

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ

BRUNO ADEMIR DIAS DA SILVA

**PREDIÇÃO DE CONSUMO DE ENERGIA ELÉTRICA NO CAMPUS MEDIANEIRA
COM O USO DE MÉTODOS ESTATÍSTICOS**

MEDIANEIRA

2023

BRUNO ADEMIR DIAS DA SILVA

**PREDIÇÃO DE CONSUMO DE ENERGIA ELÉTRICA NO CAMPUS MEDIANEIRA
COM O USO DE MÉTODOS ESTATÍSTICOS**

**Prediction of electricity consumption on the Medianeira campus using statistical
methods**

Trabalho de Conclusão de Curso de Graduação
apresentado como requisito para obtenção do título de
Bacharel em Engenharia Elétrica da Universidade
Tecnológica Federal do Paraná (UTFPR).

Orientador: Alex Lemes Guedes.

MEDIANEIRA

2023



[4.0 Internacional](https://creativecommons.org/licenses/by/4.0/)

Esta licença permite compartilhamento, remixe, adaptação e criação a partir do trabalho, mesmo para fins comerciais, desde que sejam atribuídos créditos ao(s) autor(es). Conteúdos elaborados por terceiros, citados e referenciados nesta obra não são cobertos pela licença.

BRUNO ADEMIR DIAS DA SILVA

**PREDIÇÃO DE CONSUMO DE ENERGIA ELÉTRICA NO CAMPUS MEDIANEIRA
COM USO DE MÉTODOS ESTATÍSTICOS**

Trabalho de Conclusão de Curso de Graduação
apresentado como requisito para obtenção do título de
Bacharel em Engenharia Elétrica da Universidade
Tecnológica Federal do Paraná (UTFPR).

Data de aprovação: 23/junho/2023

Alex Lemes Guedes
Mestrado
Universidade Tecnológica Federal do Paraná

Leandro Antonio Pasa
Doutorado
Universidade Tecnológica Federal do Paraná

Filipe Marangoni
Doutorado
Universidade Tecnológica Federal do Paraná

MEDIANEIRA

2023

Dedico esse trabalho a minha família, por sempre apoiar meus sonhos e a todo o momento medir esforços para que essa minha conquista fosse alcançada junto a um ótimo ensino de qualidade.

AGRADECIMENTOS

Inicialmente estes agradecimentos não conseguira atender a todos que puderam e tiveram de alguma forma passagem em minha vida enquanto universitário. Disto isso, venho me desculpar daqueles de que passaram em branco nestes agradecimentos, todavia saibam que levo todos em minhas lembranças, por fim enfatizar minha total gratidão as pessoas que tiveram diretamente ou indiretamente experiências nesta passagem da minha vida e que todos estarão eternizados.

Agradeço a meus Pais, Ademir da Silva e Eni Dias da Silva, por sempre acreditarem em meus sonhos e apoiarem em todos os projetos que tive em minha vida, amor incondicional em todos os momentos que precisei, as conversas, a todos os bons-dias que foram ditos e seus esforços para eu alcançar os meus objetivos. Serei eternamente grato por vocês e a eles devo absolutamente tudo.

As minhas irmãs e Tia, Ana Carolina Donega Rodrigues Dias, Giovana Maria Dias da Silva e Judite Dias, pois sempre cuidaram de todos as situações que ocorrem em nossas famílias.

Em especial ao meu Orientador Prof. Me. Alex Lemes Guedes, na qual tenho prazer em chamar de grande amigo, por todos os ensinamentos, conselhos, aprendizados, experiências e oportunidades que foi me concedida em minha vida academica e profissional, pois graças a ele nenhuma das minhas conquistas seriam possíveis, serei eternamente grato a o Mentor que foi para mim, e sempre terei uma dívida a qual carregarei eternamente em minha passagem nesta vida.

Aos meus amigos em especial Morita, Dionisio, Felipe, Felipe e Agnaldo, que conviveram comigo em nossa república durante esta etapa da minha vida e aos meus amigos do LAD, Anderson Camargo, Bruno Maihach, Henrique Peron, Jaílson Narcizo, Lucas Rauber, Luan Rancan, Matheus Casa, Matheus Ferreira, Marco Foltran, Nuiquer de Oliver, Paula Moura e Victor Augusto. Por todos os momentos que tivemos nunca esquecerei esta história que construímos.

A todos vocês, OBRIGADO !!

RESUMO

A projeção de consumo é uma previsão de condições futuras com a finalidade de cruzar informações, variáveis e tendências que ajudem a entender os motivos que levam uma instalação a ter um determinado perfil de consumo de energia elétrica. Uma projeção de consumo confiável pode ser uma ferramenta para propor estratégias energéticas a fim de minimizar os gastos excessivos, estando diretamente ligado a projeção de receitas, ajudando a determinar expectativas e evitar prejuízos. No campus universitários, o ambiente de uso de energia é muito disperso fisicamente, com diferentes pessoas e características de uso, de forma que sem um sistema de medição eficaz é difícil identificar ações que possam melhorar a eficiência do uso desse recurso tão necessário, e ao mesmo tempo, estabelecer uma relação com as instituições e seus custos substanciais associados ao orçamento anual. Nessas condições, este trabalho propõe o desenvolvimento do projeto de uma curva de consumo de energia elétrica do campus UTFPR Medianeira com 75% ou mais no grau de predição. Foram utilizados dados de consumos dos anos de 2014 a 2018, suas respectivas temperaturas médias mensais e quantidade de horas de utilização de todas as salas de aulas. Os dados foram separados em dois grupos, sendo os anos 2014 a 2017 utilizados para criar um modelo de regressão, que então foi testado nos dados do ano de 2018. A partir disto, chegou-se a um modelo que foi capaz de prever o consumo universitário com erro absoluto médio de 18% durante os meses do ano de teste, assim como um erro de -1,5% para o consumo anual total.

Palavras-chave: previsão; medição; consumo.

ABSTRACT

The consumption projection is a prediction of future conditions with the purpose of crossing information, variables and trends that help to understand the reasons that lead an installation to have a certain profile of energy consumption. A reliable consumption projection can be a tool to propose energy strategies in order to minimize excessive expenses, being directly linked to the revenue projection, helping to determine expectations and avoid losses. On university campuses, the energy use environment is very physically dispersed, with different people and usage characteristics, so that without an effective measurement system it is difficult to identify actions that can improve the efficiency of the use of this much needed resource, and at the same time at the same time, establish a relationship with the institutions and their substantial costs associated with the annual budget. Under these conditions, this work proposes the development of the project of an electric energy consumption curve of the UTFPR Medianeira campus with 75% or more in the degree of prediction. Consumption data from the years 2014 to 2018, their respective monthly average temperatures and number of hours of use of all classrooms were used. The data were separated into two groups, with the years 2014 to 2017 being used to create a regression model, which was then tested on the data for the year 2018. From this, a model was arrived at that was able to predict consumption university students with an average absolute error of 18% during the months of the test year, as well as an error of -1.5% for the total annual consumption.

Keywords: prediction; measurement; consumption.

LISTA DE ILUSTRAÇÕES

Figura 1 – Gráfico representativa de uma Correlação Positiva	18
Figura 2 – Gráfico representativa de uma Correlação Negativa.....	18
Figura 3 – Gráfico representativa de uma Correlação Nula	18
Figura 4 – Representação de sistema com e sem outliers.....	20
Figura 5 – Regressão Linear.....	21
Figura 6 – Vista Superior UTFPR Campus Medianeira.....	24
Figura 7 - Etapas de desenvolvimento do trabalho.	25
Figura 8 - Representação de um mapa de sala	29
Figura 9 - Representação da tabela de dados extraída do mapa de salas	29
Figura 10 - demonstrativo do calendário acadêmico.....	29
Figura 11 - Mapeamento dos dias letivos conforme calendário acadêmico.	30
Figura 12 - Índice de utilização das salas de aula em horas, para os anos de 2014 a 2017	30
Figura 13 – Gráfico de Dispersão Índice Padronizados	31
Figura 14 – Relação entre Temperatura média padronizada e Consumo Padroni- zado – Modelo 1.....	38
Figura 15 – Relação entre Índice composto e Consumo – Modelo 2.....	38
Figura 16 – Relação entre Temperatura padronizada, Índice geral padronizado e Consumo padronizado – Modelo 3.....	38
Figura 17 – Histograma, Q-Q e Correlação do Modelo 1	39
Figura 18 – Histograma, Q-Q e Correlação do Modelo 2	39
Figura 19 – Histograma, Q-Q e Correlação do Modelo 3	39
Figura 20 – Gráfico de Dispersão dos Resíduos padronizados.	42
Figura 21 – Dados padronizados de consumo de energia entre 2014 e 2017.	43
Figura 22 – Comparação entre consumo real e predito para o ano de 2018.	46

LISTA DE TABELAS

Tabela 1 – Valores Críticos de D para α igual a 5%	16
Tabela 2 – Modelo Propostos.....	32
Tabela 3– Teste de Shapiro-Wilk	33
Tabela 4 – Matriz de Correlação	33
Tabela 5 – Planilha dos dados entre 2014 e 2017	34
Tabela 6 – Planilha de Dados Padronizados.....	35
Tabela 7 – Modelos Propostos.....	37
Tabela 8 – Valores de média e desvio padrão utilizados na padronização	37
Tabela 9 – Dados estatísticos das regressões lineares de cada modelo.	41
Tabela 10 – Valores de média e desvio padrão utilizados na padronização após a retirada dos outliers.	43
Tabela 11 – Dados estatísticos do modelo final.	44
Tabela 12 – Valores reais e previstos de consumo para o ano de 2018, utili- zando o modelo final.	45

LISTA DE ABREVIATURAS E SIGLAS

r^2	Coeficiente de Determinação
r	Coeficiente de Correlação
CP	Consumo Padronizado
MAE	Erro Absoluto Médio
MSE	Erro Quadrático Médio
IC	Índice de Utilização de Sala Composto
IP	Índice de Utilização de Sala Geral Padronizado
KW	Kilowatts
SW	Shapiro-Wilk
TP	Temperatura Média Padronizada
UTFPR	Universidade Tecnológica Federal do Paraná

SUMÁRIO

1	INTRODUÇÃO	13
2	REFERENCIAL TEÓRICO.....	15
2.1	Teste paramétricos e não paramétricos	15
2.1.1	Teste de Normalidade dos Dados	15
2.2	Correlação de variáveis	17
2.3	Multicolinearidade	19
2.4	Outliers	19
2.5	Regressão Linear	20
2.6	Estimação pelo método dos mínimos quadrados	21
2.7	Fator de inflação de Variância.....	22
2.8	Padronização de Dados	22
3	PROCEDIMENTOS METODOLÓGICOS.....	24
3.1	Delimitação do Estudo	24
3.2	Etapas de desenvolvimento	25
3.3	Variáveis independentes	26
3.3.1	Temperatura Ambiente.....	27
3.3.2	Tempo de utilização de salas de aula	27
<u>3.3.2.1</u>	<u>Pesquisa no mapa de salas</u>	<u>28</u>
<u>3.3.2.2</u>	<u>Verificação de dias letivos</u>	<u>29</u>
<u>3.3.2.3</u>	<u>Quantificação de horas de utilização das salas.....</u>	<u>30</u>
<u>3.3.2.4</u>	<u>Índice Geral Padronizado</u>	<u>30</u>
3.3.3	Índice composto	31
3.4	Modelos de Regressão Linear Múltipla	32
3.5	Tratamento dos dados	32
3.5.1	Teste de Shapiro-Wilk	32
3.5.2	Tabela de Correlação de Karl Pearson	33
3.6	Padronização dos Dados	34
3.7	Verificação Erro de Modelo	36
4	RESULTADOS E DISCUSSÕES	37
4.1	Escolha do melhor modelo.....	37
4.2	Modelo final.....	42
4.3	Teste do modelo	44
5	CONCLUSÃO	47

5.1	Trabalhos futuros	47
	REFERÊNCIAS.....	49
	APÊNDICE A - ROTINA NO SOFTWARE.....	51

1 INTRODUÇÃO

A energia elétrica desempenha um papel importante e imprescindível na sociedade atual, sendo a base de funcionamento para a maior parte dos equipamentos tecnológicos existentes, proporcionando assim grande conforto à sociedade. No ano de 2001 o Brasil sofreu com uma grave crise energética, mais conhecido como “apagão”. Segundo TOLMASQUIM (2000), o consumo de energia elétrica aumentou 49% entre 1990 e 2000, enquanto a geração aumentou apenas 35%, de forma que a capacidade de geração brasileira estava abaixo do que se era esperado e necessário.

Diante deste cenário a comunidade científica passou a ter preocupações, mediante as consequências futuras que a sociedade estaria sujeita. Foi necessário exercitar o uso racional e consciente de energia e conceber ideias de eficiência energética para otimizar, projetar e prever futuros aumentos no consumo de energético devido ao crescimento populacional, condições climáticas adversas e gastos desnecessários.

Neste contexto, é importante usar procedimentos e métodos estatísticos como ferramentas análise para entender e enriquecer a pesquisa científica relacionada com o consumo de energia elétrica, tentando descrever e identificar padrões de comportamento ao longo dos anos, bem como analisar as relações presentes nas observações. Além disso, é interessante usar ferramentas ao analisar dados históricos, que permitem apresentar relatórios de acordo com os objetivos a serem alcançados e fornecer conclusões sobre o passado, bem como sobre a possível previsão de valores futuros das variáveis analisadas.

Esse tipo de previsão é especialmente importante para concessionárias de distribuição de energia, empresas que trabalham com produção e órgãos estatais que trabalham com gestão de energia elétrica no país, pois prevê o consumo em função do número de consumidores, temperatura e tempo (SANSON, DUARTE RIBEIRO, *et al.*, 2005). Com o consumo de energia elétrica crescendo acima da capacidade de geração, seu custo foi aumentando com o passar do tempo, impactando diretamente nas questões financeiras da sociedade. Com isso, o uso racional da energia passou a ser uma necessidade, colocando o foco na eficiência energética, que trata da utilização de estudos, equipamentos e até modelos matemáticos que possuem o objetivo de tornar mais consciente o uso das diferentes formas de energias. Além

disso, procura-se ainda o uso de novas fontes, a fim de minimizar o consumo hídrico e térmico, e baratear os valores a quais foram empregados nas contas energéticas a pagar.

Neste contexto, a Universidade Tecnológica Federal do Paraná, campus Medianeira (UTFPR – MD), instituiu a Comissão de Eficiência Energética do Campus Medianeira, que tem, entre seus objetivos, propor uma forma de calcular o consumo de energia elétrica com base em variáveis independentes. Isso permitirá, por exemplo, realizar projeções de consumo, propor soluções para tentar diminuir o consumo energético dentro do campus, podendo também observar os impactos dessas variáveis.

Perante o exposto, este trabalho procura determinar uma forma de prever o consumo de energia elétrica na Universidade Tecnológica Federal do Paraná - UTFPR, campus Medianeira, com o uso de variáveis independentes como temperatura e tempo de aulas por mês, de modo a auxiliar a gestão da universidade a diminuir os gastos com energia elétrica, dar ferramental para otimização do consumo e calcular impactos relacionados com mudanças no regime de operação da universidade.

1.1 Objetivos

Este trabalho tem como objetivo determinar um modelo capaz de prever o consumo de energia elétrica na UTFPR campus Medianeira com grau de predição de no mínimo 75%.

Para atingir o objetivo geral proposto, é determinado os seguintes objetivos específicos:

- Realizar estudos sobre métodos estatísticos de ajuste de curvas;
- Aquisitar dados de consumo de energia elétrica na UTFPR campus Medianeira;
- Determinar as variáveis independentes que apresentam influência sobre o consumo de energia e aquisitar estes dados;
- Desenvolver um modelo de predição através de análises estatísticas.

2 REFERENCIAL TEÓRICO

Nesta seção será realizada uma revisão bibliográfica sobre os conceitos necessários para uma melhor compreensão do trabalho.

2.1 Teste paramétricos e não paramétricos

Segundo CAMPOS (2000), os testes estatísticos podem ser divididos em dois grandes grupos: paramétricos e não paramétricos. Testes paramétricos são avaliados quando há dados contínuos e seus resultados perdem o nível de confiança devido a discrepâncias a quais podem ser submetidos. Em contraste os testes não paramétricos são mais sustentáveis com dados ordenados e menos afetados por discrepâncias em seus dados.

2.1.1 Teste de Normalidade dos Dados

Para realizar a análise de dados é importante averiguar as variáveis quantificadas, com a finalidade de garantir o uso adequado de teste paramétricos ou não paramétricos. Independentemente do objetivo principal de uma aplicação, os testes paramétricos podem fornecer resultados bons e interessantes, úteis para a tomada de decisões. O uso adequado de cada teste, baseado na escolha consciente do *software* de modelagem, deve ser sempre baseado em teorias adequadas.

Um teste possível de ser realizado para verificar a normalidade dos dados é o teste de Shapiro-Wilk. Segundo FÁVERO e BELFIORE (2017), as hipóteses empregadas são:

- H_0 : a amostra provém de uma distribuição normal;
- H_1 : a amostra provém de uma distribuição não normal.

O teste de Shapiro-Wilk é um teste de normalidade publicado em 1965 por Samuel Sanford Shapiro e Martin Wilk e se baseia na regressão dos valores amostrais ordenados com as correspondentes estatísticas de seis ordens normais que, para uma amostra de uma população normalmente distribuída, é linear (ROYSTON, 1981). A Equação (1) mostra o cálculo de um valor D, utilizado no teste de Shapiro-Wilk.

$$D_{s-w} = \frac{b^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (1)$$

Pode-se observar a existência do coeficiente b , que pode ser representado de duas formas: uma para o caso em que n é par e outro caso em que n é ímpar, sendo n o tamanho da população. A seguir, as equações (2) e (3) apresentam as formas de cálculo para n par e ímpar, respectivamente (CANTELI, 2021).

$$b = \sum_{i=1}^{n/2} a_{n-1+1} * (x_{n-i+1} - x_i) \quad (2)$$

$$b = \sum_{i=1}^{(n+1)/2} a_{n-1+1} * (x_{n-i+1} - x_i) \quad (3)$$

Analisando as equações, observa-se que suas diferenças se encontram apenas no limite superior do somatório. As amostras obtidas serão submetidas a este teste com o uso da Equação (1), a partir dos valores das parcelas de diferenças que calculadas através da Equação (2) ou Equação (3). Deste modo o valor selecionado será comparado com o valor tabelado (SHAPIRO e WILK, 1965). A Tabela 1 mostra uma compilação de alguns valores da tabela de Shapiro-Wilk de acordo com o tamanho da população utilizada no teste.

Tabela 1 – Valores Críticos de D para α igual a 5%

População	$\alpha = 0,05\%$
40	0,940
41	0,941
42	0,942
43	0,943
44	0,944
45	0,945
46	0,945
47	0,946
48	0,947
49	0,947
50	0,947

Fonte: Adaptado de Shapiro e Wilk (1965)

Se o valor calculado de D_{S-W} for maior que o tabelado, não se tem evidências para rejeitar a hipótese nula H_0 , e a amostra tem, pelo menos aproximadamente, uma distribuição normal. Caso contrário, há evidências para rejeitar a hipótese nula H_0 , e amostra tem, pelo menos aproximadamente, uma distribuição não normal.

2.2 Correlação de variáveis

Segundo VIRGILITO (2017), a teoria de correlação fundamenta-se na distribuição de probabilidade de um evento e mede o comportamento entre o evento e sua repetição e ou resultado, sendo possível observar visualmente usando um diagrama de dispersão.

Ao analisar a correlação entre variáveis há 4 resultados possíveis: correlação positiva, negativa, nula e espúria, quando as variáveis não são correlacionadas. Neste trabalho, para determinar os coeficientes de correlação será aplicado o método de Coeficiente de Correlação de Karl Pearson. O método mais utilizado para esse estudo é o de determinação do “r” de Pearson, realizando-se o comparativo em pares da amostra estudada.

A origem deste coeficiente remete a uma medida bivariada (força) do grau de relacionamento entre duas variáveis quantitativas com a finalidade de evidenciar influência mútua. Segundo VIRGILITO (2017), Pearson verificou a independência das variáveis e notou-se que o coeficiente de correlação estará sempre contido entre dois valores: $-1 \leq r_{x,y} \leq +1$

Segundo FILHO, DA ROCHA, *et al.*, 2014, a Correlação de Pearson é dada por uma associação linear entre variáveis quantitativas, tendo seu coeficiente estimado através da média e da variância das variáveis conforme a Equação (4).

$$r_{x,y} = \frac{\sum(X * Y) - \frac{(X) * (Y)}{n}}{\sqrt{\left[\sum X^2 - \frac{(\sum X^2)}{n} * \sum Y^2 - \frac{(\sum Y^2)}{n} \right]}} \quad (4)$$

Em que:

n – Número de variáveis observadas;

X – Variável 1 a ser correlacionada;

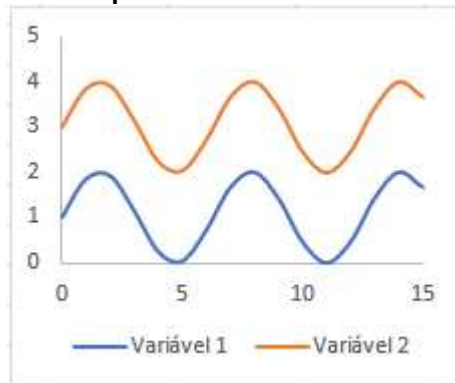
Y – Variável 2 a ser correlacionada.

Se uma variável tende a aumentar enquanto a outra também aumenta, nota-se que a correlação é positiva, de forma que se infere que há relação entre as duas. Na Figura 1 será representado um exemplo desta correlação.

Se uma variável tende a diminuir enquanto a outra aumenta, tem-se uma correlação negativa, de modo que ambas as variáveis se movendo em direções opostas. Na Figura 2 será representado um exemplo desta correlação.

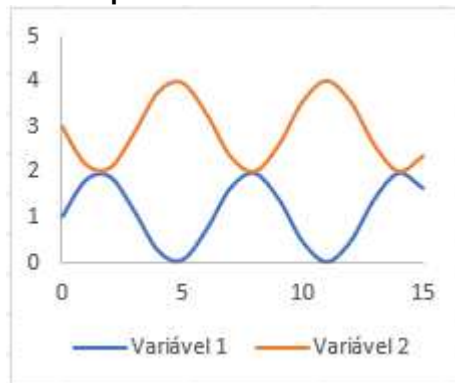
Caso exista uma grande dispersão entre os pontos e ou eles não se dispõem de uma tendência positiva ou negativa, significa que a correlação é igual a 0 ou seja as variáveis não estão linearmente correlacionadas, todavia estas variáveis podem pertencer de uma correlação não-linear. Na Figura 3 será apresentado um exemplo desta correlação.

Figura 1 – Gráfico representativa de uma Correlação Positiva



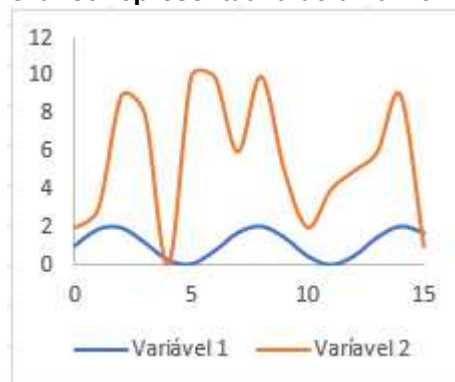
Fonte: Adaptado de Virgilito (2017)

Figura 2 – Gráfico representativa de uma Correlação Negativa



Fonte: Adaptado de Virgilito (2017)

Figura 3 – Gráfico representativa de uma Correlação Nula



Fonte: Adaptado de Virgilito (2017)

2.3 Multicolinearidade

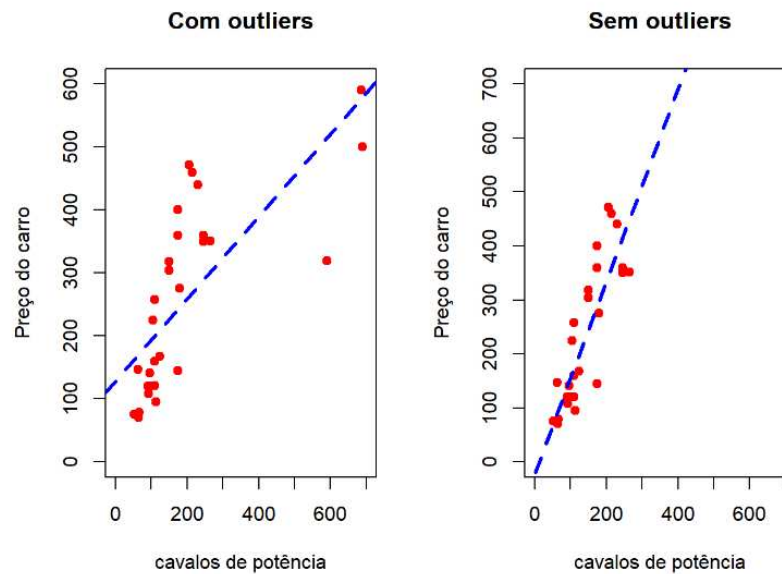
Quando se fala da relação linear (correlação) entre as variáveis explicativas de um processo, espera-se que sua relação seja forte, esse aspecto é muito comum ao se tratar do método de regressão. Essa relação deve ser forte, todavia, nunca perfeita, pois caso o coeficiente de correlação se iguale a 1, torna-se um problema devido a redução da significância estatística destas variáveis (COIMBRA, BENIN, *et al.*, 2005). Medindo o mesmo fenômeno de multicolinearidade, entretanto, ressalva que ele só é existente para casos em que esta relação se trata de mais de duas variáveis.

2.4 Outliers

Trata-se de dados que são significativamente distintos e ou diferentes de outros dados de uma mesma variável. Sendo assim, *outlier*, ou valor discrepante, é uma medida que se desvia da normalidade e pode ocasionar anomalias nos resultados obtidos por algoritmos e sistemas de análise. Segundo RIBAS e VIEIRA (2011) a análise realizada desses pontos se faz separadamente pois em montante significativo em suas amostras os impactos nos resultados esperados serão significativos, pois constituem de observações reais com algum tipo de interferência. A Figura 4 apresenta uma ideia de como *outliers* se apresentam num conjunto de amostras.

Resíduos padronizados, também conhecidos como escores Z são uma medida estatística usada para avaliar o quão longe um valor está da média em termos de desvio padrão. Os resíduos padronizados são calculados subtraindo a média e dividindo pelo desvio padrão dos dados. Ao combinar esses dois conceitos, "*outliers* residuais padronizados" são valores que se destacam como extremos após o cálculo dos resíduos padronizados para um determinado conjunto de dados, ou seja, possuem resíduos padronizados muito altos ou muito baixos em relação a outras observações. A detecção de *outliers* residuais padronizados é útil em muitas aplicações, como análise de regressão, análise de séries temporais ou detecção de anomalias. Ao identificar esses *outliers*, é possível investigar a causa raiz e determinar se são verdadeiros outliers ou simplesmente resultado de erro de medição ou erro aleatório.

Figura 4 – Representação de sistema com e sem outliers.



Fonte: Ross (2020).

Se um conjunto de amostras que contém dados discrepantes for utilizado para calcular uma regressão linear, os valores dos coeficientes podem ser modificados de forma significativa, fazendo com que ela não represente os dados reais de forma correta. Por outro lado, dados aparentemente mais dispersos podem fazer parte do comportamento esperado de um determinado sistema, de forma que se forem considerados discrepantes e retirados do conjunto de amostras, a equação encontrada também não representará de forma fidedigna o sistema original.

2.5 Regressão Linear

Segundo FÁVERO e BELFIORE (2017), o modelo de regressão linear oferece principalmente a possibilidade de examinar a relação entre uma ou mais variáveis explicativas, expostas de forma linear, e uma variável quantitativa dependente. Conseqüentemente um modelo de regressão linear geral pode ser definido da forma descrita na Equação (5).

$$\gamma_i = \alpha + \beta * x_i \quad (5)$$

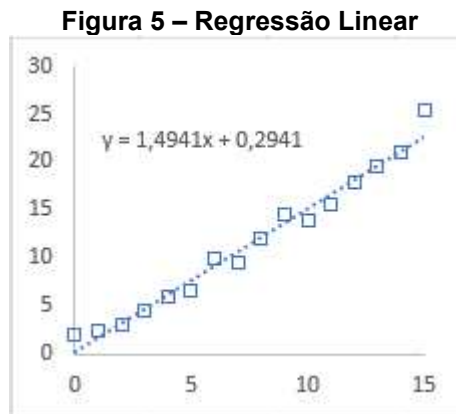
Onde:

γ – Valor Previsto;

α – Coeficiente Linear;

β – Coeficiente Angular;
 x – Variável independente.

A ideia do modelo de regressão linear é estimar uma reta que melhor descreva a relação entre variáveis. Na Figura 5 é apresentado uma representação do resultado obtido a partir da utilização dessa metodologia.



Fonte: Adaptado de Matos (1995)

Quando mais de uma variável independente é utilizada para descrever a variável dependente, tem-se uma regressão linear múltipla (DEVORE, 2005). O objetivo da regressão linear múltipla é estimar os coeficientes das variáveis preditoras para que seja possível fazer previsões sobre a variável dependente com base nos valores das variáveis independentes. Essa técnica é útil para entender como diversas variáveis se relacionam com a variável de interesse e como elas podem ser utilizadas para fazer previsões. O modelo de regressão linear múltipla é apresentado na Equação (6).

$$y = \alpha_0 + \beta_{1x_1} + \beta_{2x_2} + \dots + \beta_{n x_n} \quad (6)$$

Onde:

α_0 é o intercepto;

β_k é o coeficiente do seu predictor correspondente x_k .

2.6 Estimação pelo método dos mínimos quadrados

Trata-se de um método de otimização matemática projetado para encontrar o melhor ajuste para um conjunto de dados, tentando minimizar a soma das diferenças quadradas entre os dados estimados e observados. A ideia básica do método dos

mínimos quadrados é encontrar os coeficientes que minimizam a soma dos quadrados dos resíduos, que são as diferenças entre os valores observados e os valores previstos por um modelo (MEYER, 1965 e MINGOTI, 2005). As Equações

(7) e (8) descrevem como os coeficientes são encontrados através do método de mínimos quadrados (RIBEIRO, 2015).

$$x_0 = \frac{\sum_{i=1}^n y_i \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i y_i) \sum_{i=1}^n x_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \quad (7)$$

E

$$x_1 = \frac{n(\sum_{i=1}^n x_i y_i) - \sum_{i=1}^n y_i \sum_{i=1}^n x_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \quad (8)$$

Essa abordagem pode ser estendida para modelos de regressão mais complexos, envolvendo múltiplas variáveis independentes e termos de interação. O método dos mínimos quadrados é muito flexível e pode ser aplicado a uma ampla gama de problemas de ajuste de modelo.

2.7 Fator de inflação de Variância

A estatística do fator de inflação de variância (VIF) é uma medida usada para avaliar a multicolinearidade em modelos de regressão linear múltipla. A multicolinearidade ocorre quando duas ou mais variáveis independentes em um modelo de regressão são altamente correlacionadas entre si, o que pode afetar a precisão da estimativa dos coeficientes de regressão, o cálculo do VIF é dado pela Equação (9)(BIAGGI, MEDVID e ASSIS, 2017).

$$VIF = \frac{1}{(1 - R^2)} \quad (9)$$

2.8 Padronização de Dados

Trata-se de uma ferramenta com a finalidade de transformar um conjunto de dados para escala com média zero e desvio padrão igual a 1. Isso é feito subtraindo a média dos dados de cada valor e, em seguida, dividindo pelo desvio padrão da amostra conforme mostrado na Equação (10)(OPENINTRO STATISTICS, [s.d.]).

$$P = \frac{X - \mu}{\sigma} \quad (10)$$

Onde:

P é o Valor Padronizado;

X é o Valor da amostra;

μ é a média das amostras;

σ é o Desvio Padrão das amostras.

Quando um conjunto de dados possuem escalas muito distintas entre si, a padronização se torna uma técnica de grande utilidade para análises estatísticas, garantindo que as variáveis independentes estejam na mesma escala, evitando que uma variável tenha um peso maior apenas por estar em escala diferente quando forem sujeitas a regressão linear múltipla.

3 PROCEDIMENTOS METODOLÓGICOS

Este trabalho tem por finalidade apresentar um modelo capaz de projetar a curva de consumo de energia do campus universitário da UTFPR Medianeira, numa tentativa de prever o consumo em função de variáveis independentes. Nas próximas seções serão apresentados a forma como este objetivo será alcançado.

3.1 Delimitação do Estudo

O campus universitário da UTFPR Medianeira possui uma área total de 221.379,89 m², sendo 36.000 m² de área construída, que é utilizada por cerca de 2.200 alunos diariamente (ano de 2019).

A universidade tem seu horário letivo das 7:30 às 23h de segunda a sexta e aos sábados das 7:30 às 17:30h. Até o ano de 2018, ano em que foram realizadas as análises, cinco dos cursos de graduação disponíveis no campus eram em período integral e quatro eram ofertados à noite. Uma vista aérea da universidade pode ser visualizada na Figura 6.

Figura 6 – Vista Superior UTFPR Campus Medianeira



Fonte: Google Maps (2022).

O campus universitário é composto por um total de 17 blocos destinados a salas de aula, laboratórios, biblioteca, área administrativa, restaurante universitário, miniauditório e por fim um auditório.

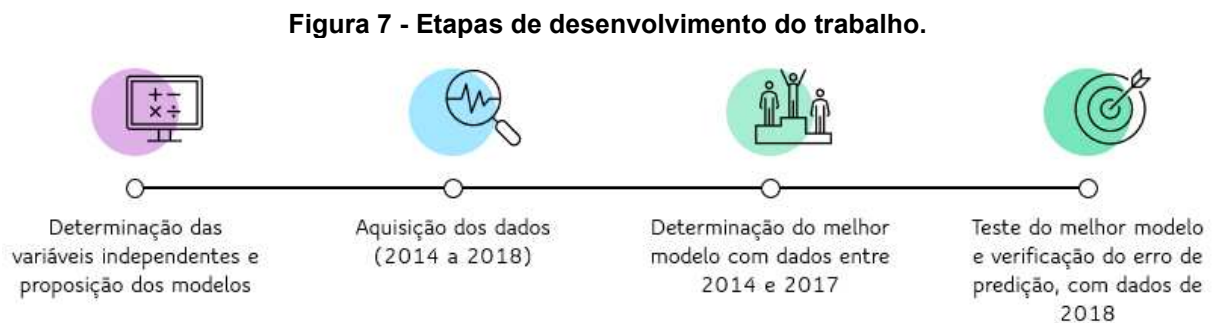
Serão observados dados de consumo de energia elétrica do campus Medianeira desde 2014 até o ano de 2017, fornecidos pela Comissão de Eficiência Energética do próprio campus. O ano de 2018 será utilizado para testar o erro apresentado pela predição.

A periodicidade de todos os dados considerados neste trabalho será mensal, pois os dados de consumo possuem esta periodicidade.

Todas as análises serão realizadas através do software Excel e R, ferramentas de uso comum para produção e desenvolvimento análises estatísticas.

3.2 Etapas de desenvolvimento

A Figura 7 mostra as etapas de desenvolvimento do trabalho.



Fonte: Autoria própria (2023)

Inicialmente é necessário determinar um conjunto de variáveis independentes que possuam correlação positiva ou negativa com o consumo de energia, que é a variável dependente e fruto de estudo deste trabalho. Procura-se uma quantidade reduzida de variáveis independentes, porém que garantam o nível de predição buscado. Com as variáveis determinadas, é então proposto 3 modelos de regressão linear.

Com isso se faz necessário a efetiva aquisição ou geração desses dados. Neste trabalho serão determinados dados relacionados com os semestres letivos entre 2014 e 2018. Com os dados em mãos, eles são tratados por métodos como normalização, ou padronização, e então, os dados entre 2014 e 2017 são aplicados aos 3 modelos propostos para verificar o índice de determinação de cada modelo. O

modelo que apresenta o melhor índice de determinação R^2 são o escolhido para prosseguir o trabalho.

Com o melhor modelo determinado, são identificados *outliers* ou dados discrepantes, que são retirados do conjunto de dados para gerar a equação final que será, por fim, aplicada aos dados do ano de 2018 para verificar o erro de predição.

3.3 Variáveis independentes

A escolha das variáveis independentes que irão compor o estudo é um passo crítico. A escolha incorreta das variáveis independentes pode levar a inviabilização do modelo a ser proposto pela possibilidade de ocorrência de correlações espúrias. Além disso, as variáveis devem ser possíveis de serem adquiridas e replicadas, além de necessariamente terem que possuir algum índice de correlação com o consumo.

Do ponto de vista de uma análise inicial, há muitas variáveis que podem impactar no consumo de energia do campus universitário. A temperatura, por exemplo, possivelmente é uma variável impactante, tanto que é utilizada no trabalho, porém existem várias abordagens possíveis, como temperatura média mensal, temperatura máxima, quantidade de tempo em que o ambiente passou acima de temperaturas consideradas como termicamente confortáveis. Dessa forma, cada trabalho de predição precisa escolher com quais variáveis irá trabalhar numa tentativa de maximizar o acerto do modelo proposto.

Entre as variáveis que inicialmente foram consideradas neste estudo, pode-se citar:

- Temperatura média mensal;
- Temperatura média mensal, calculada a partir das temperaturas máximas diárias;
- Tempo, em horas mensais, que a temperatura ficou acima de 25°C;
- Quantidade de alunos na universidade;
- Quantidade de professores e servidores;
- Quantidade de refeições vendidas no restaurante universitário;
- Semana do semestre (imagina-se que meio e finais de semestres podem apresentar maior consumo, devido à utilização de laboratórios práticos e maior tempo de estudo para as provas)

- Carga elétrica instalada na universidade.

Com a quantidade de variáveis que podem impactar no consumo de energia elétrica da universidade, houve a necessidade de aplicar filtros para ser possível o estudo. Os filtros aplicados foram:

- A variáveis apresentar empiricamente alto índice de correlação com o consumo;
- A variável ter relação direta com o período letivo, visto que são os meses com maior consumo;
- A variável ser possível de ser quantificada.

A seguir são apresentadas as variáveis independentes escolhidas para fazer parte deste estudo.

3.3.1 Temperatura Ambiente

A temperatura ambiente é claramente uma variável que influencia no consumo de energia da universidade, provavelmente com correlação positiva, ou seja, quanto maior a temperatura, maior o consumo. Este comportamento se deve principalmente à elevada carga de climatização presente no campus, de modo que esta variável independente será considerada em todos os modelos propostos.

Os valores de temperatura serão adquiridos através de uma estação Meteorológica oficial do Simepar, localizada na cidade de São Miguel do Iguaçu. Os dados adquiridos são temperaturas médias, mínimas e máximas tomadas a cada hora.

Para este trabalho, foi calculada a temperatura média mensal a partir das temperaturas médias horária adquiridas da estação meteorológica.

3.3.2 Tempo de utilização de salas de aula

Provavelmente o uso das instalações prediais da universidade possui impacto com correlação positiva com o consumo, ou seja, quanto maior o grau de utilização da estrutura predial, mais energia elétrica é demandada. Verifica-se então a necessidade de criar uma metodologia capaz de descrever numericamente a utilização do campus.

Tendo em vista os filtros citados, com destaque ao que determina que será investigado a parte relacionada à tarefa base da UTFPR que é o ensino, estima-se que o consumo de energia elétrica na universidade possua correlação positiva em relação à quantidade de alunos que a frequentam. A cada semestre uma quantidade X de alunos matriculados no campus demandam a criação de Y disciplinas, que por sua vez são ministradas em S salas de aula, ao longo de T aulas por semana. As aulas se repetem a cada semana ao longo do semestre, porém apenas em dias letivos, determinados pelo calendário acadêmico.

Foi considerado que o impacto no consumo é mais pela utilização das salas de aula em si do que pela quantidade de alunos, ou seja, é necessária uma grande variação de alunos para impactar o mesmo que uma pequena variação na quantidade de aulas, visto que turmas podem possuir entre 1 e dezenas de alunos.

Considerando tudo isso, foi proposta a criação de um índice capaz de avaliar a utilização do campus para fins acadêmicos. Este índice será calculado de acordo com a Equação (11).

$$I_{m,t} = \frac{1}{60} \times \sum_{d=\text{segunda}}^{\text{sábado}} \sum_{i=1}^{\text{Total de salas}} (s_i \times N_{i,d,t,m}) \quad (11)$$

Onde:

$I_{m,t}$ é o índice de utilização das instalações da UTFPR no mês m , no turno t . Representa a somatória da quantidade de horas que as salas de aula do campus foram utilizadas naquele turno, no mês específico;

s_i é sala específica sendo avaliada;

$N_{i,d,t}$ é o número de aulas (em minutos) na sala especificada i , no dia da semana d , no turno t , no mês m .

As subseções a seguir descrevem a metodologia para determinação do índice descrito.

3.3.2.1 Pesquisa no mapa de salas

O primeiro passo para a determinação do índice foi uma extensa pesquisa no mapa de salas, encontrado no sistema acadêmico da universidade. No ano de 2018, último ano da pesquisa, o campus possuía 87 salas de aula, de forma que para cada semestre entre o primeiro de 2014 e o segundo de 2018, cada uma dessas salas foi verificada por 10 vezes. A Figura 8 mostra uma representação do mapa de salas

fornecido pelo sistema acadêmico, enquanto a Figura 9 mostra uma representação da tabela de dados extraída de todos os mapas de sala. O tempo considerado foi de 50 minutos para cada aula.

Figura 8 - Representação de um mapa de sala

	Segunda	Terça	Quarta	Quinta	Sexta	Sábado
Manhã						
M1 07:30	CLP		CLP	Automação industrial		
M2 08:20	CLP	Redes	CLP	Automação industrial	Automação industrial	
M3 09:10	CLP	Redes			Automação industrial	
Intervalo Manhã [20 min.] - 10:00						
M4 10:20		Redes				
M5 11:10		Redes				
M6 12:00						

Fonte: Autoria própria (2023)

Figura 9 - Representação da tabela de dados extraída do mapa de salas

Semestre 1 de 2014																	
	Segunda			Terça			Quarta			Quinta			Sexta			Sábado	
	M	T	N	M	T	N	M	T	N	M	T	N	M	T	N	M	T
Sala 1	150	100	250	200	150	150	100	200	200	100	100	200	100	100	0	0	0
Sala 2																	
Sala n																	
Sala 87																	

Fonte: Autoria própria (2023)

3.3.2.2 Verificação de dias letivos

Através do calendário acadêmico da universidade foi determinado, a cada mês, quantos dias letivos foram realizados. A Figura 10 mostra um exemplo do calendário acadêmico da UTFPR para o mês de março de 2014, enquanto a Figura 11 mostra as quantidades de dias letivos para cada dia da semana, naquele mês.

Figura 10 - demonstrativo do calendário acadêmico

MARÇO							2014		
D	S	T	Q	Q	S	S			
						1		3 – Recesso	
2	3	4	5	6	7	8		4 – Carnaval	
9	10	11	12	13	14	15		5 – Recesso - Cinzas	
16	17	18	19	20	21	22		10 – Início do 1º sem. / 2014	
23	24	25	26	27	28	29			
30	31								

Fonte: Autoria própria (2023)

Figura 11 - Mapeamento dos dias letivos conforme calendário acadêmico.

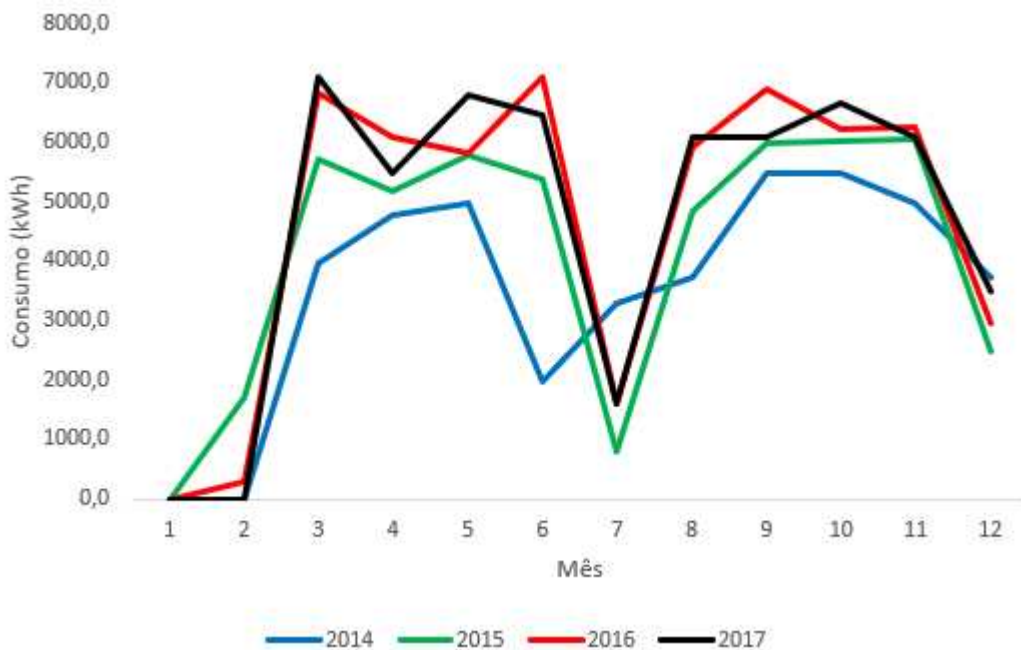
Março de 2014					
Segundas	Terças	Quartas	Quintas	Sextas	Sábados
4	3	3	3	3	3

Fonte: Autoria própria (2023)

3.3.2.3 Quantificação de horas de utilização das salas

Com posse das informações descritas, é possível calcular o tempo de utilização das salas de aula da UTFPR campus Medianeira, por dia e por turno. Como o tempo estava quantificado em minutos, o resultado obtido também estava nesta unidade. Para transformar em horas de utilização, os resultados foram divididos por 60. A Figura 12 mostra o índice calculado, em horas, para os anos de 2014 a 2017, considerando a somatória entre os turnos da manhã, tarde e noite.

Figura 12 - Índice de utilização das salas de aula em horas, para os anos de 2014 a 2017

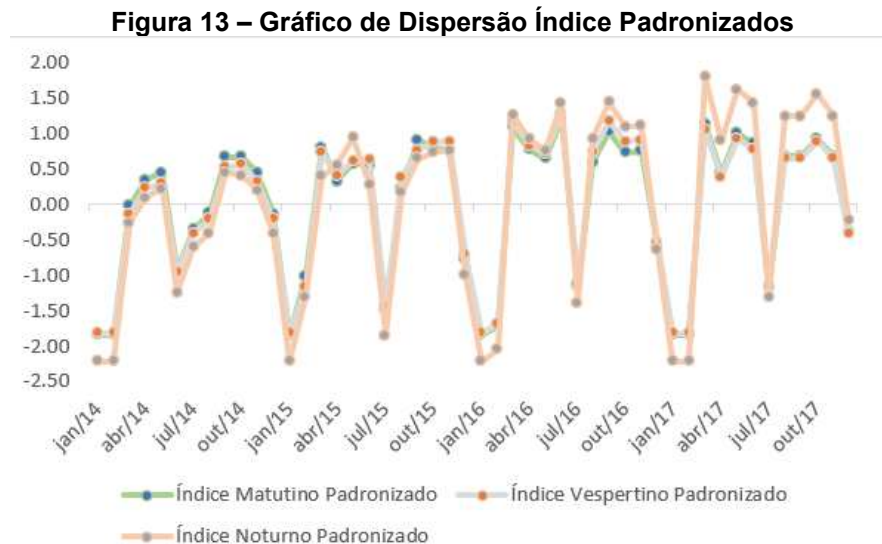


Fonte: Autoria própria (2023)

3.3.2.4 Índice Geral Padronizado

Inicialmente foi definido que seriam realizadas análises de índice por turno, ou seja, o índice de ocupação de salas seria considerado no turno da manhã, tarde e noite. Com o uso de análise gráfica, porém, observou-se que esses índices possuem alto nível de multicolinearidade, como mostrado na Figura 13, ou seja, aumentam e diminuem de forma bastante similar. Com isso a alternativa proposta foi a criação de um índice geral que une os 3 períodos em 1, através da soma simples dos índices

antes da padronização. Dessa forma, será trabalhado com o índice geral, que representa a quantidade total de horas de utilização das salas de aula do campus no mês específico.



Fonte: Autoria própria (2023)

Uma informação interessante que pode ser verificada através deste gráfico é que em 2014 o índice apresenta uma tendência leve de ser menor no período noturno, enquanto no último ano de análise, em 2017, ele se mostra maior que os demais. Isso demonstra que neste período houve um aumento na utilização das salas de aula da universidade no período noturno em relação aos períodos matutinos e vespertinos. Isso pode estar refletindo aumento de alunos ou turmas no período noturno, aumento que não está sendo refletido nos demais.

O aumento de turmas no período noturno, porém, pode impactar de forma considerável o gasto do campus com energia elétrica, visto que os momentos de aula são coincidentes com o horário de pico, entre 18 e 21 h, quando a energia possui custo mais elevado, a sensação térmica ainda é elevada em dias de calor como março e novembro, porém não há geração de energia por usinas solares.

3.3.3 Índice composto

Um dos modelos propostos utiliza um índice composto, calculado utilizando o índice geral, citado anteriormente, e a temperatura, de acordo com a equação (12). As explicações lógicas para este índice advêm do fato de que a temperatura só teria

influência real no consumo de energia caso a sala esteja sendo utilizada, de forma que nesta aproximação, as variáveis não são completamente independentes entre si.

$$I_c = I_{m,t} \times \overline{Temperatura} \quad (12)$$

3.4 Modelos de Regressão Linear Múltipla

De forma a entender as variáveis independentes e como elas se relacionam com o consumo, é proposto a aplicação de 3 tipos de modelos neste projeto. As escolhas destes modelos é formalizada com a seguinte linha de raciocínio: exercer uma verificação individual de cada variável dependente, sendo elas Temperatura e Índice de utilização de sala, avaliando cada caso separadamente e por fim juntá-las em um único modelo. Deste modo será possível observar os impactos individuais de cada uma e os impactos quando ambas estão em conjunto. A Tabela 2 apresenta os modelos considerados neste trabalho.

Tabela 2 – Modelo Propostos

Modelo 1	$Consumo = \alpha_{M1} + \overline{Temperatura} * \beta_{M1}$
Modelo 2	$Consumo = \alpha_{M2} + \overline{Índice Composto} * \beta_{M2}$
Modelo 3	$Consumo = \alpha_{M3} + \overline{Temperatura} * \beta_{1M3} + \overline{Índice Geral} * \beta_{2M3}$

Fonte: Aatoria própria (2023)

3.5 Tratamento dos dados

Para realizar o tratamento e visualização das amostras coletadas nas variáveis escolhidas para o estudo, foi utilizado o *software* Excel para criação de gráficos, tabelas para melhor visualizar os dados. Para as análises estatísticas foi utilizado o *software* R, ferramenta muito utilizada para cálculos estatísticos, com amplo fornecimento de ferramentas para manipulação de dados.

3.5.1 Teste de Shapiro-Wilk

A fim de conhecer os dados, foi aplicado o teste de Shapiro Wilk para todas as variáveis, com o auxílio do *software* R. Os resultados do teste são disponibilizados na Tabela 3.

É possível observar que os valores da estatística do teste, dado na tabela pela letra W, assumem valores próximos a 1. Como mostra a Tabela 1, para um total de amostras de 48 dados, que é caso deste trabalho, o consumo e a temperatura

assumem valor maior que 0,947, indicando distribuição normal. Percebe-se, porém que a variável “Índice Geral” não possui distribuição normal. Considerando que esta variável foi calculada baseada num modelo teórico que leva em conta o tempo previsto de utilização das salas em um semestre, é plausível que ela não tenha distribuição normal devido à baixa quantidade de semestres analisados. A análise por P-Valor confirma os resultados obtidos pelo teste, visto que o limiar para ser considerado distribuição normal é que seja maior que 0,05.

Tabela 3– Teste de Shapiro-Wilk

Consumo	W = 0.96572	P-Valor = 0.1716
Temperatura	W = 0.95343	P-Valor = 0.05486
Índice Geral	W = 0.86092	P-Valor = 0.00004295

Fonte: A autoria própria (2023)

Uma das variáveis não apresentar distribuição normal indica que as análises estatísticas realizadas com estas informações podem ser estatisticamente menos significativas do que poderiam, ou seja, a análise apresentada neste trabalho pode se tornar mais significativa estatisticamente se mais índices semestrais forem incluídos em trabalhos futuros.

3.5.2 Tabela de Correlação de Karl Pearson

Ainda com o objetivo de conhecer os dados, foi realizada uma análise de correlação entre eles através do método de Karl Pearson. O resultado, exibido pela Tabela 4, representa de forma matricial a resposta obtida através do *software* R.

Tabela 4 – Matriz de Correlação

	Consumo	Temperatura	Índice
Consumo	1	0.2161630	0.4729391
Temperatura	0.2161630	1	-0.3459944
Índice	0.4729391	-0.3459944	1

Fonte: A autoria própria (2023)

A Tabela 4 mostra as correlações entre as variáveis no conjunto de dados. O coeficiente de correlação de Pearson mede a relação linear entre duas variáveis contínuas de -1 a 1. Uma correlação de 1 indica uma correlação positiva perfeita, significando que conforme uma variável aumenta, a outra aumenta na mesma proporção. Uma correlação de -1 indica uma correlação negativa total, ou seja,

quando uma variável aumenta a outra variável é reduzida em igual proporção. Uma correlação de 0 sinaliza que não há relação linear entre as variáveis. No caso da tabela de correlação proposta pode-se observar que:

- A variável consumo apresentou correlação positiva moderada com a variável índice geral, com coeficiente de 0,47.
- A variável consumo apresentou correlação positiva fraca com a variável temperatura com coeficiente de 0,22.
- A variável temperatura apresentou correlação negativa moderada com a variável índice com coeficiente de -0,35.

As correlações apresentadas pela variável consumo são plausíveis pois significam que ela aumenta com a temperatura, assim como com o índice de utilização das salas, ou seja, apresenta coerência interna. A variável temperatura possui correlação com a variável índice, porém, provavelmente é um comportamento espúrio, visto que não há relação entre a previsão de utilização de uma sala de aula com a temperatura ambiente.

3.6 Padronização dos Dados

Com apoio e colaboração da Comissão de Eficiência Energética do Campus Medianeira, foi possível adquirir todos os dados necessários para a construção das variáveis indicadas para realização destes estudos. Através das contas de energia do campus foram adquiridos os dados de consumo de energia reais, a secretaria de ensino disponibilizou os calendários letivos dos anos de 2014 a 2018 (necessários para a construção do índice de utilização de sala). A Tabela 5 mostra os dados utilizados para determinação do modelo, entre os anos 2014 e 2017, enquanto a Tabela 6 mostra os mesmos dados padronizados.

Tabela 5 – planilha dos dados entre 2014 e 2017

Mês	Consumo	Temperatura	Índice Geral
jan/14	57069.00	24.12	0.00
fev/14	126104.00	24.10	0.00
mar/14	75724.00	23.73	3968.33
abr/14	92669.00	20.13	4765.00
mai/14	67425.00	17.08	4986.67
jun/14	54947.00	13.18	1992.50
jul/14	52251.00	14.03	3280.00
ago/14	65598.00	17.78	3742.50
set/14	89144.00	18.83	5477.50
out/14	111803.00	22.13	5489.17
nov/14	107720.00	25.78	4986.67

dez/14	104548.00	25.70	3735.83
jan/15	48290.00	25.63	0.00
fev/15	68330.00	24.83	1720.00
mar/15	106965.00	21.15	5717.50
abr/15	83072.00	18.90	5185.83
mai/15	68995.00	17.00	5780.00
jun/15	68308.00	16.25	5372.50
jul/15	42068.00	21.48	801.67
ago/15	63498.00	26.18	4845.83
set/15	94344.00	23.58	5983.33
out/15	100539.00	20.73	6023.33
nov/15	86212.00	21.38	6036.67
dez/15	65532.00	22.48	2472.50
jan/16	36946.00	21.78	0.00
fev/16	55419.00	22.95	298.33
mar/16	92820.00	23.48	6839.17
abr/16	115653.00	16.05	6097.50
mai/16	70661.00	10.80	5818.33
jun/16	62687.00	15.53	7110.83
jul/16	48884.00	20.48	1608.33
ago/16	60045.00	17.65	5902.50
set/16	73624.00	14.33	6888.33
out/16	63245.00	20.25	6227.50
nov/16	101763.00	24.25	6265.83
dez/16	91042.00	25.13	2947.50
jan/17	38113.00	23.58	0.00
fev/17	59717.00	22.58	0.00
mar/17	94889.00	22.85	7091.67
abr/17	97714.00	20.40	5476.67
mai/17	73814.00	16.00	6779.17
jun/17	69418.00	15.33	6443.33
jul/17	58948.00	18.98	1610.83
ago/17	59374.00	21.90	6081.67
set/17	95264.00	22.03	6081.67
out/17	99149.00	20.85	6660.00
nov/17	94079.00	20.43	6081.67
dez/17	108078.00	23.50	3485.83

Fonte: Aútoria pr3pria (2023)

Tabela 6 – Planilha de Dados Padronizados

Mês	Consumo Padronizado (CP)	Temperatura Padronizado (TP)	Índice Geral Padronizado (IP)
jan/14	-0,93	0,960628713	-1,81
fev/14	2,20	0,955218942	-1,81
mar/14	-0,08	0,855138189	-0,12
abr/14	0,69	-0,118620487	0,22
mai/14	-0,46	-0,943610477	0,31
jun/14	-1,02	-1,998515709	-0,96
jul/14	-1,15	-1,768600466	-0,41
ago/14	-0,54	-0,754268512	-0,22
set/14	0,53	-0,470255564	0,52
out/14	1,55	0,422356556	0,53
nov/14	1,37	1,409639658	0,31
dez/14	1,22	1,388000576	-0,22
jan/15	-1,33	1,369066380	-1,81
fev/15	-0,42	1,152675563	-1,08
mar/15	1,33	0,157277805	0,62
abr/15	0,25	-0,451321368	0,40
mai/15	-0,39	-0,965249558	0,65

jun/15	-0,42	-1,168115949	0,48
jul/15	-1,61	0,246539017	-1,47
ago/15	-0,64	1,517835066	0,25
set/15	0,76	0,814564911	0,74
out/15	1,04	0,043672626	0,75
nov/15	0,39	0,219490165	0,76
dez/15	-0,54	0,517027538	-0,76
jan/16	-1,84	0,327685573	-1,81
fev/16	-1,00	0,644157143	-1,69
mar/16	0,69	0,787516059	1,10
abr/16	1,73	-1,222213653	0,79
mai/16	-0,31	-2,642278390	0,67
jun/16	-0,67	-1,362867684	1,22
jul/16	-1,30	-0,023949504	-1,13
ago/16	-0,79	-0,789432020	0,70
set/16	-0,18	-1,687453910	1,12
out/16	-0,65	-0,086161864	0,84
nov/16	1,10	0,995792220	0,86
dez/16	0,61	1,233822119	-0,56
jan/17	-1,79	0,814564911	-1,81
fev/17	-0,81	0,544076390	-1,81
mar/17	0,79	0,617108291	1,21
abr/17	0,91	-0,045588586	0,52
mai/17	-0,17	-1,235738080	1,08
jun/17	-0,37	-1,416965389	0,93
jul/17	-0,84	-0,429682286	-1,13
ago/17	-0,82	0,360144196	0,78
set/17	0,80	0,395307703	0,78
out/17	0,98	0,076131248	1,03
nov/17	0,75	-0,037473931	0,78
dez/17	1,38	0,792925830	-0,33

Fonte: Autoria própria (2023)

3.7 Verificação Erro de Modelo

Serão realizados estudos nos modelos definidos com o objetivo de entender os impactos das variáveis independentes no consumo da UTFPR. Desse modo será analisado dois casos isolados e um caso bivariável, e, mensurando qual modelo apresenta o maior índice de determinação será escolhido o modelo que contém maiores condições de predizer o consumo real.

Após a escolha do modelo, verificação de *outliers* e a determinação da equação de carga, a mesma será aplicada com os dados do ano de 2018 a fim verificar o erro do modelo. Os dados de erro apresentados serão o erro absoluto médio (MAE), visto na Equação (13), e erro absoluto médio percentual.

$$\text{Erro Absoluto Médio} = \frac{1}{n} * \sum_{1}^n (|y - \hat{y}|) \quad (13)$$

4 RESULTADOS E DISCUSSÕES

Nas próximas sessões serão explanados os modelos e análises realizadas para validação do projeto proposto.

4.1 Escolha do melhor modelo

Nesta seção serão apresentados os 3 modelos de regressão com a finalidade de entender quais os impactos das variáveis independentes no consumo de energia do campus e qual modelo tem maior potencial de apresentar o melhor ajuste ao consumo. Os resultados foram obtidos através do *software* R, relacionados com os modelos de regressão já apresentados na Tabela 2, porém apresentados novamente na Tabela 7 para facilitar a leitura.

Tabela 7 – Modelos Propostos

Modelo 1	$Consumo = \alpha_{M1} + Temperatura * \beta_{M1}$
Modelo 2	$Consumo = \alpha_{M2} + Índice Composto * \beta_{M2}$
Modelo 3	$Consumo = \alpha_{M3} + Temperatura * \beta_{1M3} + Índice Geral * \beta_{2M3}$

Fonte: Aatoria própria (2023)

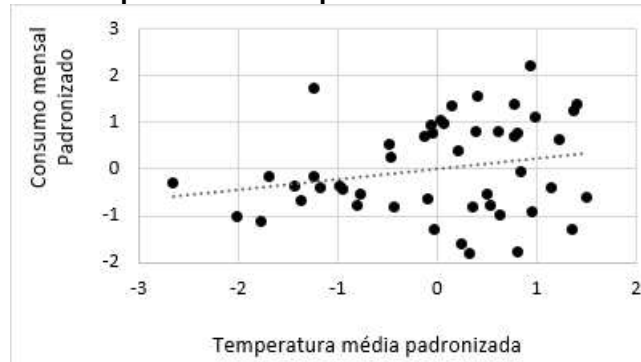
A Figura 14, Figura 15 e Figura 16 representam os dados padronizados de acordo com os modelos. É importante salientar que as figuras apresentam apenas uma mostra visual dos dados relacionados com os modelos propostos. Os valores de média e desvio padrão encontrados, úteis para a padronização e a despadronização posterior, são apresentados Tabela 8.

Tabela 8 – Valores de média e desvio padrão utilizados na padronização

	Média	Desvio padrão
Consumo	77552,10417	22062,7647
Temperatura	20,56854167	3,697014555
Índice geral	4253,333125	2345,699159
Índice composto	84488,52021	48316,20208

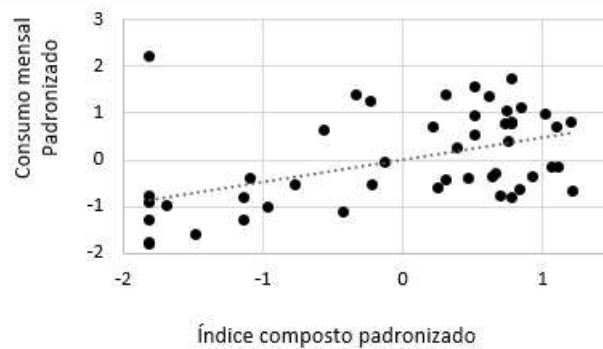
Fonte: Aatoria própria (2023)

Figura 14 – Relação entre Temperatura média padronizada e Consumo Padronizado – Modelo 1



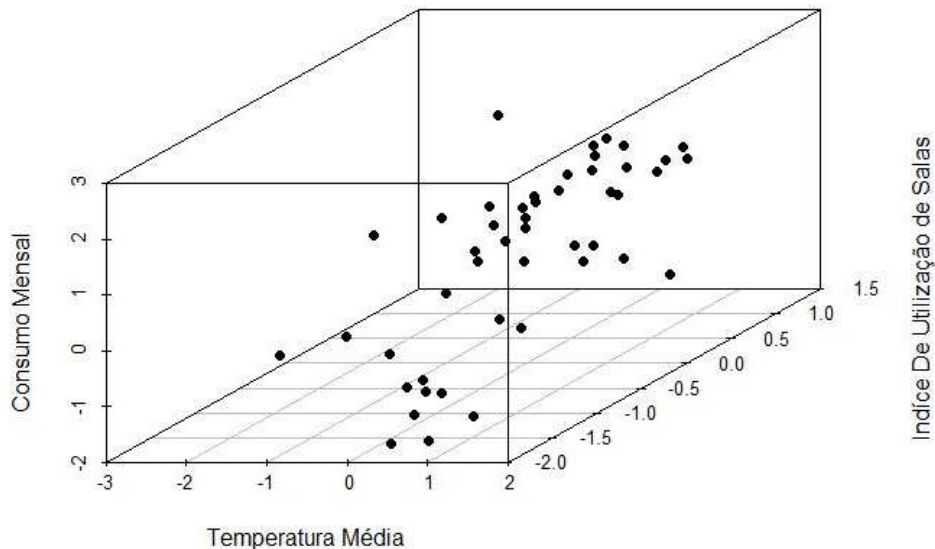
Fonte: Autoria própria (2023)

Figura 15 – Relação entre Índice composto e Consumo – Modelo 2



Fonte: Autoria própria (2023)

Figura 16 – Relação entre Temperatura padronizada, Índice geral padronizado e Consumo padronizado – Modelo 3

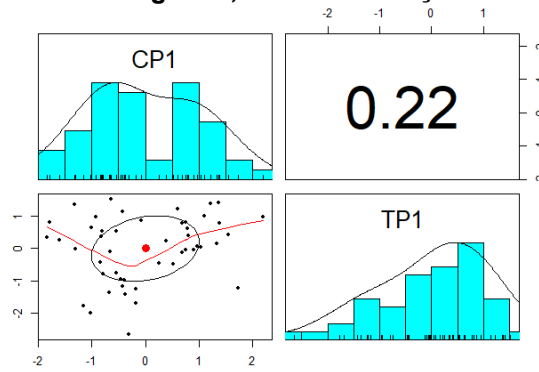


Fonte: Autoria própria (2023)

Utilizando o *software* R, foram feitos os gráficos apresentados na Figura 17, Figura 18 e na Figura 19. Apresentados de forma comparativa, as figuras mostram os histogramas da variável dependente (consumo padronizado CP), o índice de

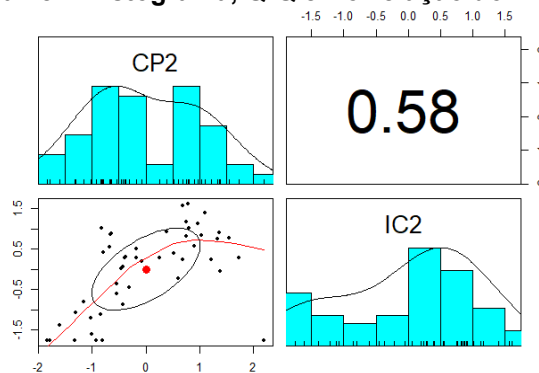
correlação entre as variáveis, gráfico quartil-quartil e por fim histograma da variável independente (Temperatura padronizada – TP, índice composto padronizado – IC, índice geral padronizado – IP).

Figura 17 – Histograma, Q-Q e Correlação do Modelo 1



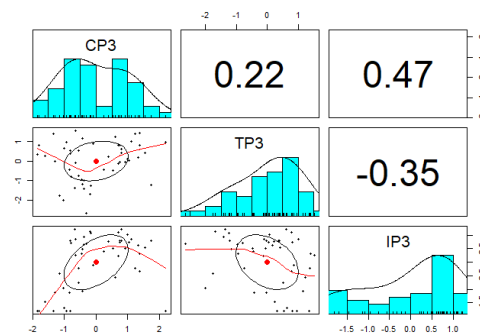
Fonte: Autoria própria (2023)

Figura 18 – Histograma, Q-Q e Correlação do Modelo 2



Fonte: Autoria própria (2023)

Figura 19 – Histograma, Q-Q e Correlação do Modelo 3



Fonte: Autoria própria (2023)

É possível observar nos histogramas que o consumo (CP) possui basicamente dois valores onde possuem maior frequência de ocorrência, ou em outras palavras, dois pontos de máximo. Possivelmente isso ocorre devido ao comportamento sazonal da universidade e à sua natureza letiva, ou seja, durante cerca de 40% do tempo não há aula, causando um comportamento de menor consumo

de energia, enquanto em cerca de 60% do tempo há aula, causando outro perfil de consumo. Esse comportamento binário pode ser o causador do formato apresentado no histograma de consumo.

O histograma de temperatura (TP) indica o clima com temperatura de maior ocorrência entre 20 e 24°C, porém, apesar de possuir frequência considerável de temperaturas mais baixas (menores que 10°C), há uma tendência para aquecimento, ou seja, a maior parte absoluta do tempo a temperatura permanece próxima ou acima de 20°C.

Nos gráficos Q – Q (Quartil – Quartil) de resíduos, pode-se observar que no caso do modelo 1 há muitos pontos extremos e pouca distribuição uniforme deles, caracterizando a dificuldade de se realizar um ajuste significativo neste modelo. É perceptível uma melhora no modelo 2, porém ainda pontos dispersos no modelo 3. Isso pode indicar também a ocorrência de *outliers* no conjunto de amostras.

A Tabela 9 apresenta os dados estatísticos obtidos através do *software* R ao determinar as curvas de regressão linear para cada um dos modelos apresentados. Há três grandes grupos de dados fornecidos pelo programa: o primeiro relacionado com os resíduos, ou seja, dados estatísticos relacionados com a diferença entre os valores da regressão e os dados. O segundo grupo de informações tem relação com a regressão linear em si, ou seja, fornece os coeficientes de intercepto e angular para as variáveis do modelo. O último grupo tem relação com a qualidade do ajuste encontrado, fornecendo variáveis importantes como o valor de R^2 , que indica o quanto o modelo de regressão é capaz de se ajustar aos dados reais.

Percebe-se que os modelos apresentam uma ordem crescente na sua capacidade de ajuste, ou seja, no primeiro modelo, com avaliação apenas da temperatura média, o valor do R^2 é de 4,6%, ou seja, estima-se que em dados sem tratamentos prévios, como a retirada de *outliers*, o modelo considerando apenas a temperatura pode explicar 4,6% da variância do consumo de energia. No modelo 2, onde se considera o índice composto por temperatura multiplicado pela taxa de utilização de salas, o coeficiente de determinação indica que o modelo é capaz de determinar 33,2% da variância do consumo. No modelo 3, onde se considera a temperatura e o índice geral de utilização de salas como variáveis independentes, o coeficiente de determinação indica que o modelo é capaz de determinar aproximadamente 39% da variância do consumo.

Tabela 9 – Dados estatísticos das regressões lineares de cada modelo.

		Modelo 1	Modelo 2	Modelo 3
Resíduos	Resíduo mínimo	-1,9636	-1,4020	-1,4647
	1 quartil	-0,6879	-0,4850	-0,3461
	Média	0,0177	-0,1389	-0,1377
	3 quartil	0,7274	0,3112	0,4150
	Resíduo Máximo	1,9942	3,2084	2,9155
Coeficiente intercepto	Estimado	0	0	-0,0001929
	Erro padrão residual	0,1439	0,1204	0,1166
	T_Valor	0	0	-0,002
	Pr(> t)	1,00	1,00	0,999
Coeficiente angular variável 1	α	0,2157	0,5762	0,4314
	Erro Padrão residual	0,1439	0,1203	0,1243
	T_Valor	1,502	4,789	3,470
	Pr(> t)	0,14	0,0000178	0,00116
Coeficiente angular variável 2	α			0,6225
	Erro Padrão residual			0,1244
	T_Valor			5,004
	Pr(> t)			0,0000
Dados estatísticos	Erro Padrão residual	0,9973	0,8343	0,8082
	R^2	0,04673	0,3327	0,3875
	R^2 ajustado	0,026	0,3182	0,3603
	F-Estatística	2,255	22,94	14,24
	P-Valor	0,14	0,0000177	0,0000169

Fonte: Autoria própria (2023)

Desta forma, foi considerado que o melhor modelo, que será utilizado para continuar com as análises, será o modelo 3, apresentado na sua forma padronizada na Equação (14).

$$\text{Consumo} = -0.000192 + 0.431202 * \overline{\text{Temperatura}} + 0.622522 * \overline{\text{Índice Geral}} \quad (14)$$

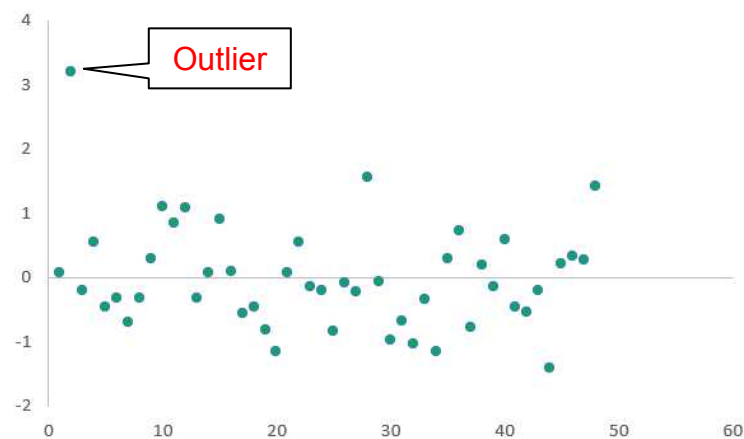
4.2 Modelo final

Antes de determinar o modelo final, é importante verificar a ocorrência de dados discrepantes.

Uma escolha inadequada do que é considerado um *outlier* pode levar a conclusões imprecisas e até mesmo errôneas. Se eles são incluídos nos dados, porém, as análises estatísticas podem ser desviadas. Neste trabalho considerou-se dois tipos de dados discrepantes: estatístico e teórico.

Na análise estatística, considerou-se que um erro residual maior ± 3 vezes o desvio padrão poderia ser considerado um *outlier*, visto que nesta janela há 99,7% da probabilidade da amostra. Dessa forma, a Figura 20 apresenta o erro, ou resíduo, do consumo predito pelo Modelo 3 em relação ao consumo real, padronizado. Destaca-se um dado, que foi considerado como discrepante por esta análise.

Figura 20 – Gráfico de Dispersão dos Resíduos padronizados.



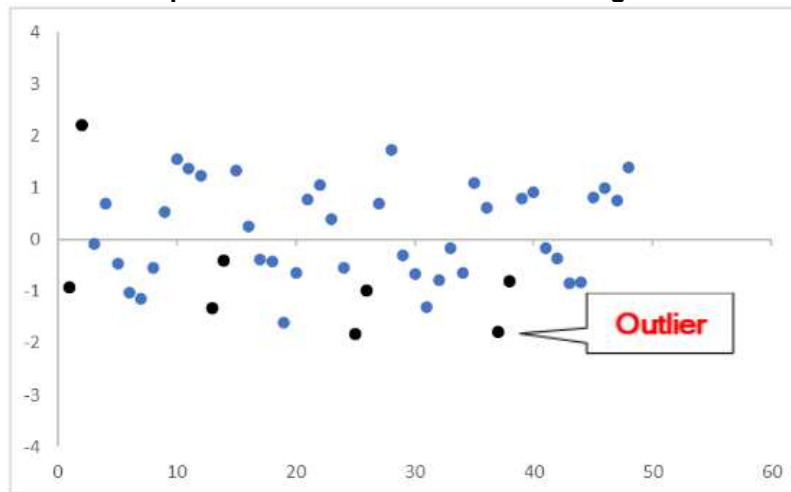
Fonte: Autoria própria (2023)

O dado considerado como discrepante é do mês de fevereiro de 2014, que apresentou um dos maiores valores de consumo da série histórica. A partir dos calendários letivos obtidos, foi detectado que o mês em questão não está no calendário acadêmico de 2014, de forma que é desconhecido o comportamento letivo daquele mês com os dados disponíveis. Servidores da universidade relatam que naquele mês houve a finalização do semestre 2 de 2013, atrasado devido à uma greve geral ocorrida no ano de 2012.

A identificação do *outlier* estatístico descrito levou à consideração que meses sem aula, ou com dados acadêmicos imprecisos, podem influenciar o modelo de forma significativa, visto que nesses casos o consumo provavelmente é determinado

majoritariamente por variáveis não consideradas neste estudo, como a ocorrência de manutenções e construções prediais e o consumo relacionado com os servidores da universidade. Desta forma, os meses maioritariamente sem aula (janeiro e fevereiro) foram retirados da base de dados para levantamento do modelo final. Por este critério, os meses de janeiro e fevereiro de 2014, que apresentam comportamento letivo desconhecido, também foram retirados da base de dados. A Figura 21 mostra todos os dados padronizados de consumo de energia do campus entre os anos de 2014 e 2017, mostrando quais foram considerados discrepantes para a determinação da regressão.

Figura 21 – Dados padronizados de consumo de energia entre 2014 e 2017.



Fonte: Autoria própria (2023)

Com a retirada dos dados discrepantes, os novos valores de médias e desvios padrões utilizados para padronização dos dados são fornecidos na Tabela 10. Como pode ser percebido, após este passo as médias foram modificadas de forma significativa, enquanto os valores dos desvios padrão diminuíram.

Tabela 10 – Valores de média e desvio padrão utilizados na padronização após a retirada dos outliers.

	Média	Desvio padrão
Consumo	80812,825	19499,37829
Temperatura	19,943	3,712450808
Índice geral	5053,5415	1642,348912
Índice composto	100147,3655	35931,22977

Fonte: Autoria própria (2023)

Após todos os ajustes, retirada de pontos considerados discrepantes, obteve-se o modelo final padronizado, mostrado na Equação (15).

$$\text{Consumo} = 0.037364 + 0.43060 * \overline{\text{Temperatura}} + 0.59499 * \text{Índice Geral} \quad (15)$$

A Tabela 11 mostra os dados estatísticos do ajuste final em relação aos anteriores.

Tabela 11 – Dados estatísticos do modelo final.

		Modelo 1	Modelo 2	Modelo 3	Modelo final
Dados estatísticos	Erro Padrão residual	0,9973	0,8343	0,8082	0,7039
	R^2	0,04673	0,3327	0,3875	0,4132
	R^2 ajustado	0,026	0,3182	0,3603	0,3816
	F-estatística	2,255	22,94	14,24	13,02
	P-Valor	0,14	0,0000177	0,0000169	0,00005219

Fonte: Autoria própria (2023)

Com estes resultados é possível observar que o valor do Erro Padrão Residual diminuiu e o nível de precisão do modelo de regressão passa a ser maior. O coeficiente de determinação (R^2) indica que cerca de 41,32% da variabilidade do consumo pode ser explicada pelo índice geral e pela temperatura.

4.3 Teste do modelo

Com a determinação do modelo final, aplicou-se então ao ano de 2018, conforme mostra a Tabela 12, onde é disponibilizado o consumo real em comparação com o consumo previsto pelo modelo, assim como os erros. A Figura 22 mostra visualmente a comparação entre os valores reais e estimados para o consumo.

Através da Tabela 12, percebe-se que em médias anuais a predição do modelo acertou 98,5% do valor real consumido pelo campus, ou seja, subestimou o consumo em apenas 1,5%. Na verificação do erro absoluto médio, porém, verifica-se que o modelo errou 18,7%, ou aproximadamente 15000 kW em média. Este nível de acerto, porém, está dentro do objetivo inicial, que era obter um grau de predição mínimo de 75%.

Os meses de janeiro, junho, julho, agosto e dezembro apresentam alto erro de predição, sugerindo que há variáveis que influenciam nesses meses, porém não foram consideradas no modelo. Entre possíveis variáveis, cita-se que em janeiro e julho há férias docentes, o mês de fevereiro, julho e início de agosto normalmente são meses aproveitados para manutenções no campus, os servidores também possuem

um índice de ocupação que foi ignorado neste trabalho, e também pode-se considerar que em finais de semestre há uma tendência maior de ocupação de salas, com prejuízo para dezembro, que é período de calor e portanto há utilização massiva de ares condicionados de forma não prevista nos sistemas acadêmicos.

Tabela 12 – Valores reais e previstos de consumo para o ano de 2018, utilizando o modelo final.

	Consumo real [kW]	Consumo previsto [kW]	Erro [kW]	Erro absoluto [kW]	Erro absoluto percentual
jan/18	37279	58297	21018	21018,16168	0,56
fev/18	67417	57438	-9979	9979,28251	0,15
mar/18	99354	99851	497	497,3827666	0,01
abr/18	116040	99089	-16951	16950,55302	0,15
mai/18	100684	89412	-11272	11271,80491	0,11
jun/18	64147	89620	25473	25472,91693	0,40
jul/18	60077	47155	-12922	12922,24403	0,22
ago/18	55705	77547	21842	21842,28387	0,39
set/18	74616	88365	13749	13748,91075	0,18
out/18	79682	75805	-3877	3876,895952	0,05
nov/18	98557	87756	-10801	10800,84277	0,11
dez/18	108456	76940	-31516	31515,99958	0,29
Valores médios	80168	78940	-1228	14991	
Valores percentuais em relação ao consumo médio real		98,5	-1,5%	18,7 %	

Fonte: Autoria própria (2023)

Através da análise dos coeficientes da Equação padronizada (15), percebe-se que o índice geral (quantidade de horas que as salas de aula da universidade são ocupadas por mês) possui uma influência no consumo mais significativa que a temperatura, de forma que se considera que a criação deste índice é uma contribuição importante deste trabalho para projetos que no futuro trabalhem na parte de predição de consumo de uma instalação.

Como a definição das aulas e as salas a serem utilizadas é um processo realizado no início do semestre, com previsões da temperatura média mensal da região é possível prever o consumo mensal da universidade, permitindo um planejamento estratégico da administração do campus.

Figura 22 – Comparação entre consumo real e predito para o ano de 2018.

Fonte: Autoria própria (2023)

Um maior controle da utilização das salas de aula, assim como conscientização das pessoas sobre o uso racional do ar-condicionado poderia melhorar ainda mais a predição dos modelos de curva de consumo

Uma informação interessante que é possível retirar a partir do termo independente que aparece no modelo de predição, é que o campus possui um consumo de energia básico de manutenção, que em 0°C seria de 737kW, enquanto em 20°C (temperatura média na região) seria de aproximadamente 46.000 kW por mês. Este valor engloba, possivelmente, a energia consumida pelos servidores de Internet, ar-condicionados que ficam acionados de forma contínua, assim como a influência do trabalho dos servidores. Observa-se, porém, que essa parcela é quase tão significativa quanto o consumo derivado das aulas. Dessa forma, para a elaboração de um plano de ação para diminuição do consumo na universidade, indica-se estudar melhor esta parcela básica, que influencia quase metade do consumo médio da universidade.

5 CONCLUSÃO

Este projeto e seu desenvolvimento têm como concepção o trabalho com métodos estatísticos para encontrar uma equação que seja capaz de prever o consumo de Energia Elétrica no campus da UTFPR Medianeira com pelo menos 75% acerto. Isso permite uma ferramenta capaz de fornecer informações que poderão ser utilizadas para minimizar os gastos energéticos excessivos, atendendo a uma melhor eficiência, planejamento e gestão da energia, pois ela está diretamente ligada a questões financeiras tendo grandes impactos nos orçamentos da universidade.

Coletou-se então informações de consumo de energia elétrica nos anos de 2014 a 2018, assim como dados de temperatura média mensal e um índice de ocupação das salas de aula da universidade nos períodos letivos. Com essas informações foram propostos 3 modelos de regressão linear, sendo dois deles univariados e um bivariado, com o objetivo de prever o consumo de energia baseando-se em grandezas independentes.

Com a construção dos modelos e a verificação de seus coeficientes de determinação, foi possível determinar que o modelo 3, que teve como variáveis independentes a temperatura média mensal e o índice de ocupação de salas como o melhor modelo, com erro absoluto médio de aproximadamente 18% quando aplicado ao ano de 2018, que não havia feito parte do ajuste do modelo.

Esta taxa de acerto é bastante promissora e mostra a importância dos termos independentes utilizados, assim como mostra um caminho para possíveis trabalhos futuros, que podem considerar variáveis não quantificadas neste trabalho, assim como utilizar outras metodologias de predição.

5.1 Trabalhos futuros

Devido as condições climáticas na região onde se localiza o Campus da UTFPR Medianeira, o faz apresentar o maior consumo energético por aluno levando em consideração os outros 12 campus da universidade.

Com isso é necessário, um levantamento não linear sobre a utilização das instalações do campus, também deve-se levar em consideração o tempo médio de permanência seja ele de servidores e alunos, até mesmo para a população que frequenta de alguma forma o campus. Isso é essencial para que os indicadores de eficiência energética e até mesmo predições futuras apresentem resultados mais

coerentes à realidade tendo em vista o comportamento de uso das unidades e de seus equipamentos.

Deste modo, para trabalhos futuros destaca-se a possibilidade da utilização de métodos não lineares para análise, utilizando-se de ferramentas tais como redes neurais, observação de novas variáveis impactantes e a colocação dos demais ambientes do campus, visando entender todas as causas possíveis e influências dentro do consumo.

REFERÊNCIAS

- CAMPOS, G. M. **Estatística Prática para Docetes e Pós-Graduando**. São paulo: [s.n.], 2000. Cap. 14. Disponível em: https://www.forp.usp.br/restauradora/gmc/gmc_livro/gmc_livro_cap14.html. Acesso em: 06 Março 2023.
- CANTELI, A. Teste de Shapiro-Wilk. **Site Github**, 2021. Disponível em: <https://andersonmdcanteli.github.io/Shapiro-Wilk/#portal-action>. Acesso em: 22 outubro 2022.
- COIMBRA, J. L. *et al.* Consequências da multicolinearidade sobre a análise de trilha em canola, Santa Maria, v. 35, p. 347-352, Março 2005. Disponível em: <https://www.scielo.br/j/cr/a/3fNHMZRFxTDgY8zNhLfsGKp/?format=pdf&lang=pt>. Acesso em: 10 Janeiro 2023.
- DEVORE, J. *Probalidade e Estatísticas: para Engenharia e Ciências*, California, 2005. Disponível em: https://www.academia.edu/38884246/Probalidade_e_Estatística_para_Engenharia_e_Ciências_6th_Jay_L_Devore_Cengage_2006_. Acesso em: 27 Agosto 2022.
- FÁVERO, L. P.; BELFIORE, P. **Manual de Análise de dados**. Rio de Janeiro: Elsevier, 2017.
- FILHO, D. B. *et al.* Desvendando os mistérios do Coeficiente de Correlação de Pearson: O retorno. **Cadernos de Pesquisa Política**, p. 66-95, Abril 2014. Disponível em: <https://www.revistas.usp.br/leviathan/article/view/132346>. Acesso em: 03 agosto 2022.
- MATOS, M. A. *Manual Operacional para a Regressão Linear*, 1995. 0-26. Disponível em: <https://web.fe.up.pt/~mam/regressao.pdf>. Acesso em: 07 Janeiro 2023.
- MEYER, P. L. **PROBABILIDADE Aplicações à Estatística**. Washington: [S.n.], 1965. Disponível em: https://www.academia.edu/29243535/Paul_L_Meyer_Probabilidade_Aplicaes_Estats_tica. Acesso em: 04 Outubro 2022.
- MINGOTI, S. A. **Análise de Dados através de Métodos de Estatística Multivariada**. Belo Horizonte: [S.n.], 2005. Acesso em: 13 Setembro 2022.
- RIBAS, J. P.; VIEIRA, P. R. D. C. **ANÁLISE MULTIVARIADA COM O USO DE SPSS**. [S.I.]: Ciencia Moderna, 2011. 0 - 272 p. Acesso em: 13 out. 2022.
- RIBEIRO, D. O MÉTODOS DOS MÍNIMOS QUADRADOS: ESTUDO E APLICAÇÕES PARA O ENSINO MÉDIO, RIO DE JANEIRO, MAIO 2015. 0-69. Disponível em: <https://uenf.br/posgraduacao/matematica/wp-content/uploads/sites/14/2017/09/28052015Renato-Neves-de-Almeida.pdf>. Acesso em: 04 mar. 2023.
- ROSS, S. D. **Manual de Análise de Dados**. [S.I.]: [S.n.], 2020. Disponível em: <https://livro.metodosquantitativos.com/docs/>. Acesso em: 25 Março 2022.

ROYSTON, J. P. An Extension of Shapiro and Wilk's W Test for Normality to Large Samples, 1981. 115-124. Disponível em:
<https://www.jstor.org/stable/2347973?seq=4>. Acesso em: 15 maio 2023.

SANSON, F. F. *et al.* Previsão de demanda por energia elétrica - Método e aplicação, 01 novembro 2005. Disponível em:
<https://www.lume.ufrgs.br/handle/10183/10443>. Acesso em: 05 Março 2023.

SHAPIRO, S.; WILK, M. **An analysis of Variance Test of Normality**. [S.l.]: [S.n.], v. 52, 1965. 591-611 p. Disponível em:
https://www.jstor.org/stable/pdf/2333709.pdf?refreqid=excelsior%3A101d2e01244e27e4c0d5c08fb04eddd4&ab_segments=&origin=&initiator=&acceptTC=1. Acesso em: 22 Novembro 2022.

TOLMASQUIM, M. **As origens da crise**. Rio de Janeiro. 2000.

VIRGILITO, S. B. **Estatística Aplicada**. [S.l.]: [S.n.], 2017. Disponível em:
<https://integrada.minhabiblioteca.com.br/reader/books/9788547214753/pageid/9>. Acesso em: 12 Agosto 2022.

APÊNDICE A - ROTINA NO SOFTWARE

ROTINA SOFTWARE R

```

# Bibliotecas usadas para análises estatísticas
library(psych)
library(car)
library(outliers)
library(readxl)
library(scatterplot3d)

# Criando uma estrutura de DADOS com 3 colunas

#Modelo ~ Consumo = Temperatura*B1
CP1 <- c(-0.93,2.20,-0.08,0.69,-0.46,-1.02,-1.15,-0.54,0.53,1.55,1.37,1.22,-1.33,-
0.42,1.33,0.25,-0.39,-0.42,-1.61,-0.64,0.76,1.04,0.39,-0.54,-1.84,-1.00,0.69,1.73,-0.31,-0.67,-1.30,-
0.79,-0.18,-0.65,1.10,0.61,-1.79,-0.81,0.79,0.91,-0.17,-0.37,-0.84,-0.82,0.80,0.98,0.75,1.38)
TP1 <- c(0.96,0.96,0.86,-0.12,-0.94,-2.00,-1.77,-0.75,-0.47,0.42,1.41,1.39,1.37,1.15,0.16,-
0.45,-0.97,-1.17,0.25,1.52,0.81,0.04,0.22,0.52,0.33,0.64,0.79,-1.22,-2.64,-1.36,-0.02,-0.79,-1.69,-
0.09,1.00,1.23,0.81,0.54,0.62,-0.05,-1.24,-1.42,-0.43,0.36,0.40,0.08,-0.04,0.79)

#Modelo ~ Consumo = IndiceComposto*B1
CP2 <- c(-0.93,2.20,-0.08,0.69,-0.46,-1.02,-1.15,-0.54,0.53,1.55,1.37,1.22,-1.33,-
0.42,1.33,0.25,-0.39,-0.42,-1.61,-0.64,0.76,1.04,0.39,-0.54,-1.84,-1.00,0.69,1.73,-0.31,-0.67,-1.30,-
0.79,-0.18,-0.65,1.10,0.61,-1.79,-0.81,0.79,0.91,-0.17,-0.37,-0.84,-0.82,0.80,0.98,0.75,1.38)
IC2 <- c(-1.75,-1.75,0.20,0.24,0.01,-1.21,-0.80,-0.37,0.39,0.77,0.91,0.24,-1.75,-
0.86,0.75,0.28,0.29,0.06,-1.39,0.88,1.17,0.84,0.92,-0.60,-1.75,-1.61,1.57,0.28,-0.45,0.54,-
1.07,0.41,0.29,0.86,1.40,-0.22,-1.75,-1.75,1.61,0.56,0.50,0.30,-1.12,1.01,1.02,1.13,0.82,-0.05)

#Modelo ~ Consumo = Temperatura*B1 + IndiceGeral*B2
CP3 <- c(-0.93,2.20,-0.08,0.69,-0.46,-1.02,-1.15,-0.54,0.53,1.55,1.37,1.22,-1.33,-
0.42,1.33,0.25,-0.39,-0.42,-1.61,-0.64,0.76,1.04,0.39,-0.54,-1.84,-1.00,0.69,1.73,-0.31,-0.67,-1.30,-
0.79,-0.18,-0.65,1.10,0.61,-1.79,-0.81,0.79,0.91,-0.17,-0.37,-0.84,-0.82,0.80,0.98,0.75,1.38)
TP3 <- c(0.96,0.96,0.86,-0.12,-0.94,-2.00,-1.77,-0.75,-0.47,0.42,1.41,1.39,1.37,1.15,0.16,-
0.45,-0.97,-1.17,0.25,1.52,0.81,0.04,0.22,0.52,0.33,0.64,0.79,-1.22,-2.64,-1.36,-0.02,-0.79,-1.69,-
0.09,1.00,1.23,0.81,0.54,0.62,-0.05,-1.24,-1.42,-0.43,0.36,0.40,0.08,-0.04,0.79)
IP3 <- c(-1.81,-1.81,-0.12,0.22,0.31,-0.96,-0.41,-0.22,0.52,0.53,0.31,-0.22,-1.81,-
1.08,0.62,0.40,0.65,0.48,-1.47,0.25,0.74,0.75,0.76,-0.76,-1.81,-1.69,1.10,0.79,0.67,1.22,-
1.13,0.70,1.12,0.84,0.86,-0.56,-1.81,-1.81,1.21,0.52,1.08,0.93,-1.13,0.78,0.78,1.03,0.78,-0.33)

#Criando um dataframe para compilar os dados
dataframe.1 <- data.frame(CP1,TP1)
dataframe.2 <- data.frame(CP2,IC2)
dataframe.3 <- data.frame(CP3,TP3,IP3)

View(dataframe.3)

# Verificar se há linhas completas de dados no dataframe
complete_rows <- complete.cases(dataframe)

# Filtrar o dataframe para manter apenas as linhas completas
dataframe_complete <- dataframe[complete_rows, ]

# Plotar o gráfico de dispersão 3D usando o dataframe completo
scatterplot3d(x = dataframe.3$TP3,
              y = dataframe.3$IP3,
              z = dataframe.3$CP3,
              xlab = "Temperatura Média",
              ylab = "Índice De Utilização de Salas",

```

```

        zlab = "Consumo Mensal",
        pch = 16)

#View(dataframe.1)
#View(dataframe.2)
#View(dataframe.3)

# Escreve os dados em um arquivo .CSV separado por colunas
options(OutDec = ".")
write.table(dataframe.1, file = "Normalizado1.csv", sep = ",", row.names = FALSE)

options(OutDec = ".")
write.table(dataframe.2, file = "Normalizado2.csv", sep = ",", row.names = FALSE)

options(OutDec = ".")
write.table(dataframe.3, file = "Normalizado3.csv", sep = ",", row.names = FALSE)

# Importa os dados do arquivo .CSV
dataframe.1 <- read.csv("Normalizado1.csv" , header = TRUE)
dataframe.2 <- read.csv("Normalizado2.csv" , header = TRUE)
dataframe.3 <- read.csv("Normalizado3.csv" , header = TRUE)

# Aplica o teste de Shapiro-Wilk á variavel de interesse

#Modelo 1
R1 <- shapiro.test(dataframe.1$CP1)
R2 <- shapiro.test(dataframe.1$TP1)

print(R1)
print(R2)

#Modelo 2
R3 <- shapiro.test(dataframe.2$CP2)
R4 <- shapiro.test(dataframe.2$IC2)

print(R3)
print(R4)

#Modelo 3
R5 <- shapiro.test(dataframe.3$CP3)
R6 <- shapiro.test(dataframe.3$TP3)
R7 <- shapiro.test(dataframe.3$IP3)

print(R5)
print(R6)
print(R7)

## Calcula a matriz de correlação de Pearson
cor(dataframe.1, method = "pearson")
cor(dataframe.2, method = "pearson")
cor(dataframe.3, method = "pearson")

# Ajusta um modelo de regressão linear múltipla
modelo1 <- lm(CP1 ~ TP1 , data=dataframe.1)
modelo2 <- lm(CP2 ~ IC2 , data=dataframe.2)
modelo3 <- lm(CP3 ~ TP3 + IP3 , data=dataframe.3)

# Exibe os resultados do modelo

```

```
summary(modelo1)
summary(modelo2)
summary(modelo3)

## Análise gráfica:
par(mfrow=c(2,2))
plot(modelo1)
par(mfrow=c(1,1))
par(mfrow=c(2,2))
plot(modelo2)
par(mfrow=c(1,1))
par(mfrow=c(2,2))
plot(modelo3)
par(mfrow=c(1,1))

## Ausência de Multicolinearidade
pairs.panels(dataframe.1)
pairs.panels(dataframe.2)
pairs.panels(dataframe.3)

### Multicolinearidade:  $r > 0.9$  (ou 0.8)

### Multicolinearidade:  $VIF > 10$ 

vif(modelo3)

## Outliers nos resíduos:
summary(rstandard(modelo1))

durbinWatsonTest(modelo1)

summary(rstandard(modelo2))

durbinWatsonTest(modelo2)

summary(rstandard(modelo3))

durbinWatsonTest(modelo3)

# Identificar outliers usando o método de Dixon
outliers <- boxplot.stats(dataframe.3)$out

# Imprimir os valores dos outliers
print(outliers)

# Remover outliers do conjunto de dados
dados_sem_outliers <- dataframe.3[!dataframe.3 %in% outliers]

# Realizar regressão com os dados sem outliers
modelo4 <- lm(CP3 ~ TP3 + IP3 , data = dados_sem_outliers)
```