

MODELO ESTATÍSTICO DE ESTIMATIVA DO CONSUMO DE ELETRICIDADE PELO SETOR AGROPECUÁRIO PAULISTA

 <https://doi.org/10.56238/sevened2024.037-211>

Monclar Nogueira Christovão

Mestre em Agronegócio e Desenvolvimento pela Universidade Estadual Paulista (UNESP)

Instituto Federal de São Paulo (IFSP), Tupã

E-mail: monclar.christovao@ifsp.edu.br

Orcid: <https://orcid.org/0000-0001-5509-0574>

Lattes: <http://lattes.cnpq.br/2038445224049133>

Mario Mollo Neto

Doutor em Engenharia Agrícola pela Universidade Estadual de Campinas

Universidade Estadual Paulista, Faculdade de Ciências e Engenharia, Tupã

E-mail: mario.mollo@unesp.br

Orcid: <https://orcid.org/0000-0002-8341-4190>

Lattes: <http://lattes.cnpq.br/6037463340047597>

RESUMO

A energia elétrica é um insumo fundamental e essencial para o desenvolvimento socioeconômico de uma região. Até o ano de 2016, o principal recurso energético utilizado no setor agropecuário de São Paulo era o óleo diesel. No entanto, nos últimos seis anos (2017 a 2022), a eletricidade se tornou a fonte de energia mais utilizada por esse setor. Nesse contexto, é vital fazer uma projeção futura da demanda de carga a fim de embasar as deliberações no planejamento e na expansão do sistema de fornecimento elétrico do estado. Os objetivos deste estudo consistem em criar um modelo estatístico capaz de prever o consumo de energia elétrica no setor rural paulista, assim como reconhecer as variáveis que influenciam na projeção desse consumo. Com base em informações quantitativas levantadas no site da Secretaria de Meio Ambiente, Infraestrutura e Logística do estado paulista, empregou-se o modelo de regressão linear múltipla. Para a construção do modelo foram selecionadas seis variáveis independentes, sendo três provenientes do agronegócio de São Paulo: número de consumidores, tarifa de energia elétrica e consumo de óleo diesel; duas relacionadas ao estado: potências instaladas das usinas térmicas de biomassa e das usinas fotovoltaicas e uma do agronegócio nacional: PIB. Devido à falta de significância estatística, duas variáveis foram rejeitadas: potência das usinas térmicas de biomassa e o PIB. Portanto, com o modelo matemático proposto, construído com quatro variáveis significativas (com diferentes níveis de influência), é possível fazer previsões futuras do consumo de energia do setor pesquisado. A potência instalada das usinas fotovoltaicas teve um impacto significativo, enquanto a quantidade de consumidores, a tarifa de energia e o consumo de óleo diesel demonstraram uma forte influência na previsão do consumo de energia do setor. O destaque ficou com a confirmação, pelo modelo estatístico ajustado, da transição do consumo de óleo diesel pela energia elétrica como a principal fonte de energia utilizada pelo setor agropecuário paulista.

Palavras-chave: Modelo Matemático. Previsão. Tendência. Setor Rural. Regressão Linear.

1 INTRODUÇÃO

No Brasil, a presença de petróleo e seus derivados na oferta interna de energia cresceu de 34% em 1970 para 46% em 2000 e estima-se que diminua para 31% em 2030 (BRASIL-MME, 2007).

No decênio de 1970, a lenha era o energético dominante, representando cerca de 46% do consumo total, mas cedeu rapidamente espaço para o uso dos derivados de petróleo, que, no citado período, alcançou uma participação de 37,85%, ao mesmo tempo que a hidroeletricidade atingiu a marca de 5,5% (BRASIL-MME, 2007).

Segundo Brasil-MME (2007), entre as décadas de 80 e 2000, a presença do petróleo e seus derivados se estabeleceu na faixa entre 40 e 50%, enquanto a lenha teve uma redução para 8% e a hidroeletricidade se consolidou em torno de 16%.

A eletricidade concorre para o progresso e a prosperidade de uma sociedade. A eletricidade é um insumo básico e fundamental e, como tal, exclui socialmente a parcela da população que tem restrições ao seu acesso (BISOGNIN; WERNER, 2020).

Segundo consta em Brasil-EPE (2024), a potência elétrica total instalada no Brasil em 2023 atingiu a marca de 226 Giga watt (GW). Alocando esse montante entre as quatro principais fontes geradoras do país, tem-se que as usinas hidroelétricas (UHE) foram responsáveis por 48,6% dessa potência, seguido pelas usinas termoeletricas (UTE) com 20,0%, as usinas solares com 16,7% e as usinas eólicas com 12,7%. O destaque ficou por conta do crescimento próximo a 55% da capacidade das plantas fotovoltaicas, em comparação com 2022, com uma potência instalada de 37.843 GW.

Quanto à produção de energia elétrica no âmbito nacional, contrapondo os anos de 2022 e 2023, de acordo com Brasil-EPE (2024), houve dois destaques. O primeiro foi o aumento de 68,1% da geração fotovoltaica, de 30.126 GWh para 50.633 GWh e o segundo foi a redução de 19,3% da geração das UTEs, que utilizam derivados de petróleo, de 7.485 GWh para 6.041 GWh.

Durante a campanha à presidência do Brasil, o governo então eleito para o quadriênio 2023-2026 disponibilizou um documento chamado “Carta para o Brasil do Amanhã” contendo propostas e compromissos nos quais está a retomada do programa “Luz para Todos” (MELLO, 2022). O objetivo desse programa é disponibilizar o acesso ao serviço público de distribuição de eletricidade aos moradores de áreas rurais distantes e carentes.

O consumo total de energia elétrica por todas as áreas no estado de São Paulo em 2021, compreendendo também os autoprodutores, foi de 151.729 Giga watt-hora (GWh), uma elevação de 4,32% em comparação com o ano anterior, que foi igual a 145.451 GWh. As áreas industrial (9,2%) e comercial (3,8%) evidenciaram os maiores aumentos na utilização de eletricidade para o período. Os setores residencial (0,5%) e agropecuário (0,2%) ostentaram baixíssimo crescimento, quase uma estabilidade, no consumo dessa fonte de energia (SÃO PAULO-SIMA, 2022).

Conforme relatam Bisognin e Werner (2020), previsões acertadas de demanda e de consumo de energia elétrica são primordiais para que corretas decisões sejam tomadas na alocação de recursos, na construção e no aprimoramento da infraestrutura elétrica de uma região. A necessidade de se planejar com antecedência resulta do fato de que a eletricidade é um insumo não passível de estoque.

Para fins de estimativas e previsões, uma das técnicas mais utilizadas na área de gestão e em universidades é a análise de regressão, cujo ponto principal é a presença da relação estatística entre uma variável intitulada dependente, ou explicada, ou prevista, com uma ou mais variáveis designadas independentes, ou preditoras, ou explanatórias (CUNHA; COELHO, 2014).

Segundo Cunha e Coelho (2014), o propósito da análise de regressão é prever os montantes da variável dependente com base nos montantes conhecidos das variáveis independentes.

Ainda de acordo com Cunha e Coelho (2014), entende-se a regressão como sendo a formação de uma relação funcional entre variáveis envolvidas na explicação de um fenômeno.

Quanto à classificação dos tipos de energia, segundo Brasil-MME (2024), as energias primárias e secundárias são categorias dentro da classificação geral das fontes de energia.

A energia primária é o resultado energético de recursos que são extraídos da natureza em sua forma original, como: lenha, petróleo, gás natural, carvão mineral e outras (BRASIL-MME, 2024). Alguns resíduos da agroindústria, como borra de café e palha de arroz, são fontes primárias usadas na geração de vapor (SÃO PAULO-SEMIL, 2024).

A energia secundária é decorrente de produtos industrializados que passam por modificações em uma ou mais transformações. São fontes energéticas secundárias o óleo diesel, óleo combustível, gás liquefeito de petróleo (GLP), querosene, eletricidade e outras (BRASIL-MME, 2024).

Pelo acima exposto, a pesquisa de previsão estatística do consumo de eletricidade do setor agropecuário paulista é justificada, visto que, além de ser uma ferramenta de diagnóstico, também é capaz de gerar um modelo matemático de estimativa do consumo de energia desse setor, contribuindo com informações regionais para atender às necessidades de melhorias e ampliação de toda a infraestrutura de transmissão e distribuição de eletricidade em virtude da tendência de crescimento no uso desse tipo de energia. Também contribui para a criação de políticas públicas para o setor do agronegócio em prol do bem-estar social da população rural e do desenvolvimento da economia.

O objetivo geral desta pesquisa foi gerar um modelo estatístico que estime o consumo de eletricidade do setor agropecuário paulista. O objetivo específico foi identificar as variáveis que contribuem para a previsão desse consumo.

2 METODOLOGIA

A ferramenta utilizada neste trabalho é a análise de regressão que, segundo Martins e Domingues (2014), é utilizada, principalmente, com a finalidade de previsão. O objetivo é a construção

de um modelo estatístico para a previsão de respostas de uma variável dependente (Y) na dependência do valor de uma variável (X) ou mais variáveis (X_1 , X_2 , ... e X_n) independente(s), estabelecendo uma relação matemática entre esses dois tipos de variáveis (MARTINS; DOMINGUES, 2014).

A variável resposta (dependente) deve ser quantitativa, enquanto a(s) variável(eis) preditor(a)s podem ser quantitativas ou qualitativas.

Na pesquisa, a variável dependente é a previsão anual do consumo de eletricidade pelo ramo agropecuário paulista, fundamentada em um modelo estatístico.

O modelo de regressão é chamado simples, ou seja, quando há somente uma variável independente (explicativa). Na existência de duas ou mais dessas variáveis, o modelo de regressão é denominado múltiplo (MARTINS; DOMINGUES, 2014).

2.1 CONSTRUINDO UM MODELO DE REGRESSÃO LINEAR SIMPLES

A relação entre os dois tipos de variáveis pode tomar diversas formas, dentre elas, um modelo simplificado, que é a relação linear simples contendo apenas uma variável independente, como a equação da reta, conforme a Equação (1).

$$Y = \beta_0 + \beta_1 \cdot X + \varepsilon \quad (1)$$

onde:

- β_0 : coeficiente linear (intercepto da reta);
- β_1 : coeficiente angular ou coeficiente de regressão (inclinação da reta);
- ε : erro da regressão ou resíduo.

Com base em dados amostrais, tem-se $\hat{Y}$, que é o resultado da previsão Y para uma observação X, de acordo com o modelo ajustado da regressão linear simples enunciado pela Equação (2).

$$\hat{Y} = b_0 + b_1 \cdot X \quad (2)$$

onde:

- b_0 : o estimador de β_0 ; e
- b_1 : o estimador de β_1 .

Os valores de b_0 e de b_1 serão coletados na coluna “Coeficientes” da ferramenta de Análise de Dados do editor de planilhas Microsoft Excel® a partir de amostras de pares ordenados de duas variáveis.

2.2 CONSTRUINDO UM MODELO DE REGRESSÃO LINEAR MÚLTIPLA

A relação entre os dois tipos de variáveis pode tomar diversas formas, dentre elas, um modelo múltiplo que tem por objetivo melhor prever e explicar o comportamento da variável dependente pesquisada (Y). Deste modo, pode-se incorporar outras variáveis explicativas ($X_1, X_2, \dots, X_n$) ao modelo, de acordo com a Equação (3).

$$Y = \beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \dots + \beta_n \cdot X_n + \varepsilon \quad (3)$$

onde:

- β_0 : coeficiente linear (intercepto/interseção da reta);
- $\beta_1, 2, \dots, n$: coeficiente angular ou coeficiente de regressão (inclinação da reta);
- ε : erro experimental de Y (pontos fora da reta).

Cunha e Coelho (2014) esclarecem que o coeficiente β_0 simboliza o valor do cruzamento (interseção) do gráfico de regressão com o eixo das ordenadas; portanto, β_0 retrata o valor de Y quando X for nulo.

A análise de regressão múltipla é empregada para estudar a relação entre uma variável dependente única, escolhida pelo pesquisador, com inúmeras variáveis independentes de valores conhecidos (GIL, 2021).

Com base em dados amostrais, tem-se $\hat{Y}$, que é o resultado da previsão Y para observações X, de acordo com o modelo ajustado fornecido pela Equação (4).

$$\hat{Y} = b_0 + b_1 \cdot X_1 + b_2 \cdot X_2 + \dots + b_n \cdot X_n \quad (4)$$

onde:

- b_0 : estimador de β_0 ;
- $b_1, 2, \dots, n$: estimadores respectivos de $\beta_1, 2, \dots, n$.

Utilizando a ferramenta designada “Regressão”, disponível em “Análise de Dados” do editor de planilhas Microsoft Excel®, a partir de amostras de pares ordenados de três ou mais variáveis, serão coletados os valores de $b_0, b_1, b_2, \dots, b_n$ na coluna “Coeficientes”.

2.3 TESTANDO A SIGNIFICÂNCIA DOS COEFICIENTES DA REGRESSÃO

Na planilha denominada “ANOVA”, criada pela ferramenta “Regressão” da Análise de Dados do Microsoft Excel®, estão presentes os montantes individuais de “valor-P” dos coeficientes (parâmetros).

Testa-se a importância de uma ou mais variáveis independentes (individualmente). A informação contida nos “valor-P” individuais permite determinar se cada variável explicativa possui relevância estatística ou não.

Hipóteses:

- $H_0: \beta_i = 0$, consiste em verificar se cada parâmetro β_i da regressão é igual a zero;
- $H_1: \beta_i \neq 0$, se $\text{valor-P} \leq \alpha$, rejeita-se a hipótese H_0 , concluindo-se que $\beta_i \neq 0$ para um risco (nível de significância) α .

2.4 TESTANDO A EXISTÊNCIA DA REGRESSÃO LINEAR

Também na planilha “ANOVA”, descrita no item anterior, está contido o valor de “F de significação”. Esse índice aponta se o modelo de regressão linear como um todo é estatisticamente relevante ou não. Ele indica se a combinação de variáveis explicativas tem uma associação estatisticamente significativa com a variável resposta.

Para a condição de nível descritivo ou probabilidade de significância (p-valor: F de significação) $\leq \alpha$ (nível de significância) adotado na pesquisa, pode-se concluir que existe regressão, ou seja, o modelo pode explicar e prever Y (variável resposta).

2.5 FERRAMENTAS PARA A ANÁLISE DOS RESULTADOS

2.5.1 Qualidade do Ajuste

2.5.1.1 Coeficiente de Correlação Linear

Aplicando-se a ferramenta “Regressão” contida na Análise de Dados do Excel®, a partir de amostras de pares ordenados de duas ou mais variáveis, obtém-se o chamado Coeficiente de Correlação Linear (R múltiplo) presente na tabela denominada “Estatística de Regressão”. Esse coeficiente simboliza o nível de ligação (associação) entre as variáveis autônomas e a dependente (CUNHA; COELHO, 2014).

2.5.1.2 Coeficiente de Determinação ou de Explicação

Na tabela “Estatística de Regressão”, encontra-se também o chamado Coeficiente de Determinação ou Explicação (R^2).

O R^2 é uma medida da proporção da variação em Y (resposta), que é explicada por X (explicativa) pelo ajuste do modelo linear, sendo, assim, uma medida descritiva da qualidade do ajuste. O valor desse coeficiente situa-se entre $0 < R^2 < 1$, sendo que quanto mais próximo do valor unitário, melhor será a qualidade do ajuste desse modelo linear.

2.5.1.3 Coeficiente de Determinação Ajustado

O Coeficiente de Determinação Ajustado (R^2 ajustado) está relacionado na tabela “Estatística de Regressão”, sendo utilizado na comparação entre equações de regressão que contêm quantidades distintas de variáveis independentes ou de dimensões amostrais (CUNHA; COELHO, 2014).

A equação de regressão que melhor explica a variável dependente estudada é aquela cujo montante do coeficiente R^2 ajustado é o mais próximo do valor unitário.

2.5.1.4 Erro Padrão

O “Erro Padrão” é o valor médio, na unidade da variável dependente, que a equação ajustada está errando para mais ou para menos; em outras palavras, é a distância média entre os montantes observados e a reta da regressão linear.

Também presente na tabela citada anteriormente, o “Erro Padrão” é um indicador de precisão da regressão, sendo que quanto menor for o seu valor, melhor será a estimativa do modelo (CUNHA; COELHO, 2014).

2.5.2 Comparando Equações de Regressão

Equações de regressão podem ser comparadas, utilizando-se para tal o Coeficiente de Determinação Ajustado (R^2 ajustado) de ambas, sendo que a regressão que melhor explica a variável dependente é aquela cujo montante do coeficiente R^2 ajustado é o mais próximo do valor unitário.

Outro fator de comparação, entre regressões que explicam a mesma variável dependente, é o “Erro Padrão” que, quanto menor for seu valor, mais precisa será a estimativa do modelo.

2.5.3 Análise de Resíduos (ϵ)

Sugere-se a verificação de alguns atributos da distribuição dos resíduos, tais como:

- Averiguar se é nula a média da distribuição de probabilidade da variável ϵ ;
- Testar a homoscedasticidade. Em análise de variância (ANOVA), os erros devem ter variância comum;
- Verificar a normalidade da distribuição de probabilidade da variável ϵ empregando o Teste de Shapiro-Wilk.

Hipóteses:

- H_0 : a variável em estudo é normalmente distribuída;
- H_1 : a variável em estudo não é normalmente distribuída.

Se o p-valor $\leq \alpha$, rejeita-se a hipótese H_0 para um risco (nível de significância) α .

Se o $p\text{-valor} > \alpha$, não se rejeita a hipótese H_0 para um risco (nível de significância) α .

- d) Ausência de autocorrelação serial. De acordo com Cunha e Coelho (2014), o modelo deve presumir que a correlação entre os resíduos por toda a extensão do espectro das variáveis independentes é nula; isso significa que o impacto de uma observação em uma determinada variável X não afeta as observações subsequentes. Portanto, não há relação causal entre os resíduos e a variável X e, conseqüentemente, a variável Y é afetada apenas pela própria variável X em questão, e não por efeitos defasados de X_1 sobre X_2 e, conseqüentemente, sobre Y .

2.5.4 Multicolinearidade Entre As Variáveis Independentes

A multicolinearidade acontece quando duas ou mais variáveis independentes, presentes no modelo e que tentam explicar o mesmo fenômeno, possuem informações semelhantes. Por consequência, a alta correlação entre duas ou mais variáveis independentes pode dificultar a distinção dos seus efeitos individuais sobre a variável dependente, provocando uma sobreposição de informações na explicação e estimação, o que pode levar à perda de significância de uma delas na elucidação do comportamento do fenômeno (CUNHA; COELHO, 2014).

Conforme Cunha e Coelho (2014), a multicolinearidade tem o efeito de distorcer os coeficientes angulares calculados das variáveis afetadas, prejudicando a precisão preditiva do modelo e o entendimento do verdadeiro impacto da variável independente no comportamento da variável dependente.

As implicações do acima exposto são presumíveis e listadas a seguir:

- Aumento nos erros padrão;
- Redução da eficiência dos estimadores;
- Estimativas menos precisas;
- Maior sensibilidade a pequenas oscilações nos dados.

Essas inferências descritas anteriormente dificultam a separação dos efeitos de cada variável.

A multicolinearidade entre as variáveis independentes será investigada, empregando-se o Teste de Correlação de Pearson com o intuito de não utilizar variáveis redundantes.

Hipóteses:

- $H_0: \rho = 0$ (não existe correlação linear entre as variáveis em estudo);
- $H_1: \rho \neq 0$ (existe correlação linear entre as variáveis em estudo).

Se o $p\text{-valor} \leq \alpha$, rejeita-se a hipótese H_0 para um risco (nível de significância) α .

Se o $p\text{-valor} > \alpha$, não se rejeita a hipótese H_0 para um risco (nível de significância) α .

Na sequência, são apresentadas a aplicação do método proposto, bem como a exploração e discussão dos resultados.

3 RESULTADOS E DISCUSSÃO

A coleta dos dados procedeu-se, principalmente, do relatório digital anual intitulado Balanço Energético do Estado de São Paulo (BEESP) disponibilizado pela Secretaria de Meio Ambiente, Infraestrutura e Logística (SEMIL) do governo paulista, em que o setor agropecuário está incluído na classe denominada “demais” (juntamente com os serviços públicos, poderes públicos, iluminação pública e consumo próprio) nas tabelas do referido documento.

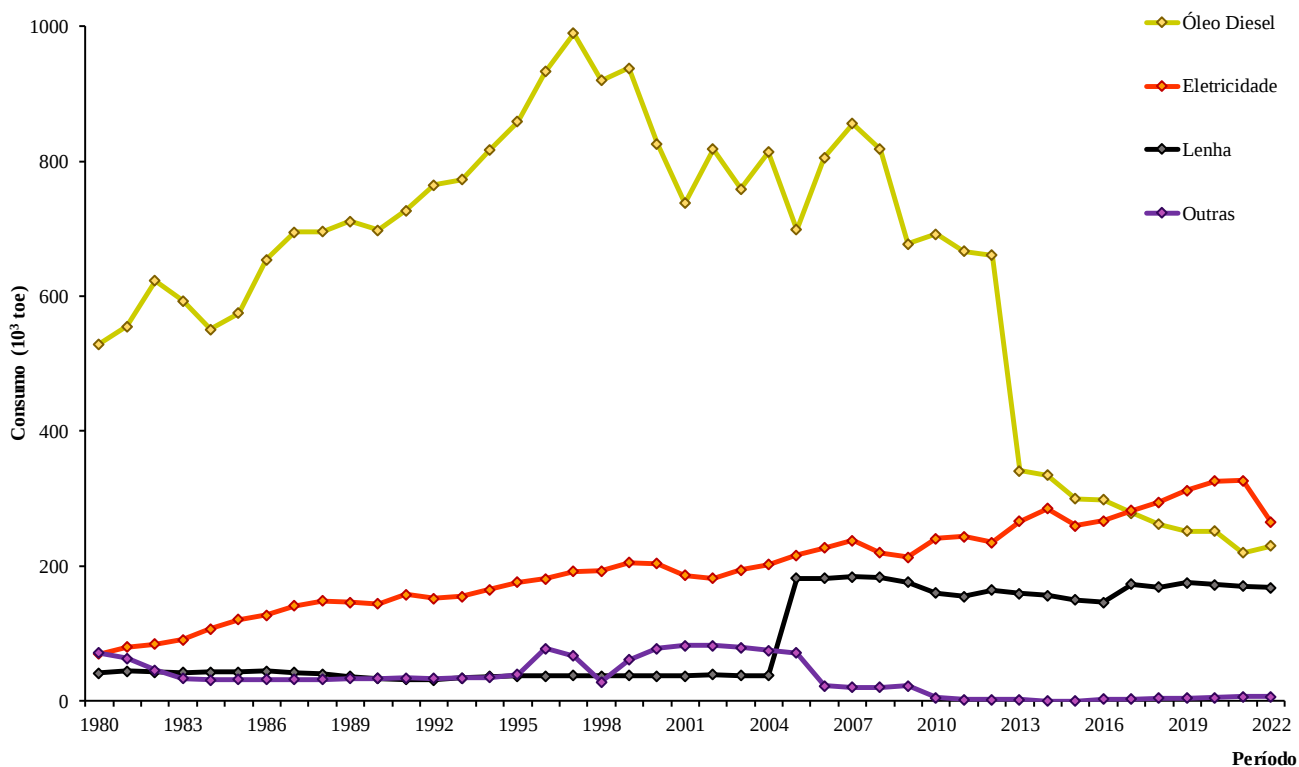
Nesta pesquisa, a abordagem é quantitativa pela natureza de todas as variáveis envolvidas. E estas variáveis – variável dependente Y e variáveis independentes X1 a X6 – estão relacionadas e detalhadas a seguir.

- A variável dependente Y – consumo de eletricidade, em Megawatt-hora (MWh), do agronegócio paulista – teve sua série histórica (2000 a 2023) coletada do *site* da SEMIL (<https://dadosenergeticos.energia.sp.gov.br/portalsev2/intranet/Eletricidade/index.html>) em “Dados de Energia/Sistema de Distribuição/Consumo Anual”.
- Variável independente X1 (quantidade de consumidores do agronegócio paulista). As informações foram reunidas do mesmo *site* da SEMIL citado anteriormente (<https://dadosenergeticos.energia.sp.gov.br/portalsev2/intranet/Eletricidade/index.html>) em “Dados de Energia/Sistema de Distribuição/Consumo Anual”.
- Variável independente X2 – tarifa de eletricidade, em reais por Megawatt-hora (R\$/MWh), do setor rural paulista, com os valores das tarifas convencionais de energia elétrica [Tarifa de Uso do Sistema de Distribuição (TUSD) + Tarifa de Energia (TE)], entre os anos de 2000 e 2009, foi determinada pela média aritmética, ano a ano, de três subgrupos, que são: B2-Rural, B2-Cooperativa de Eletrificação Rural e B2-Serviço Público de Irrigação, contidos no “Histórico anterior a 2010/Histórico das Resoluções Homologatórias”, publicadas pela Agência Nacional de Energia Elétrica (ANEEL) no *site* (<https://portalrelatorios.aneel.gov.br/luznatarifa/basestarifas>) das seguintes distribuidoras de energia: Companhia Paulista de Força e Luz (CPFL Paulista), Companhia Piratininga de Força e Luz (CPFL Piratininga), Companhia Luz e Força Santa Cruz (CLFSC) e Elektro – Eletricidade e Serviços S/A. Quanto aos montantes das tarifas de energia (TUSD + TE), a partir de 2010, no *link*: <https://portalrelatorios.aneel.gov.br/luznatarifa/basestarifas> em “TUSD e TE/Dados das Tarifas Homologadas”, também foram calculados pela média aritmética, ano a ano, dos três subgrupos B2 citados anteriormente das distribuidoras

subsequentes: CPFL Paulista, CPFL Piratininga, CPFL Santa Cruz, Energias de Portugal (EDP) São Paulo, Enel Distribuição São Paulo, Energisa Sul Sudeste (ESS) e Neoenergia Elektro.

- Variável independente X3 (consumo de óleo diesel). Do BEESP 2019 (ano base 2018), associado ao BEESP 2023 (ano base 2022), foram obtidas informações das fontes energéticas consumidas pelo citado setor, a partir das quais foi plotado o Gráfico 1 com a finalidade de avaliar a evolução do consumo de cada fonte de energia e sua dependência pelo setor. O destaque ficou por conta da eletricidade, óleo diesel e lenha. As demais fontes de energia consumidas (denominadas no gráfico como “outras”) é composta pela soma de três energéticos: GLP, querosene e óleo combustível. No Gráfico 1, é possível observar a diminuição do consumo de óleo diesel a partir de 2007, sendo este, até 2016, o energético mais usado pelo setor rural paulista. Em oposição, nota-se a crescente utilização da eletricidade, sendo o energético mais consumido nos últimos seis anos.

Gráfico 1 - Evolução do consumo de energéticos, em toe¹, pelo setor agropecuário paulista



Fonte: Elaborado pelo autor a partir de São Paulo-SIMA (2019) e São Paulo-SEMIL (2023).

Em função do acima exposto, foi atribuída a variável independente X3, o consumo de óleo diesel, em mil (10³) metros cúbicos (m³), do agronegócio paulista, sendo os seus volumes obtidos em duas etapas. A primeira, entre os anos de 2000 e 2012, do BEESP 2019 (ano base 2018) disponível no [link: https://smastr16.blob.core.windows.net/2001/2023/12/BEESP2019ab2018.pdf](https://smastr16.blob.core.windows.net/2001/2023/12/BEESP2019ab2018.pdf). A segunda,

¹ De acordo com São Paulo-SEMIL (2023), a unidade de energia utilizada em alguns relatórios de balanços energéticos é a tonelada de óleo equivalente (toe), sendo que um toe equivale a 10x10⁹ calorias ou a 10 Giga calorias (10 Gcal).

de 2013 a 2022, foi coletada do BEESP 2023: Ano Base 2022 (<https://smastr16.blob.core.windows.net/2001/2024/02/BEESP2023ab2022-2a-edicao-2.pdf>).

- Variável independente X4 [UTE Biomassa², em quilowatt (kW), do estado de São Paulo]. É o somatório das potências elétricas instaladas das usinas termoeletricas geradoras paulistas que utilizam os seguintes compostos energéticos: bagaço de cana-de-açúcar; biogás; biogás-AGR³; biogás-RU⁴; carvão-RU; casca de arroz; lenha; licor negro (lixívia⁵); resíduos de madeira e resíduos florestais. Essas informações foram colhidas dos arquivos digitais “Resumo Executivo: Dados de produção de energia a partir de fontes renováveis” dos anos de 2013 a 2023 (https://dadosenergeticos.energia.sp.gov.br/portalcerv2/intranet/BiblioVirtual/renovaveis/resumo_executivoRE.pdf).
- Variável independente X5 (Geração Fotovoltaica, em kW, do estado de São Paulo): utilizou-se a mesma fonte citada na variável anterior (X4) para reunir as informações do somatório das potências elétricas instaladas das usinas solares paulistas.
- Variável independente X6 [Produto Interno Bruto (PIB) do agronegócio brasileiro]: as informações do PIB do Agronegócio do Estado de São Paulo ainda não estão disponíveis no site; por essa razão, foram coletados os valores do PIB do Agronegócio Brasileiro no link: [https://view.officeapps.live.com/op/view.aspx?src=https%3A%2F%2Fcepea.esalq.usp.br%2Fupload%2Fkceditor%2Ffiles%2FPlanilha_PIB_Cepea_Portugues_Site%2520\(1\)\(4\).xlsx&wdOrigin=BROWSELINK](https://view.officeapps.live.com/op/view.aspx?src=https%3A%2F%2Fcepea.esalq.usp.br%2Fupload%2Fkceditor%2Ffiles%2FPlanilha_PIB_Cepea_Portugues_Site%2520(1)(4).xlsx&wdOrigin=BROWSELINK) na aba “PIB” da planilha do Centro de Estudos Avançados em Economia Aplicada (CEPEA). O resultado desse PIB é a soma, em milhões de reais correntes, de quatro fatores, que são:
 - Insumos utilizados na agropecuária;
 - Produção básica da agropecuária;
 - Produção agroindustrial;
 - Serviços agropecuários.

Os dados coletados produziram séries históricas para o setor alvo do estudo, entre 2000 e 2022. Os montantes referentes aos gastos com recursos energéticos dos anos de 2020 e 2021 podem ter sido afetados pela pandemia da COVID-19.

A variável dependente Y (consumo de eletricidade), bem como as variáveis independentes X1 (quantidade de consumidores), X2 (tarifa de energia elétrica) e X3 (consumo de óleo diesel) são pertinentes ao agronegócio paulista, enquanto as variáveis independentes X4 (geração de energia

² Unidades geradoras que empregam fontes renováveis para a produção de eletricidade.

³ AGR: resíduos agrícolas.

⁴ RU: resíduos sólidos urbanos.

⁵ Resíduo líquido oriundo do setor de papel & celulose.

elétrica pelas usinas termoeletricas que usam biomassa) e X5 (geração de energia elétrica pelas usinas fotovoltaicas) são referentes ao estado de São Paulo; por fim, a variável independente X6 (produto interno bruto) é alusiva ao agronegócio nacional.

3.1 TESTE DA SIGNIFICÂNCIA DOS COEFICIENTES

Em “Dados” (guia da faixa de opções) no interior do editor de planilhas Microsoft Excel®, escolhe-se o suplemento “Análise de Dados” e, na sequência, é selecionada a ferramenta de análise “Regressão” onde a variável Y é inserida, bem como de X1 a X6, obtendo-se, deste modo, a Tabela 1.

Na Tabela 1, em vermelho, temos a variável independente X4 com alto montante para valor-P de aproximadamente 81,81%, muito superior ao nível de significância (α) adotado de 5%. Além disto, na 2ª parte da Tabela 1, em “Coeficientes”, são apresentados os limites inferiores e superiores do intervalo de confiança, nas colunas “95% inferiores” e “95% superiores”, os quais podem ser expressos como: $P(-0,192 \leq \beta_4 \leq +0,240)$. O teste descrito no item 2.3 indica a não rejeição da hipótese $H_0: \beta_4 = 0$. A não rejeição dessa hipótese é confirmada pelo intervalo de confiança que abrange o valor zero.

Considerando o anteriormente exposto, admite-se que não existe regressão do consumo de eletricidade do setor do agronegócio paulista sobre a potência instalada nas UTEs Biomassas do estado de São Paulo. Portanto, X4 não é estatisticamente significativa para este modelo de regressão, sendo descartada.

Tabela 1 - Avaliação dos valores-P das seis variáveis independentes

Coeficientes		valor-P	95% inferiores	95% superiores
Interseção	358.982,066	0,692462	-1.530.548,003	2.248.512,135
X1: Quantidade de Consumidores	13,674	0,003160	5,320	22,029
X2: Tarifa de Energia Elétrica (R\$/MWh)	-5.017,753	0,011741	-8.758,809	-1.276,696
X3: Consumo de Óleo Diesel (10^3 m^3)	-807,850	0,222354	-2.156,661	540,961
X4: UTE Biomassa (kW)	0,024	0,818148	-0,192	0,240
X5: Usina Fotovoltaica (kW)	1,783	0,043915	0,055	3,511
X6: PIB (milhões R\$)	0,536	0,270721	-0,460	1,531

Fonte: Elaborado pelo autor.

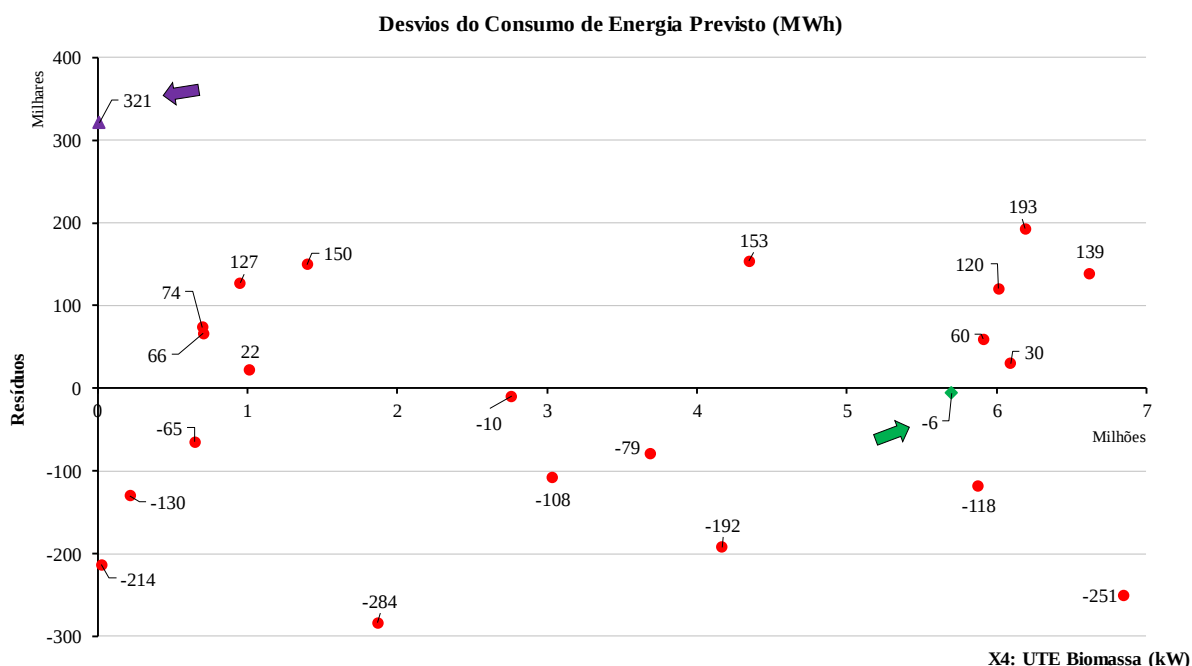
No Gráfico 2, estão plotados os Resíduos (MWh) em relação à potência instalada (kW) das UTEs paulistas que utilizam biomassas como combustível. Os resíduos são as diferenças entre os valores reais do consumo de energia e os valores estimados pelo modelo ajustado.

Nota-se a alternância entre os pontos positivos (12) e os pontos negativos (11), num total de 23 pontos (anos coletados), sinalizando a ausência de um padrão, o que possibilita que sejam considerados aleatórios.

O ponto mais distante (em roxo), demonstrado no Gráfico 2, está com um consumo estimado aproximado de 321.000MWh inferior ao real, enquanto o ponto mais próximo (em verde) entre o

consumo real e o estimado é de 6.000MWh negativo, ou seja, o consumo previsto, neste caso, é superior ao real.

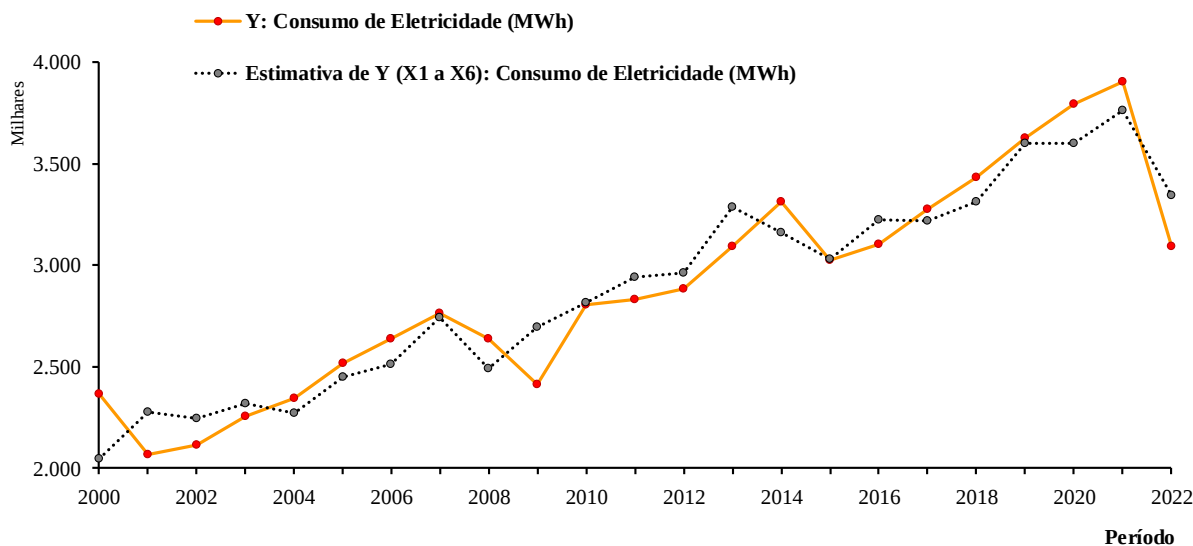
Gráfico 2 - Dispersão dos resíduos do consumo de energia com a potência das UTEs Biomassa



Fonte: Elaborado pelo autor.

No Gráfico 3, têm-se duas linhas gráficas para o mesmo intervalo de tempo de 23 anos, sendo a linha gráfica contínua (em laranja) o consumo de energia real ou observado pelo setor rural paulista, em MWh, em contraste com a linha pontilhada (em preto), que é uma estimativa de consumo desse setor, em função das seis variáveis independentes pesquisadas (X1 a X6), utilizando-se a técnica da regressão linear múltipla. É possível observar erros entre os pontos da linha real e da linha estimada.

Gráfico 3 - Gráficos dos consumos real e estimado de energia para seis variáveis



Fonte: Elaborado pelo autor.

A ferramenta de análise “Regressão” foi novamente executada para a variável Y, bem como todas as variáveis independentes, com exceção de X4 que foi descartada, gerando a Tabela 2. Nela, descrito em vermelho, está a variável X6 com valor aproximado para valor-P de 11,48%, sendo superior ao nível de significância praticado de 5%.

Observando também os limites das colunas "95% inferiores" e "95% superiores" para o intervalo de confiança de X6 igual a $P(-0,162 \leq \beta_6 \leq +1,367)$, isto significa que β_6 passa por zero, apontando a não rejeição da hipótese $H_0: \beta_0 = 0$, descrita no item 2.3.

Pelas razões elencadas anteriormente, concebe-se que não existe regressão do consumo de eletricidade do agronegócio paulista sobre o PIB do agronegócio brasileiro.

Tabela 2 - Avaliação dos coeficientes para cinco variáveis independentes

Coeficientes		valor-P	95% inferiores	95% superiores
Interseção	353.073,910	0,688524	-1.473.692,129	2.179.839,948
X1: Quantidade de Consumidores	14,102	0,000660	6,945	21,259
X2: Tarifa de Energia Elétrica (R\$/MWh)	-5.053,089	0,008820	-8.658,034	-1.448,145
X3: Consumo de Óleo Diesel (10 ³ m ³)	-921,762	0,033122	-1.760,465	-83,058
X5: Usina Fotovoltaica (kW)	1,696	0,027905	0,208	3,185
X6: PIB (milhões R\$)	0,602	0,114887	-0,162	1,367

Fonte: Elaborado pelo autor.

O modelo foi reconstruído pela última vez, agora, inserindo-se a variável Y, assim como todas as variáveis X, com exceção das variáveis X4 e X6 que foram excluídas por não serem, estatisticamente, significativas.

O programa gerou a Tabela 3, cujos montantes de valor-P estão destacados em laranja, com valores aproximados iguais a: 0,014% para X1, 3,124% de X2, 0,264% para X3 e 1,169% de X5. Essas quatro variáveis foram estatisticamente aceitas, considerando uma margem de erro (nível de significância) de 5%, reconhecendo-se a existência da regressão do consumo de energia sobre essas quatro variáveis.

Tabela 3 - Análise de dados com quatro variáveis independentes

RESUMO DOS RESULTADOS

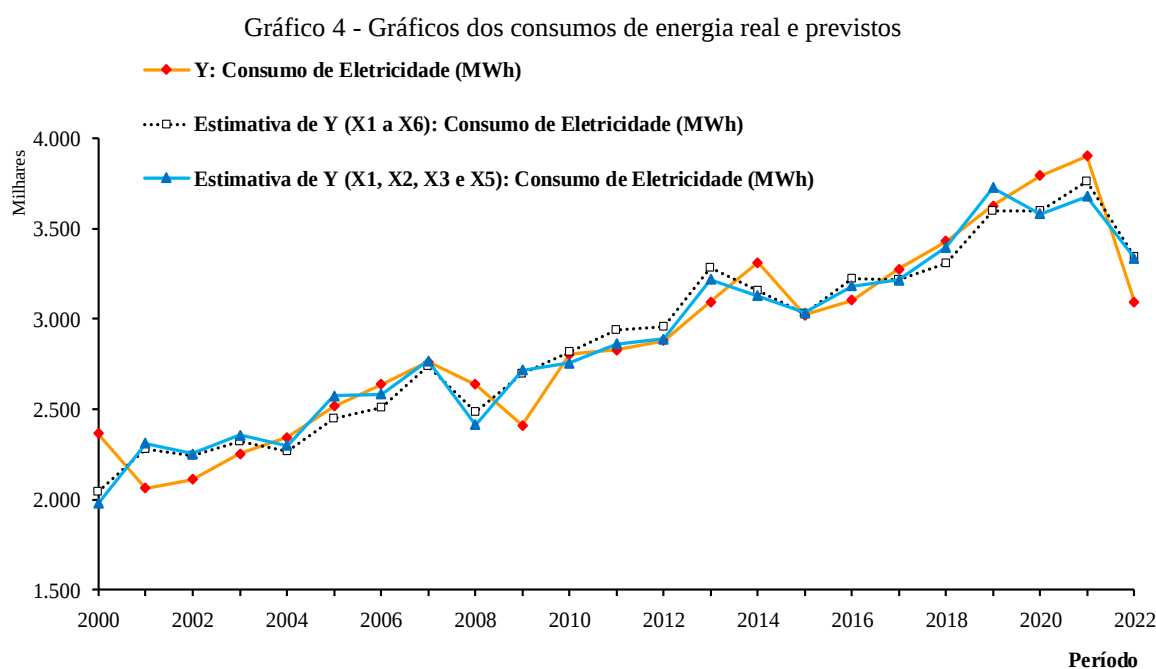
Estatística de regressão	
R múltiplo	0,9465
R-Quadrado	0,8958
R-quadrado ajustado	0,8726
Erro padrão	185.374,5251
Observações	23
ANOVA	
	F de significação
Regressão	4
Resíduo	18
Total	22

Coeficientes		valor-P
Interseção	319.305,420	0,728883
X1: Quantidade de Consumidores	16,044	0,000141
X2: Tarifa de Energia Elétrica (R\$/MWh)	-3.529,563	0,031241
X3: Consumo de Óleo Diesel (10 ³ m ³)	-1.254,294	0,002643
X5: Usina Fotovoltaica (kW)	2,002	0,011692

Fonte: Elaborado pelo autor.

No Gráfico 4, estão plotados três linhas gráficas para o período de 2000 a 2022 (23 anos). A linha gráfica em laranja revela a variação do consumo de eletricidade real ou observada (MWh) do setor agropecuário paulista, enquanto a linha pontilhada na cor preta é uma previsão do consumo do setor, em MWh, em função de todas as variáveis independentes selecionadas (X1 a X6).

A linha em azul no gráfico é a previsão do consumo de eletricidade (MWh) do setor agropecuário paulista, para o mesmo período em questão, considerando quatro variáveis independentes (X1, X2, X3 e X5). Os pontos plotados, tanto na linha pontilhada em preto como na linha contínua em azul, foram obtidos com a técnica estatística da Regressão Linear Múltipla.



Fonte: Elaborado pelo autor.

3.2 GERAÇÃO DO MODELO DE REGRESSÃO LINEAR MÚLTIPLA

Os coeficientes de interseção (b_0) e das variáveis independentes X1 (b_1), X2 (b_2), X3 (b_3) e X5 (b_5) foram coletados da Tabela 3 (em azul), conforme segue:

- $b_0 = + 319.305,420 \Rightarrow$ consumo de energia estimado, não associado às variáveis independentes;
- $b_1 = + 16,044 \Rightarrow$ acréscimo estimado no consumo de energia para cada consumidor adicional, mantidas constante as demais variáveis independentes;

- $b_2 = - 3.529,563 \Rightarrow$ decréscimo estimado no consumo de energia para cada aumento da tarifa de energia elétrica, em R\$/MWh), mantendo inalteradas as outras variáveis;
- $b_3 = - 1.254,294 \Rightarrow$ decréscimo estimado no consumo de energia para cada 10^3 m^3 de óleo diesel adicional, mantendo constantes as demais variáveis independentes;
- $b_5 = + 2,002 \Rightarrow$ acréscimo estimado no consumo de energia para cada quilowatt (kW) de potência adicional nas usinas fotovoltaicas paulistas, mantendo constantes as demais variáveis independentes.

Substituindo os cinco coeficientes acima na Equação (4), temos a Equação (5):

$$\hat{Y} = 319305,42 + 16,044 \cdot X_1 - 3529,563 \cdot X_2 - 1254,294 \cdot X_3 + 2,002 \cdot X_5 \quad (5)$$

A Equação (5) é o modelo estatístico ajustado da regressão linear múltipla para a previsão do consumo de eletricidade do setor agropecuário paulista, considerando as variáveis selecionadas nesta pesquisa.

Este modelo de regressão linear múltipla prevê para o setor pesquisado:

- Aumento aproximado de 16,044MWh no consumo de energia elétrica para cada incremento de 1 (um) consumidor;
- Redução de cerca de 3.529,563MWh no consumo de energia para cada aumento de R\$ 1,00 por MWh na tarifa de eletricidade;
- Redução em torno de 1.254,294MWh no consumo de energia para cada incremento de 1 (um) 10^3 m^3 de óleo diesel utilizado;
- Aumento próximo de 2,002MWh no consumo de energia para cada unidade de kW, acrescida de potência nas usinas fotovoltaicas paulistas.

3.3 TESTE DA EXISTÊNCIA DA REGRESSÃO

Da Tabela 3 (em verde), é coletado o valor aproximado de “F de significação” de $1,314 \times 10^{-6}$ %. Como “F de significação” é menor que α (nível de significância adotado em 5%), conclui-se que existe regressão, ou seja, as variáveis explicativas (X_1 , X_2 , X_3 e X_5), em conjunto, têm uma associação estatisticamente relevante que pode explicar e prever Y (consumo de eletricidade), segundo descrição do item 2.4.

3.4 ANÁLISE DOS RESULTADOS DO MODELO DE REGRESSÃO LINEAR MÚLTIPLA

3.4.1 Qualidade do Ajuste

3.4.1.1 Coeficiente de Correlação Linear

O Coeficiente de Correlação (R múltiplo) está destacado em rosa na Tabela 3, de valor igual a 0,9465. Esse montante sugere uma grande semelhança de comportamento entre as variáveis, apontando para o mesmo sentido. Em outras palavras, existe uma correlação positiva considerável entre as variáveis (MARTINS; DOMINGUES, 2014).

3.4.1.2 Coeficiente de Determinação ou de Explicação

Extraindo da Tabela 3 o Coeficiente de Determinação ou Poder Explicativo (R-Quadrado), em violeta, de valor igual a 0,8958, e empregando o arredondamento de uma casa decimal, o montante de R^2 torna-se 0,9.

Constata-se que a proporção em Y (consumo de eletricidade), que é explicada, em conjunto, por X1 (quantidade de consumidores), X2 (tarifa de energia elétrica), X3 (consumo de óleo diesel) e X5 (geração de energia elétrica pelas usinas fotovoltaicas) por meio do modelo de regressão linear obtido, possui alto poder de explicação ($0,9 \leq R^2 < 1,0$), de acordo com Martins e Domingues (2014).

Portanto, as variáveis X1, X2, X3 e X5, em conjunto, por intermédio do modelo, explicam 89,58% das alterações ou modificações ocorridas em Y. O restante (10,42%) pode ser conferido a outras variáveis eventuais (não inseridas neste modelo).

3.4.1.3 Erro Padrão da Estimativa

Na Tabela 3, está disponível o valor do “Erro Padrão” próximo de 185.374MWh (destacado em dourado). Nesta pesquisa, os valores reais (observados) estão em média 185.374MWh afastados da reta de regressão linear deste modelo.

Na Tabela 4, estão descritos os erros percentuais dos consumos estimados pela Equação (5) do modelo ajustado em comparação com os valores do consumo de energia reais (observados) no setor.

Tabela 4 - Erros entre os consumos reais e os estimados pelo modelo

Observação	Erro (%)
1	-16,26
2	12,04
3	6,73
4	4,61
5	-1,83
6	2,34
7	-2,03
8	0,15
9	-8,50
10	12,63
11	-1,81
12	1,17

13	0,37
14	4,05
15	-5,55
16	0,40
17	2,56
18	-1,80
19	-0,96
20	2,77
21	-5,56
22	-5,69
23	7,82

Fonte: Elaborado pelo autor.

O erro padrão é mais uma maneira de medir a precisão do cálculo, sendo uma espécie de desvio padrão em relação à linha de regressão. Quanto menor for o erro padrão da estimativa, mais acurado será o modelo de previsão.

E por fim, na linha “Observações”, verifica-se o número 23, que se refere ao tamanho da amostragem pesquisada (23 anos).

3.4.2 Análise de Resíduos (ϵ)

O período abrangido pela pesquisa é de 23 anos (23 observações), sendo as informações da coluna “Resíduos”, destacada na cor marrom (Tabela 5), retiradas da ferramenta “Regressão” do Microsoft Excel®.

Tabela 5 - Teste de Shapiro-Wilk

Observação	Resíduos	Teste de Shapiro-Wilk	
1	384.528,969	Tamanho da amostra	23
2	-248.358,280	Média	0,000
3	-142.160,421	Desvio padrão	167.677,567
4	-103.973,389	W	0,9725
5	42.852,549	p	0,7206
6	-58.843,459		
7	53.570,456		
8	-4.245,696		
9	224.231,042		
10	-304.571,083		
11	50.691,540		
12	-32.989,903		
13	-10.712,205		
14	-125.362,612		
15	183.952,766		
16	-12.102,422		
17	-79.460,857		
18	59.001,669		
19	32.973,947		
20	-100.334,068		
21	210.936,122		
22	222.182,276		
23	-241.806,940		

Fonte: Adaptado pelo autor do software BioEstat®.

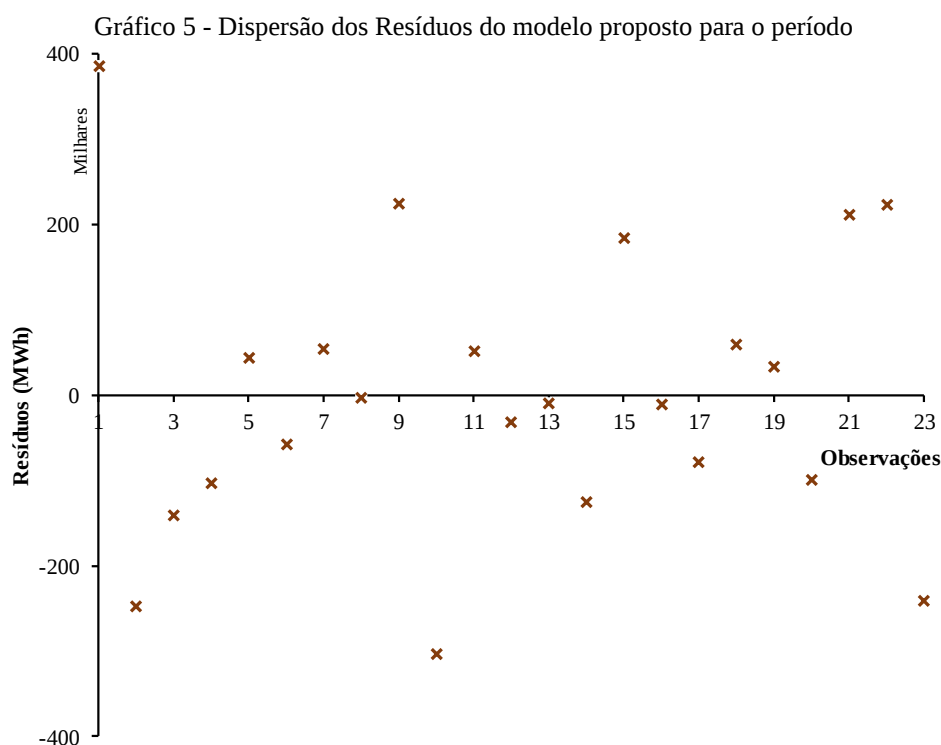
O Teste de Shapiro-Wilk, por meio do *software* BioEstat[®] versão 5.3⁶, é executado para calcular a média da distribuição de probabilidade da variável ε (item 2.5.3, parte a), bem como verificar a sua normalidade (item 2.5.3, parte c).

O resultado da média da distribuição de probabilidade da variável ε é de valor igual zero (destaque em vermelho) na Tabela 5.

Quanto à verificação da normalidade da distribuição de probabilidade da variável ε , de acordo com a Tabela 5, o montante de p-valor é de 0,7206 (destaque em verde), ou seja, 72,06%. Como p-valor $> \alpha$, para um nível de significância (α) de 5%, não se rejeita a hipótese H_0 . Portanto, há indícios de que a distribuição dos erros ou resíduos é normalmente distribuída.

Nota-se também, de acordo com a Tabela 5 (em marrom), a alternância de sinais positivos (dez) e negativos (treze) dos “Resíduos”, confirmando um bom ajustamento do modelo.

No Gráfico 5, os valores dos resíduos estão plotados aleatoriamente ao redor do eixo Observações (23 anos), sem mostrar qualquer tipo de padrão. Portanto, deduz-se que os resíduos são independentes entre si, notando-se apenas o efeito das variáveis independentes sobre a dependente, o que significa a inexistência de autocorrelação residual.



Fonte: Elaborado pelo autor.

⁶ <https://www.mamiraua.org.br/downloads/programas/>

3.4.3 Averiguando a Existência de Multicolinearidade Entre a Variável Dependente e as Variáveis Independentes Seleccionadas

As informações das variáveis Y, X1, X2, X3 e X5 foram transportadas para o editor de planilhas Microsoft Excel®, onde usou-se a configuração de execução do menu: Dados => Análise de Dados => Correlação => OK, concebendo a Tabela 6.

Analisando a tabela, nota-se que a variável dependente (Y) possui, em ordem decrescente de intensidade de correlação (em módulo), uma forte correlação com a variável X3 de 87,28%, seguida pela variável X1, com 80,87%, X2, com 75,63% e, por fim, a menor correlação, com percentual de 60,89%, com a variável X5.

De acordo com Martins e Domingues (2014), a ordenação das variáveis independentes descritas no parágrafo anterior será a mesma em nível de importância de contribuição para uma boa resposta da análise de regressão em estudo, ou seja, a variável X3 é a mais significativa, seguida por X1 e X2, determinando para X5 a menor participação na regressão.

Segundo Gil (2021), as correlações da Tabela 6 podem ser descritas por frases, tais como:

- A correlação entre Y e X3 de 87,28% (negativa) é caracterizada como uma correlação negativa muito forte por estar no intervalo entre -70% e -99%;
- As correlações positivas entre Y e X1 de 80,87% e entre Y e X2 igual a 75,63% são retratadas como correlações positivas muito fortes por estarem inseridas entre 70% e 99%;
- A correlação entre Y e X5 de 60,89% (positiva) é especificada como uma correlação positiva substancial por encontrar-se entre 50% e 69%.

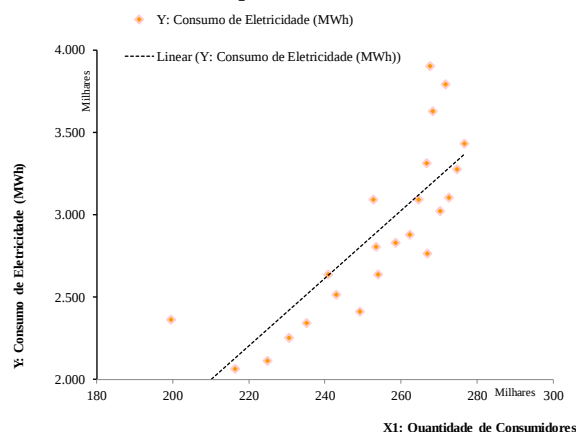
Tabela 6 - Correlação entre a variável dependente em relação às independentes

Correlação	Y:
	Consumo de Eletricidade (MWh)
Y:	1
Consumo de Eletricidade (MWh)	
X1:	+ 0,8087
Quantidade de Consumidores	
X2:	+ 0,7563
Tarifa de Energia Elétrica (R\$/MWh)	
X3:	- 0,8728
Consumo de Óleo Diesel (10 ³ m ³)	
X5:	+ 0,6089
Usina Fotovoltaica (kW)	

Fonte: Elaborado pelo autor.

Pode-se constatar no Gráfico 6, que representa o consumo de energia pelo setor do agronegócio paulista relativo à quantidade de seus consumidores, uma correlação positiva, isto é, as variáveis estão relacionadas de forma positiva e crescem no mesmo sentido (MARTINS; DOMINGUES, 2014), com poucos pontos dispersos em torno da linha de tendência linear pontilhada preta.

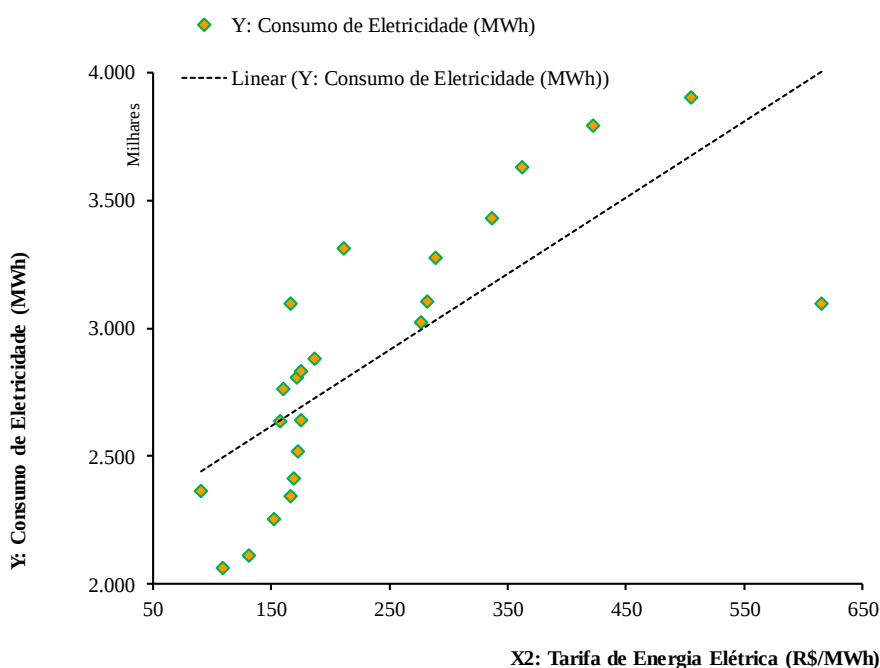
Gráfico 6 - Dispersão de Y relativo a X1



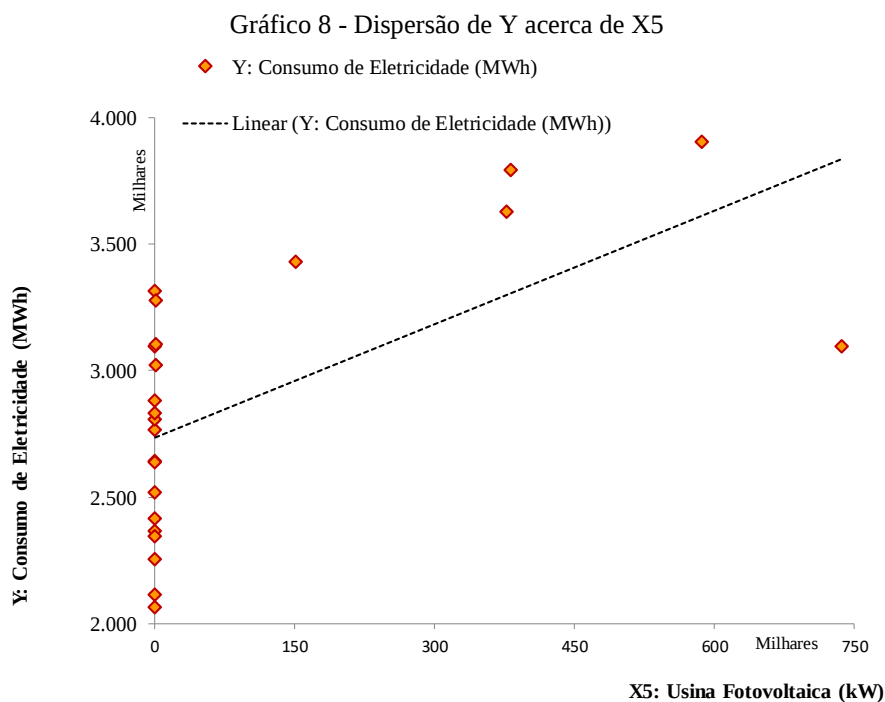
Fonte: Elaborado pelo autor.

Quanto ao Gráfico 7, que ilustra o consumo de energia do setor rural paulista com relação a tarifa de energia elétrica, e o Gráfico 8 sobre o consumo de energia desse setor acerca da potência instalada das usinas fotovoltaicas, ambos possuem uma correlação positiva, porém com muitos pontos dispersos em referência às linhas de tendência lineares pontilhadas (SPIEGEL; STEPHENS, 2009).

Gráfico 7 - Dispersão de Y com relação a X2



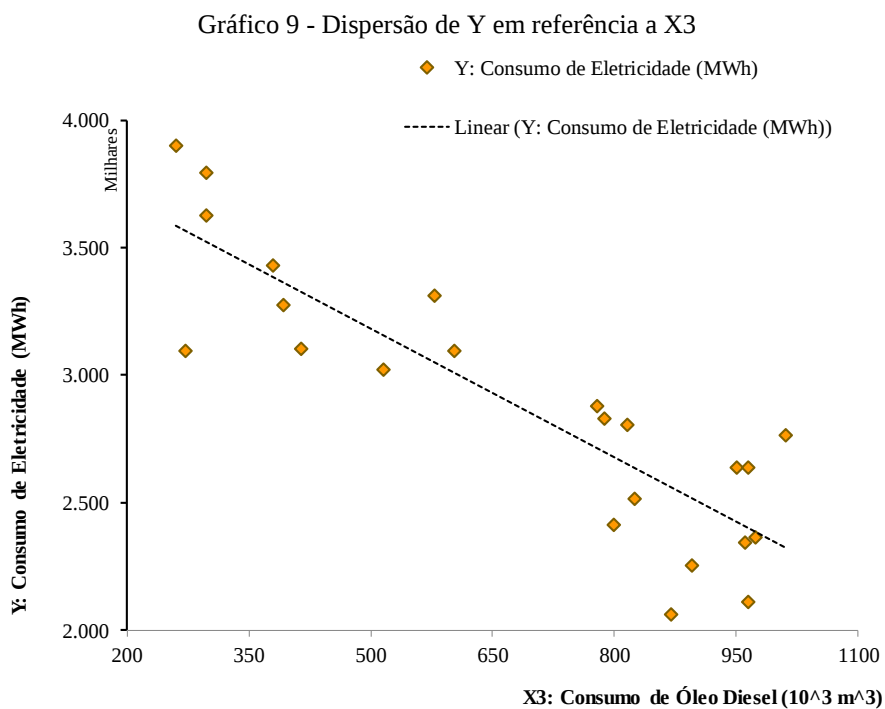
Fonte: Elaborado pelo autor.



Fonte: Elaborado pelo autor.

No caso do Gráfico 9, que trata do consumo de eletricidade pelo setor do agronegócio paulista relacionado ao consumo de óleo diesel desse setor, a correlação é negativa. Isto significa que o crescimento de uma variável se dá com o decréscimo da outra (SPIEGEL; STEPHENS, 2009).

Há dispersão dos pontos em referência à linha de tendência linear de cor preta pontilhada.



Fonte: Elaborado pelo autor.

3.4.4 Verificando a Existência de Multicolinearidade Entre as Variáveis Independentes

Executando novamente a ferramenta de análise “Correlação”, nesta ocasião, somente com as informações das variáveis independentes X1, X2, X3 e X5, reproduziu-se a Tabela 7.

De acordo com Gil (2021), a variável X2 possui uma correlação positiva muito forte com X5 igual a 92,34%, por estar inserida entre 70% e 99%, enquanto os valores das correlações entre X2 e X3 (-87%) e entre X3 e X5 (-70,53%) são caracterizados como correlações negativas muito fortes por estarem localizados entre -70% e -99%.

Pelo anteriormente exposto, constata-se que X2 está fortemente correlacionada com X3 (87%) e, principalmente, com X5 (92,34%). Essa falha se chama Multicolinearidade que, segundo Martins e Domingues (2014), apresenta um obstáculo extra na análise, já que, os resultados finais ficam mascarados, sobretudo, os testes de hipóteses devido à influência de uma variável independente sobre a outra. Ademais, essa distorção também acarreta a violação das premissas do modelo.

A variável X1 tem uma correlação negativa substancial com X3 (-67,90%) e positiva substancial com X2 (53,20%), conforme observado na Tabela 7, por seus valores estarem no intervalo de -50% a -69% e 50% a 69%, respectivamente (GIL, 2021).

A respeito da correlação de 26,07% entre X1 e X5 (Tabela 7), Gil (2021) relata que é uma correlação baixa por estar entre 10% e 29%.

Tabela 7 - Correlação entre as variáveis independentes

Correlação	X1:	X2:	X3:	X5:
X1:	1			
Quantidade de Consumidores				
X2:	+ 0,5320	1		
Tarifa de Energia Elétrica (R\$/MWh)				
X3:	- 0,6790	- 0,8700	1	
Consumo de Óleo Diesel (10 ³ m ³)				
X5:	+ 0,2607	+ 0,9234	- 0,7053	1
Usina Fotovoltaica (kW)				

Fonte: Elaborado pelo autor.

O Teste de Correlação de Pearson, via *software* BioEstat[®], também foi usado na investigação da correlação entre as variáveis independentes.

Avaliando a correlação entre as variáveis X1 e X2, observou-se com destaque em verde (Tabela 8) o coeficiente de correlação de Pearson, r (Pearson), de valor igual a 0,5320. Estando o coeficiente na faixa $0 < r < 1$, significa que se trata de uma correlação linear positiva, em que as variáveis se alteram no mesmo sentido.

Como o valor de r (Pearson) está dentro do intervalo $0,50 < r < 0,69$, é possível concluir que se trata de uma correlação substancial. Quanto ao p-valor (em azul claro na Tabela 8), tem-se que: $0,0089 < 0,05$, rejeitando-se a hipótese H_0 para um risco (nível de significância) de 5%, conforme descrito no item 2.5.4; logo, há indícios de que as variáveis X1 e X2 estão correlacionadas.

Tabela 8 - Correlação entre a quantidade de consumidores e a tarifa de eletricidade

X1	X2 (R\$/MWh)	Teste de Correlação Linear	
199.539	90,30	n (pares)	23
216.269	109,45	r (Pearson)	+ 0,5320
224.846	130,92	(p)	0,0089
230.601	152,55		
235.127	166,19		
243.022	172,67		
253.998	175,00		
266.882	160,58		
240.722	158,03		
249.125	169,32		
253.394	171,71		
258.543	174,96		
262.243	186,66		
264.528	166,73		
266.735	211,42		
270.180	276,62		
272.649	281,68		
274.610	288,71		
276.707	336,77		
268.339	362,69		
271.668	421,97		
267.680	505,28		
252.614	615,37		

Fonte: Adaptado pelo autor do software BioEstat®.

Quanto à correlação acerca das variáveis X1 e X3, de acordo com o destaque em roxo da Tabela 9, o coeficiente de correlação r (Pearson) é de 0,6790 (negativo). Neste caso, o coeficiente está na faixa $-1 < r < 0$, evidenciando uma correlação linear negativa, na qual as variáveis se alteram em sentido contrário.

O montante de r (Pearson) está entre $-0,50 < r < -0,69$, mostrando que se trata de uma correlação substancial. Considerando o p-valor (em rosa claro na Tabela 9) igual a 0,0004, sendo $0,0004 < 0,05$, rejeita-se a hipótese H_0 para um risco de 5%, segundo especificado no item 2.5.4; portanto, há indícios de que as variáveis X1 e X3 estão correlacionadas.

Tabela 9 - Correlação entre a quantidade de consumidores e o consumo de óleo diesel

X1	X3 (10^3 m^3)	Teste de Correlação Linear	
199.539	974	n (pares)	23
216.269	870	r (Pearson)	- 0,6790
224.846	965	(p)	0,0004
230.601	895		
235.127	960		
243.022	824		
253.998	950		
266.882	1.010		
240.722	965		
249.125	799		
253.394	816		
258.543	787		
262.243	779		
264.528	602		
266.735	578		

270.180	515		
272.649	414		
274.610	392		
276.707	379		
268.339	297		
271.668	297		
267.680	260		
252.614	271		

Fonte: Adaptado pelo autor do software BioEstat®.

Verificando a correlação entre as variáveis X1 e X5, nota-se, com destaque em laranja (Tabela 10), o coeficiente de correlação r (Pearson) de valor igual a 0,2607. Como o coeficiente encontra-se na faixa de $0 < r < 1$, tem-se uma correlação linear positiva e, conseqüentemente, as variáveis convertem para o mesmo caminho.

O valor do coeficiente está no intervalo de $0,10 < r < 0,29$; portanto, trata-se de uma correlação baixa. Com destaque em cinza na Tabela 10, tem-se o p -valor de 0,2294, ou seja, $0,2294 > 0,05$; logo, não se rejeita a hipótese H_0 para um risco (nível de significância) de 5%, indicando que as variáveis X2 e X3 não estão correlacionadas.

Tabela 10 - Correlação entre a número de consumidores e a potência das usinas fotovoltaicas

X1	X5 (kW)	Teste de Correlação Linear	
199.539	0	n (pares)	23
216.269	0	r (Pearson)	+ 0,2607
224.846	0	(p)	0,2294
230.601	0		
235.127	0		
243.022	0		
253.998	0		
266.882	0		
240.722	0		
249.125	0		
253.394	0		
258.543	0		
262.243	0		
264.528	0		
266.735	0		
270.180	1.100		
272.649	1.100		
274.610	1.100		
276.707	151.217		
268.339	377.426		
271.668	382.426		
267.680	587.064		
252.614	736.887		

Fonte: Adaptado pelo autor do software BioEstat®.

O montante do coeficiente de correlação r (Pearson), realçado em marinho na Tabela 11, é de 0,8700 (negativo) e está na faixa $-0,50 < r < -0,69$, comprovando uma correlação linear negativa muito forte (as variáveis caminham em sentidos opostos).

Com um p-valor (em verde na Tabela 11) menor que 0,0001 e, por conseguinte, menor que 0,05, rejeita-se a hipótese H_0 para um risco de 5% (detalhado no item 2.5.4); isto posto, há evidência de que as variáveis X2 e X3 estão correlacionadas.

Tabela 11 - Correlação entre a tarifa de eletricidade e o consumo de óleo diesel

X2 (R\$/MWh)	X3 (10 ³ m ³)	Teste de Correlação Linear	
90,30	974	n (pares)	23
109,45	870	r (Pearson)	- 0,8700
130,92	965	(p)	< 0,0001
152,55	895		
166,19	960		
172,67	824		
175,00	950		
160,58	1.010		
158,03	965		
169,32	799		
171,71	816		
174,96	787		
186,66	779		
166,73	602		
211,42	578		
276,62	515		
281,68	414		
288,71	392		
336,77	379		
362,69	297		
421,97	297		
505,28	260		
615,37	271		

Fonte: Adaptado pelo autor do *software* BioEstat[®].

Segundo a Tabela 12, o coeficiente de correlação de Pearson, r (Pearson), tem valor igual a 0,9234 (destaque em vermelho). Como este valor está dentro do intervalo $0 < r < 1$, conclui-se que se trata de uma correlação linear positiva muito forte, pois o coeficiente está no intervalo entre 0,70 e 0,99.

Sendo p-valor < 0,0001 (em amarelo na Tabela 12) menor que 5%, rejeita-se a hipótese nula para o nível de significância adotado; portanto, há indícios de que as variáveis X2 e X5 estão correlacionadas.

Tabela 12 - Correlação entre a tarifa de eletricidade e a potência das usinas fotovoltaicas

X2 (R\$/MWh)	X5 (kW)	Teste de Correlação Linear	
90,30	0	n (pares)	23
109,45	0	r (Pearson)	+ 0,9234
130,92	0	(p)	< 0,0001
152,55	0		
166,19	0		
172,67	0		
175,00	0		
160,58	0		
158,03	0		
169,32	0		

171,71	0		
174,96	0		
186,66	0		
166,73	0		
211,42	0		
276,62	1.100		
281,68	1.100		
288,71	1.100		
336,77	151.217		
362,69	377.426		
421,97	382.426		
505,28	587.064		
615,37	736.887		

Fonte: Adaptado pelo autor do software BioEstat®.

Observando a Tabela 13, com realce em vermelho, tem-se o coeficiente de correlação r (Pearson) de X_3 e X_5 igual a 0,7053 (negativo) inserido no intervalo: $-0,70 < r < -0,99$, apresentando uma correlação linear negativa muito forte (variam em sentido contrário).

Como $p\text{-valor} < \alpha$, ou seja, $0,02\% < 5\%$ (em azul na Tabela 13), rejeita-se a hipótese nula para o nível de significância adotado (discriminado no item 2.5.4); portanto, há evidência de que existe correlação linear entre elas.

Tabela 13 - Correlação entre o consumo de óleo diesel e a potência das usinas fotovoltaicas

X3 (10 ³ m ³)	X5 (kW)	Teste de Correlação Linear	
		n (pares)	23
974	0	r (Pearson)	- 0,7053
870	0	(p)	0,0002
965	0		
895	0		
960	0		
824	0		
950	0		
1.010	0		
965	0		
799	0		
816	0		
787	0		
779	0		
602	0		
578	0		
515	1.100		
414	1.100		
392	1.100		
379	151.217		
297	377.426		
297	382.426		
260	587.064		
271	736.887		

Fonte: Adaptado pelo autor do software BioEstat®.

A variável X_1 (quantidade de consumidores) foi a que apresentou a menor correlação (associação) com as demais variáveis, com os seguintes coeficientes de correlação r (Pearson):

- +53,20% (Tabela 8) com X_2 (tarifa de energia elétrica, em R\$/MWh);

- -67,90% (Tabela 9) com X3 (consumo de óleo diesel, em 10^3m^3);
- +26,07% (Tabela 10) com X5 (Usina Fotovoltaica, em kW).

De modo oposto, foi apontado que a variável X2 ostentou as maiores correlações (associações) com as variáveis pesquisadas, especialmente, com X3 e X5, conforme os coeficientes de correlação r (Pearson) exibidos a seguir:

- -87,00% (Tabela 11) com X3 (consumo de óleo diesel, em 10^3m^3);
- +92,34% (Tabela 12) com X5 (Usina Fotovoltaica, em kW).

Segundo Cunha e Coelho (2014), o problema da multicolinearidade refere-se, principalmente, ao grau, e não à natureza. É inevitável a presença de correlação entre as variáveis independentes, sendo preferível selecionar aquelas com menor grau para evitar complicações na interpretação dos resultados.

O cenário ideal seria possuir múltiplas variáveis independentes alta ou perfeitamente correlacionadas com a variável dependente, porém com ínfima ou nenhuma correlação entre si.

Portanto, pode-se considerar descartar a variável X2 por sua forte associação com as demais variáveis independentes.

4 CONCLUSÃO

O desenvolvimento desta pesquisa forneceu contribuições consideráveis para o alcance dos objetivos aqui propostos.

A geração de um modelo estatístico de estimativa do consumo de eletricidade pelo setor agropecuário paulista, que é o objetivo geral, foi alcançada, bem como o objetivo específico, que era identificar as variáveis que contribuem para essa estimativa.

Foram escolhidas seis variáveis para construir o modelo, sendo três do setor do agronegócio paulista: a quantidade de consumidores, a tarifa de energia elétrica e o consumo de óleo diesel; duas referentes ao estado de São Paulo: potências instaladas das usinas térmicas de biomassa e das usinas fotovoltaicas e uma do agronegócio brasileiro: PIB.

Dessas, duas variáveis foram descartas: a potência instalada das usinas térmicas de biomassa paulistas e o PIB do agronegócio brasileiro, pois o modelo de regressão linear as considerou estatisticamente não significativas. Uma possível causa para a exclusão dessa variável é pelo fato de ter-se considerado o PIB nacional, e não o paulista, de um país continental como o Brasil, que é um grande exportador de *commodities* agrícolas e pecuárias.

As demais variáveis integraram o modelo estatístico, porém com níveis de influência distintos. A potência total instalada das usinas fotovoltaicas paulistas apresentou uma influência substancial, enquanto as demais – quantidade de consumidores, tarifa de energia elétrica e consumo de óleo diesel

do âmbito rural paulista –, apresentaram influências muito fortes na estimativa do consumo de energia elétrica do setor estudado.

Constatou-se que o consumo de óleo diesel foi a variável estudada mais significativa, porém com ação inversa (sinal negativo no modelo estatístico ajustado) na regressão linear múltipla, comprovando, desse modo, sua substituição pela energia elétrica como principal fonte energética do setor rural paulista.

Em trabalhos futuros, a metodologia aqui aplicada poderia empregar a técnica de seleção de variáveis *stepwise*, conhecida também como método passo a passo ou por etapas, que é uma abordagem frequente na busca sequencial. É possível, por esse método, examinar o impacto adicional de cada variável independente no modelo, já que a inclusão de cada variável é avaliada antes da formulação da equação matemática.

A presente pesquisa pode contribuir para o entendimento de elementos que estimulam o crescimento do consumo de energia do setor agropecuário paulista, disponibilizando ferramentas de análise e de estimação futura aos responsáveis pelo setor elétrico do estado para sua manutenção e ampliação permanentes, possibilitando, dessa forma, a continuidade do fornecimento de energia elétrica, um serviço público primordial, para a realização de tarefas cotidianas da comunidade rural paulista.

AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) – Código de Financiamento 001.

Agradeço ao Instituto Federal de São Paulo (IFSP) por me conceder o afastamento remunerado para qualificação *Stricto Sensu*.

REFERÊNCIAS

BISOGNIN, C.; WERNER, L. Análise do consumo mensal de energia elétrica no Estado de São Paulo. REP – Revista de Engenharia de Produção, Brasil, v. 2, p. 59-72, jan./abr. 2020. Disponível em: <https://periodicos.ufms.br/index.php/REP/article/view/9397>. Acesso em: 28 nov. 2024.

BRASIL. Empresa de Pesquisa Energética (EPE). Anuário Estatístico de Energia Elétrica 2024: Ano Base 2023. Rio de Janeiro: EPE, 2024. 6 p. Disponível em: <https://www.epe.gov.br/sites-pt/publicacoes-dados-abertos/publicacoes/PublicacoesArquivos/publicacao-160/topico-168/anuario-factsheet-2024.pdf>. Acesso em: 28 nov. 2024.

BRASIL. Ministério de Minas e Energia (MME). Balanço Energético Nacional 2024: Ano Base 2023. Rio de Janeiro: EPE, 2024. 274 p. Disponível em: <https://www.epe.gov.br/sites-pt/publicacoes-dados-abertos/publicacoes/PublicacoesArquivos/publicacao-819/topico-723/BEN2024.pdf>. Acesso em: 28 nov. 2024.

BRASIL. Ministério de Minas e Energia (MME). Plano Nacional de Energia 2030: Combustíveis Líquidos. Brasília: MME, 2007. v. 12, 98 p. Disponível em: <https://www.epe.gov.br/sites-pt/publicacoes-dados-abertos/publicacoes/PublicacoesArquivos/publicacao-165/topico-173/PNE%202030%20-%20Combust%C3%ADveis%20L%C3%ADquidos.pdf>. Acesso em: 28 nov. 2024.

CUNHA, J. V. A. da; COELHO, A. C. Regressão linear múltipla. In: CORRAR, L. J.; PAULO, E.; DIAS FILHO, J. M. (Coord.). Análise multivariada: para os cursos de administração, ciências contábeis e economia. 1. ed. 7. reimpr. São Paulo: Atlas, 2014. p. 131-231. ISBN: 978-85-224-4707-7.

GIL, A. C. Métodos e técnicas de pesquisa social. 7. ed. 3. reimpr. São Paulo: Atlas, 2021. 230 p. ISBN: 978-85-970-2057-1.

MARTINS, G. de A.; DOMINGUES, O. Estatística Geral e Aplicada. 5. ed. rev. e ampl. São Paulo: Atlas, 2014. 416 p. ISBN: 978-85-224-8677-9.

MELLO, D. Lula diz que é preciso melhorar salários de professores. Agência Brasil, São Paulo, out. 2022. Disponível em: <https://agenciabrasil.ebc.com.br/politica/noticia/2022-10/lula-diz-que-e-preciso-melhorar-salarios-de-professores>. Acesso em: 28 nov. 2024.

SÃO PAULO (Estado). Secretaria de Infraestrutura e Meio Ambiente (SIMA). Balanço Energético do Estado de São Paulo (BEESP) 2019: Ano Base 2018. São Paulo: SIMA, 2019. 274 p. Disponível em: <https://smastr16.blob.core.windows.net/2001/2023/12/BEESP2019ab2018.pdf>. Acesso em: 29 nov. 2024.

SÃO PAULO (Estado). Secretaria de Infraestrutura e Meio Ambiente (SIMA). Balanço Energético do Estado de São Paulo (BEESP) 2022: Ano Base 2021. São Paulo: SIMA, 2022. 270 p. Disponível em: <https://dadosenergeticos.energia.sp.gov.br/portalecv2/intranet/BiblioVirtual/diversos/BalancoEnergetico.pdf>. Acesso em: 28 nov. 2024.

SÃO PAULO (Estado). Secretaria de Meio Ambiente, Infraestrutura e Logística (SEMIL). Balanço Energético do Estado de São Paulo (BEESP) 2023: Ano Base 2022. 2. ed. São Paulo: SEMIL, jan. 2024. 160 p. Disponível em: <https://smastr16.blob.core.windows.net/2001/2024/02/BEESP2023ab2022-2a-edicao-2.pdf>. Acesso em: 29 nov. 2024.



SPIEGEL, M. R.; STEPHENS, L. J. Estatística. Tradução: NASCIMENTO, J. L. do, 4. ed. Porto Alegre: Bookman, 2009. 597 p. ISBN: 978-85-7780-461-0. Título original: Schaum's Outlines: Theory and Problems of Statistics, 4th Edition ISBN: 9780071485845.