: n. 118, p. 175-192, 2010.

KONISHI, Fabio, SOARES, Paula Meyer, GALVÃO, Reny Aparecida, SILVA, Marcelo Santana, ROCHA, Angela Machado. **Uma década da tecnologia bicombustível**: análise do segmento automobilístico e a sua correlação com o setor sucroalcooleiro. Revista Energia na Agricultura, vol. 29, nº 4, 2014. Disponível em: <[http://energia.fca.unesp.br/index.php/energia/article/view/999/pdf\\_5](http://energia.fca.unesp.br/index.php/energia/article/view/999/pdf_5)>. Acesso em: Janeiro de 2017.

LARSON, Ron, FARBER, Betsy. **Estatística Aplicada**. São Paulo: Pearson Prentice Hall, 2010.

MACHADO, Gleysson B. Fabricação de Açúcar e Etanol a partir da Cana-de-açúcar. [Publicação online], 2014. Disponível em: <<http://www.portaldobiogas.com/fabricacao-de-acucar-e-etanol-partir-da-cana-de-acucar/>>. Acesso em: Outubro de 2016.

MANNARELLI FILHO, Teucle. **Análise da Correlação e Regressão da Expansão Açucareira da Região Oeste de São Paulo**. Revista Economia & Pesquisa, Fundação Educacional de Araçatuba, Volume 7, ISSN Impresso: 1516-7666, 2005.

MELO, André de Souza, SAMPAIO, Yony de Sá Barreto. **Impactos dos preços da gasolina e do etanol sobre a demanda de etanol no Brasil**. Revista de economia contemporânea, vol.18, nº 1, 2014. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S1415-98482014000100056](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1415-98482014000100056)>. Acesso em: Janeiro de 2017.

MINGOTI, Sueli Aparecida. **Análise de dados através de métodos de estatística multivariada**: uma abordagem aplicada. Belo Horizonte: Editora UFMG, 2005.

**Minitab**. Tipos de Análise de Regressão. 2017. Disponível em: <<https://support.minitab.com/pt-br/minitab/18/help-and-how-to/modeling-statistics/regression/supporting-topics/basics/types-of-regression-analyses/>>. Acesso em: Setembro de 2017.

NASTARI, Plínio. M. (2012). **A importância do setor sucroenergético no Brasil**. Revista Agroanalysis. Disponível em: <<http://www.agroanalysis.com.br/3/2012/mercado-negocios/producao-sustentavel-a-importancia-do-setor-sucroenergetico-no-brasil>>. Acesso em: Janeiro de 2017.

**NovaCana.** Preservação do solo a partir do plantio de cana. Curitiba: NovaCana. Disponível em: <<https://www.novacana.com/sustentabilidade/preservacao-solos-plantio-cana/>>. Acesso em: Fevereiro de 2017.

**NovaCana.** Distribuição das usinas de etanol no Brasil. Curitiba: NovaCana. Disponível em: <<https://www.novacana.com/usina/distribuicao-usinas-etanol-brasil/>>. Acesso em: Setembro de 2017.

**Portal Action.** Cálculo e interpretação do p-valor. 2017. Disponível em: <<http://www.portalaction.com.br/inferencia/512-calculo-e-interpretacao-do-p-valor>>. Acesso em: Setembro de 2017.

RIBEIRO, José Luis Duarte, CATEN, ten Carla. **Regressão Linear Simples.** Disponível em: <<http://wiki.nosdigitais.teia.org.br/images/c/c8/RLS.pdf>>. Acesso em: Setembro de 2017.

RODRIGUES, Délcio, ORTIZ, Lúcia. **Em direção à sustentabilidade da produção de etanol de cana de açúcar no Brasil.** 2006. Disponível em: <<http://riosvivos.org.br/arquivos/444052181.pdf>>. Acesso em: Fevereiro de 2017.

RODRIGUES, José Augusto R. **Do engenho à biorrefinaria. A usina de açúcar como empreendimento industrial para a geração de produtos bioquímicos e biocombustíveis.** Revista Química Nova, vol. 34, nº 7, 2011. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0100-40422011000700024](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-40422011000700024)>. Acesso em: Março de 2017.

TRIOLA, Mario F. **Introdução à Estatística.** Rio de Janeiro: Livros Técnicos Científicos, 1999.

TURRIONI, João Batista, MELLO, Carlos Henrique Pereira. **Metodologia de Pesquisa em Engenharia de Produção.** Universidade Federal de Itajubá – UNIFEI, Programa de Pós-graduação em Engenharia de Produção – Curso de Especialização em Qualidade e Produtividade, 2012. Disponível em: <[http://www.marco.eng.br/adm-organizacao-I/Apostila\\_Metodologia\\_Completa\\_2012\\_%20UNIFEI.pdf](http://www.marco.eng.br/adm-organizacao-I/Apostila_Metodologia_Completa_2012_%20UNIFEI.pdf)>. Acesso em: Outubro de 2017.

UNICA - União da Indústria de Cana-de-Açúcar. **A Sustentabilidade no Setor Sucroenergético Brasileiro.** São Paulo: Unica. Disponível em: <<http://www.unica.com.br/sustentabilidade/>>. Acesso em: Novembro de 2016.

UNICA - União da Indústria de Cana-de-Açúcar. **Carro flex representa melhor custo-benefício para brasileiro.** São Paulo: Unica, 2017. Disponível em: <<http://www.unica.com.br/noticia/9851879920320450396/carro-flex-representa-o-melhor-custo-beneficio-para-brasileiro/>>. Acesso em: Fevereiro de 2017.

UNICA - União da Indústria de Cana-de-Açúcar. **Competitivo, etanol cresce em vendas no estado de Minas Gerais.** São Paulo: Unica, 2013. Disponível em: <<http://www.unica.com.br/noticia/40728229920323122056/competitivo-por-cento2C-etanol-cresce-em-vendas-no-estado-de-minas-gerais/>>. Acesso em: Outubro de 2017.

UNICA - União da Indústria de Cana-de-Açúcar. **Consumo de etanol hidratado cresceu em mais de 37% em 2015.** São Paulo: Unica, 2016. Disponível em: <<http://www.unica.com.br/noticia/14883647920325965467/consumo-de-etanol-hidratado-cresceu-mais-de-37-por-cento-em-2015/>>. Acesso em: Novembro de 2016.

UNICA - União da Indústria de Cana-de-Açúcar. **Histórico de Produção e Moagem.** São Paulo: Unica. Disponível em: <<http://www.unicadata.com.br/historico-de-producao-e-moagem.php?idMn=31&tipoHistorico=2>>. Acesso em: Janeiro de 2017.

UNICA - União da Indústria de Cana-de-Açúcar. **Lideranças do setor recebem comenda em Goiás.** São Paulo: Unica, 2017. Disponível em: <<http://www.unica.com.br/noticia/25292487920311735640/liderancas-do-setor-recebem-comenda-em-goias/>>. Acesso em: Outubro 2017.

UNICA - União da Indústria de Cana-de-Açúcar. **Setor sucroenergético do estado de São Paulo frente à crise econômica.** São Paulo: Unica, 2016. Disponível em: <<http://www.unica.com.br/convidados/29063662920344564516/setor-sucroenergetico-do-estado-de-sao-paulo-frente-a-crise-economica/>>. Acesso em: Outubro de 2017.

UNICA - União da Indústria de Cana-de-Açúcar. **Tocantins tem maior produção de cana-de-açúcar da Região Norte.** São Paulo: Única, 2013. Disponível em: < <https://www.novacana.com/n/cana/safra/tocantins-maior-producao-cana-acucar-norte-040113/>>. Acesso em: Outubro de 2017.