

MODELO MATEMÁTICO DE ESTIMATIVA DE CONSUMO DE ENERGIA PELO SETOR RURAL PAULISTA

MONCLAR NOGUEIRA CHRISTOVÃO, MARIO MOLLO NETO, SANDRA CRISTINA DE OLIVEIRA, ANA FLÁVIA DE LUCA OLIVEIRA CHRISTOVÃO

RESUMO: A energia elétrica é um item fundamental e indispensável para o desenvolvimento econômico de uma região. No setor agropecuário paulista, até o ano de 2018, o óleo diesel foi a matriz energética mais usada. Porém, nos anos seguintes (2019 e 2020), a energia elétrica, que vinha em uma crescente constante, passou a ser o energético mais consumido pelo referido setor. Dessa forma, o presente trabalho tem por objetivo criar um método matemático de previsão do consumo de eletricidade para a categoria rural do estado de São Paulo. Para tanto, foram utilizadas as teorias de análise de regressão em conjunto com a ferramenta Regressão do editor de planilhas Microsoft Excel®. Foram empregados o Teste de Shapiro-Wilk para apurar a normalidade dos resíduos e o Teste de Correlação de Pearson para avaliar a multicolinearidade das variáveis. Como resultado, foi obtido um modelo matemático de regressão linear múltipla com pouca influência das variáveis “quantidade de consumidores do setor agropecuário” e “energia elétrica gerada no estado”, enquanto que a variável mais significativa, mas com influência inversa, foi o uso do óleo diesel, ratificando sua substituição como principal fonte energética do setor rural pela energia elétrica. Foi possível concluir que a escalada do consumo de energia elétrica pelo setor pesquisado é sólida, permanente e irreversível, e que necessitará de investimentos em benfeitorias e aperfeiçoamento de toda infraestrutura elétrica, em especial, da transmissão e distribuição de energia, o que promoverá o crescimento e desenvolvimento regional, bem como a robustez econômica do estado de São Paulo.

Palavras-chave: Previsão de consumo. Modelo matemático. Consumo de eletricidade. Consumo de energia rural. Demanda de energia.

INTRODUÇÃO

A disponibilidade e o fornecimento contínuo de energia elétrica em um país são uma preocupação permanente de seus governantes, visto que tal fonte energética é vital para o crescimento econômico de uma região (MARANGONI; MAGATÃO; DE ARRUDA, 2020).

A crescente demanda por energia em escala mundial tem provocado apreensão internacional em razão de possíveis flutuações no abastecimento, além do aumento da devastação ambiental e instabilidades geopolíticas (FAN; XIAO; WANG, 2014).

O agravante para o problema acima descrito é a falta de recursos financeiros disponíveis em países pobres ou ditos em “desenvolvimento”, que são necessários para a implantação e manutenção da infraestrutura do setor elétrico dessas nações (TAGHIZADEH-YAZDI; MOHAMMADI-BALANI, 2020).

A estimativa futura da demanda por eletricidade é um recurso fundamental para a gestão do consumo de energia, programação de manutenções e planejamento de investimentos em mercados futuros de energia (VU; MUTTAQI; AGALGAONKAR, 2015).

As estimativas futuras de consumo de eletricidade podem ser desenvolvidas por métodos matemáticos para diferentes periodicidades, a depender da necessidade da programação ou da gestão de companhias do ramo elétrico em curta, média ou longa previsões (ABDULKAREEM *et al.*, 2019).

O consumo total de eletricidade por setor no estado de São Paulo em 2020, compreendendo também os autoprodutores, foi de 145.451 Giga Watts hora (GWh), uma redução de 3,78% em comparação com o ano anterior, que foi igual a 151.120 GWh. Os ramos comercial (-12,34%) e industrial (-2,35%) evidenciaram um decréscimo para o referido período. Em contrapartida, os setores agropecuário (1,62%) e residencial (3,08%) apresentaram crescimento no consumo (SÃO PAULO, 2021).

O objetivo deste trabalho é gerar um modelo matemático para a previsão do consumo de eletricidade do setor agropecuário do estado de São Paulo. Especificamente, pretende-se identificar as variáveis ou a combinação delas que contribuem para a estimativa do consumo de energia desse setor.

Dado o contexto apresentado, justifica-se a necessidade de estudos relacionados a essa temática relevante ao segmento do agronegócio.

MATERIAIS E MÉTODOS

Coleta de dados

Trata-se de uma pesquisa descritiva quanto aos objetivos, com abordagem quantitativa. A coleta de dados foi realizada no arquivo digital do relatório anual intitulado “Balanço Energético do Estado de São Paulo” (BEESP) de 2021, disponibilizado pela Secretaria de Infraestrutura e Meio Ambiente (SIMA) do governo paulista.

Os dados coletados produziram séries históricas entre 2012 e 2021. No entanto, os dados referentes aos gastos com eletricidade de 2020 e 2021 foram afetados pela pandemia da COVID-19.

O recurso escolhido foi a análise de regressão múltipla, que é um dos métodos mais usados em estudos analíticos, principalmente, em previsões. O objetivo é criar um modelo estatístico para estimativas de respostas futuras de uma variável dependente (Y) em função de uma variável (X) ou mais variáveis (X_1 , X_2 , ... e X_n) independente(s), concebendo uma associação matemática entre essas variáveis (MARTINS; DOMINGUES, 2014).

A regressão é simples quando há somente uma variável independente. Na existência de duas ou mais dessas variáveis, o modelo de regressão é múltiplo.

Neste trabalho, considerou-se a variável dependente Y (consumo anual de eletricidade do setor agropecuário paulista, em GWh) como função das variáveis X_1 (quantidade de consumidores) e X_2 (gasto de óleo diesel em 10^3 m^3), que estão vinculadas ao agronegócio, bem como da variável X_3 (eletricidade em GWh), que está relacionada ao estado de São Paulo.

Construção de um modelo de regressão linear múltiplo

O modelo de regressão linear múltiplo busca prever e explicar o comportamento da variável Y em função das variáveis X_1, X_2, \dots, X_n , de acordo com a Equação (1).

$$Y = \beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \dots + \beta_n \cdot X_n + \varepsilon \quad (1)$$

Onde:

- β_0 : coeficiente linear (intercepto da reta); $\beta_1, \beta_2, \dots, \beta_n$: coeficiente angular ou coeficiente de regressão (inclinação da reta); e
- ε : erro experimental de Y (pontos fora da reta).

Com base em dados amostrais, tem-se que \hat{Y} é o resultado da previsão Y para observações X , conforme o modelo ajustado da Equação (2). Assim, a partir de amostras de pares ordenados de três ou mais variáveis, obtêm-se os valores de $b_0, b_1, b_2, \dots, b_n$.

$$\hat{Y} = b_0 + b_1 \cdot X_1 + b_2 \cdot X_2 + \dots + b_n \cdot X_n \quad (2)$$

Onde:

- b_0 : estimador de β_0 ; e
- b_1, b_2, \dots, b_n : estimadores respectivos de $\beta_1, \beta_2, \dots, \beta_n$.

Teste da significância dos coeficientes, da existência da regressão e análise dos resultados

A importância das variáveis independentes é testada individualmente. O teste verifica se cada parâmetro β_i da regressão é igual a zero (hipótese H_0). Assim, se o p-valor for menor ou igual ao nível de significância, rejeita-se a hipótese H_0 e conclui-se que $\beta_i \neq 0$ para um risco α .

Em seguida, verifica-se a existência da regressão linear, considerando o teste F (Análise de Variância - ANOVA). Se o p-valor (F de significação) for menor ou igual ao nível de significância do teste, rejeita-se a hipótese H_0 de que todos os parâmetros β_i da

regressão são iguais a zero e conclui-se que existe regressão, ou seja, que o modelo ajustado pode explicar e prever Y.

O coeficiente de Determinação ou Explicação (R^2) é uma medida da proporção da variação em Y (resposta), que é explicada por X (explicativa) pelo ajuste do modelo linear, sendo, portanto, 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 aos dados.

Recomenda-se ainda a checagem de alguns atributos da distribuição dos resíduos, como:

- a) Averiguar se a média dos resíduos é aproximadamente igual a zero, e
- b) Verificar a normalidade da distribuição de probabilidade da variável ϵ , empregando o Teste de Shapiro-Wilk. Se o p-valor for menor ou igual a α , rejeita-se a hipótese H_0 de existência de normalidade dos dados para um risco (nível de significância) α .

Sugere-se também a investigação da correlação entre as variáveis independentes, duas a duas, empregando o Teste de Correlação de Pearson para verificar o efeito de multicolinearidade, ou seja, se existem variáveis redundantes ao modelo. O teste verifica se a correlação linear entre duas variáveis é nula (hipótese H_0). Logo, se o p-valor for menor ou igual a α (nível de significância do teste), rejeita-se a hipótese H_0 e conclui-se que existe correlação linear significativa entre as duas variáveis.

Neste trabalho, utilizou-se o suplemento “Análise de Dados” do Excel para o ajuste do modelo de regressão aos dados e um nível de significância (α) de 10% para os testes de hipóteses aplicados.

RESULTADOS E DISCUSSÃO

Com a finalidade de explicar o quanto da variação no consumo anual de eletricidade do setor agropecuário paulista se deve ao conjunto considerado de variáveis independentes ou explicativas, ajustou-se um modelo de regressão linear múltiplo a partir das informações coletadas.

A variável X_3 (energia elétrica gerada no estado), com p-valor = 0,522, não foi considerada significativa e, portanto, foi a única variável descartada. Desse modo, para um nível de significância de 10%, o modelo de regressão final ajustado aos dados, conforme Tabela 1, é dado pela Equação (3):

$$\hat{Y} = 11958,88 - 0,03 \cdot X_1 - 2,15 \cdot X_2 \quad (3)$$

Tabela 1 - Resumo da análise de regressão linear múltipla

Estatística de Regressão	
R-Quadrado	81,70%
R-quadrado ajustado	76,47%

Análise de Variância (ANOVA)				
Variação	gl	SQ	F	p-valor (F de significação)
Regressão	2	768782,6	15,62	0,003
Erro Residual	7	172221,02		
Total	9	941003,6		

Preditores	Coefficientes	Estatística-t	p-valor
Interseção	11958,88	2,90	0,023
X_1	-0,03	-1,91	0,098
X_2	-2,15	-5,29	0,001

Fonte: Elaborado pelos autores.

Quanto à significância do modelo de regressão, como p-valor é menor que 10% (F-significação e respectivo p-valor na ANOVA da Tabela 1), há evidências de que o modelo pode explicar e prever a variável resposta Y.

Da mesma forma, como p-valor de X_1 e de X_2 são menores que 10% (vide estatísticas t e respectivos p-valores da Tabela 1), os parâmetros de tais variáveis confirmaram-se como significativos ao modelo.

Avaliou-se o grau de ajustamento do modelo baseado no coeficiente de determinação múltipla (R-quadrado e R-quadrado ajustado na Tabela 1), cujos valores 81,70% e 76,47%, respectivamente, estabelecem porcentagens aceitáveis de variação de Y explicada pelas variáveis independentes significativas (X_1 e X_2).

O teste de Shapiro-Wilk para os resíduos confirmou a normalidade desses, uma vez que o p-valor é igual a 0,639 e não se rejeita a hipótese H_0 . Logo, há indícios de que a distribuição dos resíduos é normalmente distribuída.

O Teste de Correlação de Pearson, via BioEstat[®], foi usado para investigar a correlação entre as variáveis independentes.

Avaliando a correlação entre X_1 e X_2 , o coeficiente de correlação de Pearson “r (Pearson)” tem valor igual a 0,6256 (negativo). Como o valor está no intervalo $-0,50 < r < -0,69$, conclui-se que se trata de uma correlação negativa substancial. Quanto ao p-valor, temos que $0,0530 < 0,10$, portanto, rejeita-se a hipótese H_0 para um risco (nível de significância) de 10%, logo há indícios de que X_1 e X_2 estão correlacionadas.

Quanto à correlação entre X_1 e X_3 , o coeficiente “r (Pearson)” foi de 0,2593 (negativo). Como este valor está entre $-0,10 < r < -0,29$, nota-se uma correlação negativa baixa. Considerando o p-valor de 0,4693 ($> 0,10$), não se rejeita a hipótese H_0 , logo há indícios de que não há correlação linear entre X_1 e X_3 .

Verificando a correlação entre X_2 e X_3 , o coeficiente “r (Pearson)” foi de 0,7222. Este valor está no intervalo de $0,70 < r < 0,99$, portanto, trata-se de uma correlação positiva muito forte. O p-valor foi de 0,0183 ($< 0,10$), logo rejeita-se a hipótese H_0 . Há indícios de que X_2 e X_3 estão correlacionadas.

O resultado anteriormente exposto colabora e fortalece a deliberação de descartar X_3 .

Analisando os sinais dos coeficientes disponíveis na Tabela 1, tanto a variável X_1 (quantidade de consumidores) quanto a variável X_2 (gasto de óleo diesel em 10^3 m^3) apresentaram coeficientes negativos, indicando que essas se movimentam em sentido contrário ao da variável resposta Y (consumo anual de eletricidade do setor agropecuário paulista), quando se observa uma delas e mantém a outra variável constante.

De fato, houve um crescimento contínuo na quantidade de consumidores até o ano de 2018, com uma redução brusca no ano seguinte e uma elevação em 2020. Ainda assim, seu quantitativo foi superior aos anos anteriores a 2016. O óleo diesel foi a fonte energética mais consumida pelo setor agropecuário paulista até 2018, apresentando queda a partir de 2007. Já em 2019 e 2020, o consumo maior do setor foi de energia elétrica (SÃO PAULO, 2021).

CONSIDERAÇÕES FINAIS

Este trabalho alcançou os objetivos propostos e identificou as variáveis que contribuíram para a estimativa desse consumo: o “número de consumidores” com baixa influência, a “energia elétrica gerada” com baixíssima influência, sendo descartada por não ser significativa, e o “consumo de óleo diesel” (com alta influência) na estimativa do consumo de energia do setor agropecuário paulista.

Constatou-se que o consumo de óleo diesel foi a variável estudada mais significativa, porém com ação inversa – sinal negativo no modelo de regressão linear ajustado –, comprovando, desse modo, sua substituição pela eletricidade como principal fonte energética do setor estudado.

Outro item a ser considerado em pesquisas futuras é a inserção de parâmetros climáticos e meteorológicos para obter a curva estimada de consumo de energia em um determinado período. É notório que as condições climáticas influenciam o consumo por eletricidade, já que a principal causa do efeito estufa é a elevação da temperatura que, por sua vez, provoca o aumento do uso de uma vasta gama de tipos e tamanhos de refrigeradores elétricos.

A pesquisa propõe o reconhecimento da tendência de crescimento do consumo de energia do ramo agropecuário paulista, contribuindo com dados para que gestores de políticas públicas e empresas de geração, transmissão e distribuição de eletricidade possam gerar melhores estimativas de investimentos, de forma a dar sustentabilidade ao desenvolvimento no curto, médio e longo prazo.

REFERÊNCIAS

ABDULKAREEM, A.; OKOROAFOR, E.J.; AWELEWA, A.; ADEKITAN, A. Pseudo-inverse matrix model for estimating long-term annual peak electricity demand: The covenant university's experience. *International Journal of Energy Economics and Policy*, 4. ed., Nigéria, v. 9, p. 103-109, maio 2019. DOI: 10.32479/ijeep.7566. Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0->

85067405770&doi=10.32479%2fijeeep.7566&partnerID=40&md5=cf35b54993edabfab6f3fad136fc64d1>. Acesso em: 08 out. 2021.

FAN, C.; XIAO, F.; WANG, S. Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques. *Applied Energy*, Hong Kong, v. 127, p. 1-10, 2014. DOI: 10.1016/j.apenergy.2014.04.016. Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-84899701114&doi=10.1016%2fj.apenergy.2014.04.016&partnerID=40&md5=9384074adcfe1d6622e290ed54d625>>. Acesso em: 08 out. 2021.

MARANGONI, F.; MAGATÃO, L.; DE ARRUDA, L.V.R. Demand response optimization model to energy and power expenses analysis and contract revision. *Energies*, 11. ed., Brasil, v. 13, p. 1-23, jun. 2020. DOI: 10.3390/en13112803. Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85086370091&doi=10.3390%2fen13112803&partnerID=40&md5=45f9bc9a41402f2b673c99d8123bf344>>. Acesso em: 08 out. 2021.

MARTINS, G. de A.; DOMINGUES, O. *Estatística Geral e Aplicada*. 5.ed. rev. e ampl. São Paulo: Atlas, 2014. 416p.

SÃO PAULO (Estado). Secretaria de Infraestrutura e Meio Ambiente. *Balanco Energético do Estado de São Paulo (BEESP) 2021: Ano Base 2020*. São Paulo, 2021. 195p. Disponível em: <<http://dadosenergeticos.energia.sp.gov.br/portalev2/intranet/BiblioVirtual/diversos/BalancoEnergético.pdf>>. Acesso em: 15 março 2022.

TAGHIZADEH-YAZDI, M.; MOHAMMADI-BALANI, A. A mathematical model for multi-region, multi-source, multi-period generation expansion planning in renewable energy for country-wide generation-transmission planning. *Journal of Information Technology Management*, 4. ed., Irã, v. 12, p. 215-231, 2020. DOI: 10.22059/jitm.2020.298258.2476. Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85099813647&doi=10.22059%2fjitm.2020.298258.2476&partnerID=40&md5=6491aae48693ec897d81f15b2ab1c4e0>>. Acesso em: 08 out. 2021.

VU, D.H.; MUTTAQI, K.M.; AGALGAONKAR, A.P. A variance inflation factor and backward elimination based robust regression model for forecasting monthly electricity

demand using climatic variables. *Applied Energy*, Austrália, v. 140, p. 385-394, 2015.

DOI: 10.1016/j.apenergy.2014.12.011. Disponível em:

<[https://www.scopus.com/inward/record.uri?eid=2-s2.0-](https://www.scopus.com/inward/record.uri?eid=2-s2.0-84919935314&doi=10.1016%2fj.apenergy.2014.12.011&partnerID=40&md5=b2876ae6c31888837a79074bc7356a1)

[84919935314&doi=10.1016%2fj.apenergy.2014.12.011&partnerID=40&md5=b2876ae](https://www.scopus.com/inward/record.uri?eid=2-s2.0-84919935314&doi=10.1016%2fj.apenergy.2014.12.011&partnerID=40&md5=b2876ae6c31888837a79074bc7356a1)

[e6c31888837a79074bc7356a1](https://www.scopus.com/inward/record.uri?eid=2-s2.0-84919935314&doi=10.1016%2fj.apenergy.2014.12.011&partnerID=40&md5=b2876ae6c31888837a79074bc7356a1)>. Acesso em: 08 out. 2021.