



Universidade de Brasília
Departamento de Estatística

Aplicação de métodos de reamostragem computacional na regressão linear múltipla para o estudo da composição física e do poder energético da madeira de Eucalipto durante o processo de torrefação

João Pedro Moreira Pupe

Projeto apresentado para o Departamento de Estatística da Universidade de Brasília como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

Brasília
2023

João Pedro Moreira Pupe

Aplicação de métodos de reamostragem computacional na regressão linear múltipla para o estudo da composição física e do poder energético da madeira de Eucalipto durante o processo de torrefação

Orientador(a): Frederico Machado Almeida
Coorientador(a): Edgar Amaral Silveira

Projeto apresentado para o Departamento de Estatística da Universidade de Brasília como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

**Brasília
2023**

Sumário

1 Introdução	10
2 Revisão de Literatura	12
2.1 Regressão Linear Múltipla	12
2.1.1 Método dos Mínimos Quadrados	13
2.1.2 Estimação dos Parâmetros	14
2.1.3 Intervalo de Confiança dos Parâmetros	14
2.1.4 Valores Ajustados	14
2.1.5 Resíduos	15
2.1.6 Coeficiente de Determinação na Regressão	15
2.2 Testes de Hipóteses	16
2.2.1 Teste t	16
2.2.2 Teste de Shapiro-Wilk	16
2.2.3 Teste de Durbin-Watson	17
2.2.4 Teste de Breusch-Pagam	18
2.3 Métodos de Reamostragem Computacional	18
2.3.1 Bootstrap	18
2.3.2 Jackknife	19
2.4 Gráficos	20
2.4.1 Gráfico de Dispersão	20
2.4.2 Histograma	21
2.4.3 Boxplot	22
2.4.4 Gráfico QQ	22
3 Metodologia	24
3.1 Material	24
3.2 Análise dos Dados	24
4 Resultados	26
4.1 Análises Descritivas	26
4.1.1 Variável X_1	26

4.1.2	Variável X_2	27
4.1.3	Variável X_3	28
4.1.4	Variável Y_1	29
4.1.5	Variável Y_2	29
4.1.6	Correlações	30
4.2	Modelo I	32
4.2.1	Modelo Original	32
4.2.2	Métodos Bootstrap e Jackknife	33
4.3	Modelo II	40
4.3.1	Modelo Original	40
4.3.2	Métodos Bootstrap e Jackknife	41
5	Conclusão	48
6	Apêndice	50
6.1	Banco de Dados	50
6.2	Códigos R	50

Lista de Tabelas

1	Medidas descritivas de X_1	26
2	Medidas descritivas de X_2	27
3	Tabela de frequência de X_3	28
4	Medidas descritivas de Y_1	29
5	Medidas descritivas de Y_2	30
6	Avaliação das estimativas originais do Modelo I	32
7	Resultado dos Testes para o Modelo I	33
8	Avaliação das estimativas de Bootstrap e Jackknife do Modelo I, onde β_j^{or} é o β_j original, $\hat{\beta}_j^{Btr}$ é a média das estimativas dos $\hat{\beta}_j$'s obtida usando o método de Bootstrap e $\hat{\beta}_j^{Jk}$ a média das estimativas dos $\hat{\beta}_j$'s obtida usando o método de Jackknife.	34
9	P-valores dos testes de shapiro-wilk para \bar{e} das reamostragens do modelo I	38
10	Avaliação das estimativas originais do Modelo II	40
11	Resultado dos Testes para o Modelo II	41
12	Avaliação das estimativas de Bootstrap e Jackknife do Modelo II, onde β_j^{or} é o β_j original, $\hat{\beta}_j^{Btr}$ é a média das estimativas dos $\hat{\beta}_j$'s obtida usando o método de Bootstrap e $\hat{\beta}_j^{Jk}$ a média das estimativas dos $\hat{\beta}_j$'s obtida usando o método de Jackknife.	42
13	P-valores dos testes de shapiro-wilk para \bar{e} das reamostragens do modelo II	46
14	Banco de Dados	50

Lista de Figuras

1	Exemplo de Gráfico de Dispersão	21
2	Exemplo de Histograma	21
3	Exemplo de Boxplot	22
4	Exemplo de Gráfico QQ	23
5	Boxplot e Histograma de X_1	27
6	Boxplot e Histograma de X_2	28
7	Boxplot e Histograma de Y_1	29
8	Boxplot e Histograma de Y_2	30
9	Matriz de correlações	31
10	Gráfico de dispersão dos Erros do Modelo I	33
11	Histogramas e gráficos QQ das estimativas de $\hat{\beta}_0$ para o Modelo I	35
12	Histogramas e gráficos QQ das estimativas de $\hat{\beta}_1$ para o Modelo I	36
13	Histogramas e gráficos QQ das estimativas de $\hat{\beta}_2$ para o Modelo I	36
14	Histogramas e gráficos QQ das estimativas de \hat{R}^2 para o Modelo I	37
15	Histogramas e gráficos QQ das estimativas de \bar{e} para o Modelo I	37
16	Boxplots dos Vieses das estimativas de Bootstrap do modelo I	39
17	Boxplots dos Vieses das estimativas de Jackknife do modelo I	39
18	Gráfico de dispersão dos Erros do Modelo II	41
19	Histogramas e gráficos QQ das estimativas de $\hat{\beta}_0$ para o Modelo II	43
20	Histogramas e gráficos QQ das estimativas de $\hat{\beta}_1$ para o Modelo II	43
21	Histogramas e gráficos QQ das estimativas de $\hat{\beta}_2$ para o Modelo II	44
22	Histogramas e gráficos QQ das estimativas de \hat{R}^2 para o Modelo II	44
23	Histogramas e gráficos QQ das estimativas de \bar{e} para o Modelo II	45
24	Boxplots dos Vieses das estimativas de Bootstrap do modelo II	46
25	Boxplots dos Vieses das estimativas de Jackknife do modelo II	47

Agradecimentos

Agradeço sinceramente a todos aqueles que contribuíram para a realização deste trabalho e que estiveram ao meu lado durante toda essa jornada acadêmica. Em especial, gostaria de expressar minha gratidão à minha família, que foi fundamental em cada etapa desse processo.

Ao meu amigo, Rafael Araruna, você foi o melhor veterano que eu poderia ter encontrado. Sua orientação e apoio foram inestimáveis, e sou grato por todas as vezes que você compartilhou seus conhecimentos e experiências comigo.

Aos meus colegas de semestre, João Victor Melo, Guilherme Silva, Caio Cavalcante e Lucas Menezes, gostaria de agradecer por estarem sempre dispostos a ajudar. Nossa parceria e troca de conhecimentos foram essenciais para o desenvolvimento deste trabalho e também para o meu crescimento pessoal ao longo do curso.

Aos meus amigos de vida, João Victor Carvalho e Rafael Boaventura. Vocês estiveram presentes em todos os momentos, tanto acadêmicos quanto pessoais, e foram fundamentais para o meu amadurecimento e evolução. Suas conversas inspiradoras e apoio constante foram um verdadeiro incentivo para seguir em frente.

À minha parceira, Maria Clara, você foi uma parte essencial em todo esse processo e merece todo meu reconhecimento e gratidão. Sua presença constante, incentivo e apoio incondicional foram verdadeiramente inspiradores. Obrigado por estar ao meu lado, me motivar nos momentos de desafio e compartilhar alegrias nos momentos de conquista. Sua compreensão e amor inabaláveis tornaram essa jornada acadêmica ainda mais significativa e gratificante.

A todos os demais amigos, familiares e pessoas queridas que não foram mencionados, saibam que cada gesto de carinho, encorajamento e suporte foi fundamental para minha jornada acadêmica e para a conclusão deste trabalho.

Por fim, quero expressar minha gratidão aos professores, orientadores e demais membros da instituição pelo ensino de qualidade e pelas oportunidades proporcionadas ao longo do curso.

A todos vocês, meu mais sincero obrigado. Sem o apoio e a colaboração de cada um, este trabalho não seria possível. Vocês foram pilares essenciais no meu caminho até aqui, e sou imensamente grato por tudo que fizeram por mim.

Resumo

A crescente população mundial e o aumento do consumo de recursos fósseis têm causado impactos ambientais significativos, como o aquecimento global. Por isso, a busca por fontes de energia renovável, como solar, eólica e biomassa, torna-se crucial. No Brasil, a biomassa, especialmente proveniente de florestas energéticas de eucalipto, desempenha um papel importante na matriz energética, sendo utilizada na indústria de papel e celulose e na produção de carvão vegetal. No entanto, a biomassa *in natura* possui desvantagens como alto teor de umidade e baixo teor de carbono, o que compromete seu uso como biocombustível. Para superar esses desafios, a torrefação de biomassa tem sido utilizada como tratamento termoquímico para aumentar a densidade energética do produto, otimizando os processos de manejo, transporte e estocagem, além de mitigar custos desnecessários. Nesse contexto, o estudo propõe a aplicação dos métodos de reamostragem computacional de Bootstrap e de Jackknife em modelos de regressão linear múltipla para analisar o rendimento de massa e o poder energético da madeira de eucalipto durante a torrefação. Os resultados das análises indicaram que o método de Bootstrap apresentou melhor desempenho que o de Jackknife, ele foi capaz de indicar que apesar da não aleatoriedade dos resíduos dos modelos de regressão, as estimativas são consistentes, o que permite entender melhor a relação de causa e efeito dos parâmetros operacionais do processo de torrefação nos seus resultados.

Palavras-chave: Biomassa; Torrefação; Eucalipto; Reamostragem Computacional; Bootstrap; Jackknife; Regressão Linear Múltipla.

Abstract

The growing world population and increasing consumption of fossil resources have caused significant environmental impacts, such as global warming. Therefore, the quest for renewable energy sources, such as solar, wind, and biomass, becomes crucial. In Brazil, biomass, especially from energy forests of eucalyptus, plays an important role in the energy matrix, being used in the pulp and paper industry and charcoal production. However, raw biomass has disadvantages such as high moisture content and low carbon content, which compromise its use as a biofuel. To overcome these challenges, biomass torrefaction has been used as a thermochemical treatment to increase the energy density of the product, optimizing handling, transportation, and storage processes while mitigating unnecessary costs. In this context, the study proposes the application of the computational resampling methods of Bootstrap and Jackknife in multiple linear regression models to analyze the mass yield and energy content of eucalyptus wood during torrefaction. The analysis results indicated that the Bootstrap method outperformed Jackknife, as it was capable of indicating that despite the non-randomness of the regression models residuals, the estimates are consistent, allowing for a better understanding of the cause and effect relationship of the operational parameters of the torrefaction process on its results.

Keywords: Biomass; Torrefaction; Eucalyptus; Computational Resampling; Bootstrap; Jackknife; Multiple Linear Regression.

1 Introdução

O planeta terra e o ser humano precisam de energia renovável. É fato que a população mundial cresceu exponencialmente na era que muitos geólogos gostam de chamar de Antropoceno, em que o homem passou, por meio de suas atividades, a ser o principal fator de transformações ecossistêmicas globais (DA VEIGA, 2019). Dessa forma, o consumo de recursos fósseis (não renováveis) aumentou significativamente, o que resultou na ampliação de emissões de gases nocivos (CO₂, NO_x e SO₂) e em consequentes impactos ambientais como o aquecimento global. Portanto, a procura por vetores energéticos renováveis como solar, eólica e biomassa se faz necessária e de suma importância (SILVEIRA et al., 2021).

O Brasil se destaca com uma matriz energética composta por 51,6% de fontes renováveis e 48,4% de não renováveis (MATOS; SOARES, 2022). Dentro dos vetores renováveis, a biomassa tem papel importante, constituindo 25,1% (dos quais lenha e carvão representam 8,7%) da matriz energética brasileira (MATOS; SOARES, 2022). Nesse contexto, as florestas energéticas de eucalipto apresentam um enorme potencial, pois possuem acelerado desenvolvimento e adaptação à diferentes ambientes, sendo vastamente explorado na indústria de papel e celulose, assim como na produção de carvão como biocombustível.

A biomassa, na sua forma *in natura*, apresenta certas desvantagens como combustível sólido. Entre elas, pode-se pontuar o alto teor de umidade e o alto conteúdo de hidrogênio e oxigênio. Essas desvantagens intrínsecas podem dificultar a utilização da biomassa como biocombustível, uma vez que compromete o processo de armazenamento (material propício a ataque de fungos) e reduz sua capacidade energética (baixo teor de carbono). Uma das soluções para superar essas desvantagens são os conhecidos tratamentos termoquímicos, como a torrefação de biomassa (SILVEIRA et al., 2021).

A torrefação é um tratamento termoquímico que objetiva a densificação energética de uma biomassa. Este processo acontece em uma atmosfera inerte com temperaturas variando entre 200 e 300 °C (SILVEIRA et al., 2021). Os parâmetros operacionais desse processo definem a severidade do tratamento e o nível de modificação da estrutura da biomassa. Os principais parâmetros são a temperatura e o tempo de tratamento. Durante a torrefação, compostos químicos formados por oxigênio e hidrogênio são volatilizados e o carbono fica retido, aumentando assim a densidade energética do produto torrefado.

O atrativo da torrefação é que, durante o processo, há uma redução da massa em 30%. Ou seja, ocorre o aumento do poder calorífico superior (PCS) da biomassa devido a retenção de carbono e, ao mesmo tempo, reduz a sua massa, aumentando assim sua densidade energética. Dessa maneira, o tratamento otimiza os processos de manejo,

transporte e estocagem, benefícios cruciais para quem investe no setor de carvão vegetal.

Após a torrefação de uma determinada biomassa, é necessário realizar experimentos para avaliar o efeito da severidade do tratamento nas características de interesse. Por exemplo, a análise elementar (composição de % carbono, % hidrogênio, % oxigênio) e a calorimétrica (PCS). O levantamento de todas as características, a partir do tratamento da torrefação e posterior análise do biocarvão, é complexo e custoso. Portanto, engenheiros e pesquisadores buscam métodos numéricos e/ou estatísticos que possam ser úteis para prever seus resultados sem a necessidade de executar diversos experimentos. Dessa forma, otimizam e mitigam custos desnecessários, aumentando o valor agregado do produto e permitindo a inserção dele no mercado com preços mais competitivos. (SILVEIRA et al., 2021).

A análise de regressão é uma metodologia estatística que utiliza uma relação matemática para avaliar a força que um conjunto de variáveis independentes (explicativas ou simplesmente covariáveis) exercem, ou que parecem exercer, sobre outra, a chamada variável dependente (ou explicada). Essa metodologia é amplamente utilizada em negócios, ciências sociais e comportamentais, engenharias e muitas outras áreas de conhecimento, apresentando enorme potencial no contexto de transição energética e biocombustível. Nesse contexto, o presente trabalho tem como objetivo aplicar métodos de amostragem computacional nos modelos de regressão linear múltipla no estudo do rendimento sólido e do poder energético da madeira de eucalipto durante o processo de torrefação e dessa maneira avaliar o desempenho dos estimadores dos mínimos quadrados para diferentes métodos de amostragem bem como entender a influência e a relação de causa e efeito dos parâmetros tempo e temperatura do processo de torrefação nos resultados de poder calorífico superior e rendimento de massa e por fim apresentar uma discussão geral sobre os resultados obtidos durante os processos de amostragem e estimação.

2 Revisão de Literatura

2.1 Regressão Linear Múltipla

A origem da teoria de regressão remonta do século XIX, quando Galton estudou em seus trabalhos, a relação entre a altura dos pais e dos filhos, digamos $(X_i$ e $Y_i)$, procurando saber como a altura do pai influenciava a altura do filho (DEMÉTRIO; ZOCCHI, 2006). No experimento, Galton notou que se o pai fosse muito alto ou muito baixo, o filho teria uma altura tendendo à média. Por isso, ele chamou de regressão, ou seja, existe uma tendência de os dados regredirem à média.

Em suma, a utilização dos modelos de regressão, pode ter por objetivos: (i) prever valores da variável depende por meio de um conjunto de variáveis independentes; (ii) selecionar as principais variáveis que exercem força significativa sobre a variável dependente; (iii) estimar os parâmetros do modelo; por fim, (iv) fazer inferências sobre os parâmetros estimados. É importante salientar que, mesmo que não exista relação causal entre as variáveis que compõem a estrutura de regressão, podemos relacioná-las por meio de uma expressão matemática, que pode ser útil para se estimar o valor de uma das variáveis quando conhecemos os valores das outras, sob determinadas condições.

A análise de regressão linear múltipla é um dos métodos estatísticos mais usados na atualidade. A relação funcional entre a variável dependente, digamos Y , e um conjunto de variáveis explicativas $\mathbf{X} = (X_{11}, \dots, X_{np})^\top$ é regida por uma fórmula matemática $Y = f(X_{11}, \dots, X_{np})$. Porém, como na prática a relação entre Y e X não é perfeita, i.e., em situações reais vão existir fatores que podem intervir na relação, resultando na existência de alguns pontos fora da reta então, é conveniente escrever tal relação da seguinte forma:

$$Y = f(X_{11}, \dots, X_{np}) + \epsilon, \quad (2.1.1)$$

observe que, o modelo em (2.1.1) é composto por duas partes: a parte determinística representada por $f(X_{11}, \dots, X_{np})$ e a parte aleatória, caracterizada por ϵ (conhecido por termo de erro). O termo de erro irá captar todas as influências sobre Y não devidas ou explicadas pela parte determinística (reta de regressão). De forma geral, se x_1, x_2, \dots, x_p denota a coleção de variáveis independentes (constantes conhecidas), o modelo de regressão linear múltipla para a i -ésima observação da variável resposta é dado por:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \epsilon_i, \quad (2.1.2)$$

para $i = 1, 2, \dots, n$. Sendo $\beta_0, \beta_1, \dots, \beta_p$ parâmetros desconhecidos, comumente desig-

nados por coeficientes de regressão, e ϵ_i são independentes e identicamente distribuídas da normal com média 0 e variância constante, i.e, $\epsilon_i \sim N(0, \sigma^2)$. Observe que, quando $p = 1$, ou seja, quando temos uma variável explicativa na estrutura de regressão, o modelo em (2.1.2) reduz-se para uma equação de regressão linear simples, veja (NETER et al., 1996) para mais detalhes. Em termos matriciais, a equação de regressão linear apresentado em (2.1.2) pode ser expressa da seguinte forma:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad (2.1.3)$$

onde:

\mathbf{Y} é o vetor $n \times 1$ de variáveis respostas.

$\boldsymbol{\beta}$ é o vetor $(p + 1) \times 1$ dos coeficientes de regressão.

\mathbf{X} é a matriz experimental com dimensão $n \times p$, e

$\boldsymbol{\epsilon}$ é o vetor de erros que são independentes e identicamente distribuídas tal que $\mathbf{E}(\boldsymbol{\epsilon}) = \mathbf{0}$ e variância-covariância $\boldsymbol{\sigma}(\boldsymbol{\epsilon}) = \sigma^2 \mathbf{I}_n$. Sendo \mathbf{I}_n uma matriz identidade.

Desta forma, segue imediatamente que $\mathbf{E}(\mathbf{Y}|\mathbf{X}) = \mathbf{X}\boldsymbol{\beta}$, e $\sigma^2(\mathbf{Y}|\mathbf{X}) = \boldsymbol{\sigma}^2(\boldsymbol{\epsilon}) = \sigma^2 \mathbf{I}_n$, e $\mathbf{Y}|\mathbf{X} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n)$. A linearidade do modelo em (2.1.3) se refere ao fato de que os parâmetros ou coeficientes de regressão devem aparecer em uma estrutura linear no modelo, e portanto, no caso mais simples, espera-se que o diagrama de dispersão entre as variáveis X e Y sugerira um comportamento linear no espalhamento dos pontos. Além disso, para que as aplicações dos modelos sejam adequadas aos dados disponíveis, outras suposições associadas aos erros, tais como: homoscedasticidade, independência dos erros, linearidade devem ser observadas. A estimativa do vetor associado aos coeficientes de regressão $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^\top$ é obtida por meio do chamado método dos mínimos quadrados (MMQ).

2.1.1 Método dos Mínimos Quadrados

Este método consiste em uma técnica de otimização matemática, e tem como objetivo encontrar o melhor ajuste para um conjunto de dados, tentando, assim, minimizar a soma dos quadrados das diferenças entre o valor estimado e os dados observados (tais diferenças são chamadas resíduos) (FISHER, 1925).

Nesse sentido, se trata de um estimador que minimiza a soma dos quadrados dos resíduos da regressão, de forma a maximizar o grau de ajuste do modelo aos dados observados.

2.1.2 Estimação dos Parâmetros

Mantendo a notação matricial e utilizando os conhecimentos de estimação por mínimos quadrados, os coeficientes do modelo, ou seja, β_j , com $j=0,1,\dots,p$, são estimados por:

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \quad (2.1.4)$$

2.1.3 Intervalo de Confiança dos Parâmetros

Os intervalos de confiança para os parâmetros, também conhecidos como ICs, geralmente são calculados com 95% de nível de confiança.

Para o cálculo deles, primeiramente precisa-se da variância estimada dos parâmetros. Essa variância é calculada da seguinte maneira para o intercepto do modelo:

$$Var(\hat{\beta}_0) = \sigma^2 \left(\frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right), \quad (2.1.5)$$

e de outra maneira para o restante dos parâmetros do modelo:

$$Var(\hat{\beta}_j) = \frac{\hat{\sigma}^2}{\sum (x - \bar{x})^2}. \quad (2.1.6)$$

Feito isso, é calculado o intervalo de confiança para cada $\hat{\beta}_j$:

$$IC_{1-\alpha}(\hat{\beta}_j) = \hat{\beta}_j \pm t_{(n-p,\alpha/2)} \sqrt{Var(\hat{\beta}_j)}, \quad (2.1.7)$$

em que $t_{(n-p,\alpha/2)}$ são os quantis da distribuição t-student com $n - p$ graus de liberdade e $j = 0,1,2,3,\dots$

2.1.4 Valores Ajustados

Seja o vetor de valores ajustados \hat{Y}_i denotado por \hat{Y} , então, em notação matricial, tem-se que:

$$\hat{Y} = \mathbf{X}\hat{\beta}. \quad (2.1.8)$$

Nesse sentido, utilizando a definição da equação 2.1.4, a equação 2.1.8 pode ser reescrita como:

$$\hat{Y} = \mathbf{X}\hat{\beta} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}. \quad (2.1.9)$$

2.1.5 Resíduos

Seja o vetor de resíduos $\hat{e}_i = Y_i - \hat{Y}_i$ denotado por \hat{e} , então, em notação matricial, tem-se que:

$$\hat{e} = \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{X}\hat{\beta}. \quad (2.1.10)$$

2.1.6 Coeficiente de Determinação na Regressão

O coeficiente de determinação, também chamado de R^2 , indica o quanto da variação da variável Y é explicado pelas variáveis independentes (x_1, x_2, \dots, x_p) . Esse coeficiente varia entre 0 e 1, indicando em porcentagem quanto está sendo explicado pelo modelo, ou seja, quanto mais perto de 1, mais as variáveis independentes explicam sobre a variação de Y . Seu valor é obtido a partir da fórmula:

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = \frac{SQE}{SQT} \quad (2.1.11)$$

com:

- p = número de variáveis explicativas;
- n = tamanho da amostra;
- \bar{y} = média amostra da variável resposta Y ;
- \hat{y}_i = i -ésimo valor predito pela regressão;
- $\sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = SQE$ = soma de quadrados explicada;
- $\sum_{i=1}^n (y_i - \bar{y})^2 = SQT$ = soma de quadrados total.

2.2 Testes de Hipóteses

Nessa seção, foi utilizado os livros BUSSAB e MORETTIN (2018) e Shapiro e Wilk (1965) para obter as informações acerca das técnicas utilizadas.

2.2.1 Teste t

Esse teste consiste em verificar se o valor do parâmetro β_j é diferente de zero, isto é, testa se a variável X_j tem alguma influência sobre o valor esperado de Y_i . Para isso, estabelece-se as seguintes hipóteses:

$$\begin{cases} H_0 : \beta_j = 0 \\ H_1 : \beta_j \neq 0. \end{cases}$$

O objetivo do teste é dizer se H_0 é ou não aceitável. Para isso, primeiramente, define-se o nível de significância α , o qual é definido da seguinte forma:

$$\alpha = P(\text{rejeitar } H_0 | H_0 \text{ é verdadeira}), \quad (2.2.1)$$

ou seja, é a probabilidade de rejeitar a hipótese nula dada que ela é verdadeira. Usualmente, o valor de α é fixado como 5%, 10% ou 1%.

Em seguida, calcula-se o p-valor, o qual indica a probabilidade de ocorrer valores da estatística mais extremos do que o observado, sob a hipótese de H_0 ser verdadeira. Assim, se o p-valor for menor que α , H_0 é rejeitada.

2.2.2 Teste de Shapiro-Wilk

O teste de Shapiro-Wilk é um teste estatístico utilizado para verificar se uma determinada amostra de dados segue uma distribuição normal. Ele é amplamente utilizado para testar a normalidade dos dados antes de aplicar certas análises estatísticas. Suas hipóteses são:

$$\begin{cases} H_0 : \text{A variável segue uma distribuição Normal} \\ H_1 : \text{A variável não segue uma distribuição Normal.} \end{cases}$$

O teste consiste na comparação entre os valores observados na amostra e os valores esperados de uma distribuição normal (SHAPIRO; WILK, 1965). A hipótese nula do teste é que a amostra segue uma distribuição normal. Se o p-valor resultante do teste for menor

do que um nível de significância pré-determinado (geralmente 0,05), então a hipótese nula é rejeitada, indicando que os dados não seguem uma distribuição normal.

O procedimento do teste envolve o cálculo de uma estatística de teste que leva em consideração as diferenças entre os valores observados e esperados. Essa estatística de teste é comparada a uma distribuição conhecida para determinar o p-valor correspondente.

O cálculo do p-valor no teste de Shapiro-Wilk envolve a comparação da estatística de teste com a distribuição conhecida da estatística de teste para a hipótese nula de normalidade. A interpretação do p-valor é feita em relação a um nível de significância pré-determinado, já que ele significa a probabilidade de obter uma estatística de teste igual ou mais extrema do que a observada, assumindo que a hipótese nula seja verdadeira. Quanto menor o p-valor, maior a evidência contra a hipótese nula.

O teste de Shapiro-Wilk é considerado um dos métodos mais poderosos para testar a normalidade dos dados, especialmente para amostras de tamanho moderado a grande. Ele é preferível a outros testes, como o teste de Kolmogorov-Smirnov, quando se trata de detectar desvios da normalidade em pequenas amostras.

2.2.3 Teste de Durbin-Watson

Esse teste serve para verificar os desvios na direção de:

$$\varepsilon_i = \rho\varepsilon_{i-1} + v_i.$$

Explicando de uma maneira melhor, esse teste serve para detectar correlação de ordem 1, auto-correlação de ordem 1, ou seja, seria verificar se o erro na unidade presente (ε_i) está correlacionado com o erro na unidade passada (ε_{i-1}).

De maneira resumida, seu objetivo é verificar se os erros são independentes ou não.

Suas hipóteses são:

$$\begin{cases} H_0 : \text{Os erros são independentes} \\ H_1 : \text{Os erros não são independentes.} \end{cases}$$

A análise para verificar se H_0 é ou não aceitável é feita da mesma maneira como foi explicado no teste t.

2.2.4 Teste de Breusch-Pagam

Para verificar a homogeneidade da variância, pode-se aplicar o teste de Breusch-Pagan, que é utilizado para testar se a variância do erro de um modelo de regressão é constante. É indicado para grandes amostras e é sensível quanto à normalidade dos resíduos.

O teste possui as seguintes hipóteses:

$$\begin{cases} H_0 : \text{As variâncias dos erros são iguais} \\ H_1 : \text{As variâncias dos erros são diferentes.} \end{cases}$$

A análise para verificar se H_0 é ou não aceitável é feita da mesma maneira como foi explicado no teste t.

2.3 Métodos de Reamostragem Computacional

A reamostragem computacional é uma técnica estatística que envolve a geração de amostras adicionais a partir de uma amostra original, utilizando métodos computacionais. Essas amostras adicionais permitem estimar parâmetros, avaliar incertezas e construir intervalos de confiança, fornecendo uma abordagem flexível e poderosa para lidar com problemas estatísticos complexos, especialmente quando não é possível obter soluções analíticas. A reamostragem computacional, incluindo métodos como o Bootstrap e o Jackknife, desempenha um papel importante na inferência estatística moderna, fornecendo uma maneira de obter informações mais confiáveis sobre os parâmetros populacionais e a variabilidade associada a eles.

2.3.1 Bootstrap

O Bootstrap, é uma técnica estatística utilizada para estimar a incerteza associada a uma medida estatística ou para avaliar a distribuição de uma estatística desconhecida. O método envolve a criação de várias amostras de dados, chamadas de amostras de Bootstrap, a partir de uma única amostra original (BUSSAB; MORETTIN, 2018).

A ideia por trás do Bootstrap é que as amostras sejam representativas da população original. A partir dessas amostras, é possível calcular estatísticas repetidamente, obtendo assim uma distribuição empírica delas. Essa distribuição pode ser usada para calcular intervalos de confiança, realizar testes de hipóteses ou fornecer estimativas de variância (BUSSAB; MORETTIN, 2018).

No Bootstrap clássico, a reamostragem é realizada com reposição, ou seja, uma observação pode ser selecionada várias vezes em uma amostra de Bootstrap. Essa é a abordagem padrão do Bootstrap e a mais comumente utilizada.

No entanto, é possível fazer uma variação do Bootstrap chamada de "Bootstrap sem reposição" ou "Bootstrap de reamostragem balanceada". Nesse caso, a reamostragem é feita sem reposição, o que significa que cada observação é selecionada apenas uma vez na amostra de Bootstrap. Essa abordagem é usada quando se deseja manter o mesmo tamanho para todas as reamostras e garantir que cada observação seja representada uma única vez em cada reamostra de Bootstrap (DAVISON; HINKLEY, 1997).

O procedimento para obtenção de amostras Bootstrap, considerando uma amostra original $X = \{x_1, x_2, \dots, x_n\}$ e a estatística de interesse $\hat{\theta} = F(X)$, está descrito a seguir:

1. Geram-se amostras Bootstrap $X(1), X(2), \dots, X(B)$, com ou sem reposição de X , em que B é o número de reamostras selecionadas.
2. Calculam-se as estimativas da estatística de interesse: $\hat{\theta}^{(b)} = F(X_b)$, $b = 1, 2, \dots, B$.
3. Calcula-se o erro padrão *Bootstrap*, \hat{S}_{Boot} como:

$$\hat{S}_{Boot} = \sqrt{\sum_{b=1}^B \frac{(\hat{\theta}^{(b)} - \hat{\theta}^{(*)})^2}{B-1}}$$

em que,

$$\hat{\theta}^{(*)} = \sum_{b=1}^B \frac{\hat{\theta}^{(b)}}{B}.$$

2.3.2 Jackknife

O método do Jackknife é uma técnica de reamostragem relacionada ao bootstrap. Ele é usado para estimar a variância ou outros estimadores da precisão de uma estatística. A ideia dele é criar uma série de amostras "deixando um de fora". Para cada amostra, calcula-se a estatística de interesse. Em seguida, calcula-se a variância dessas estatísticas, que fornece uma estimativa da variância da estatística original (EFRON, 1979).

Essa metodologia consiste em um método não-paramétrico de reamostragem destinado a estimar o viés e, assim, reduzir a variância dos estimadores em condições teoricamente complexas, ou em que não se tenha confiança no modelo especificado.

O procedimento para obtenção da amostra Jackknife é:

1. Seleciona-se uma amostra original de tamanho n : $X = \{x_1, x_2, \dots, x_n\}$

2. Define-se a estatística de interesse: $\hat{\theta} = F(X)$
3. Gera-se a amostra Jackknife 1: $X^{(1)} = \{x_2, x_3, \dots, x_{n-1}, x_n\}$ e $\hat{\theta}(1) = F(X^{(1)})$
4. Gera-se a amostra Jackknife 2: $X^{(2)} = \{x_1, x_3, \dots, x_{n-1}, x_n\}$ e $\hat{\theta}(2) = F(X^{(2)})$
5. Gera-se o restante de amostras Jackknife, sendo a última dada por:

$$X^{(n)} = \{x_1, x_2, \dots, x_{n-2}, x_{n-1}\}$$

e

$$\hat{\theta}^{(n)} = F(X^{(n)}).$$

6. Estima-se o erro padrão da estatística de interesse definida no passo 2 como:

$$\hat{S}_{Jack} = \sqrt{\frac{n-1}{n} \sum_{i=1}^n (\hat{\theta}^{(i)} - \hat{\theta}^{(\cdot)})^2}$$

em que,

$$\hat{\theta}^{(\cdot)} = \sum_{i=1}^n \frac{\hat{\theta}^{(i)}}{n}.$$

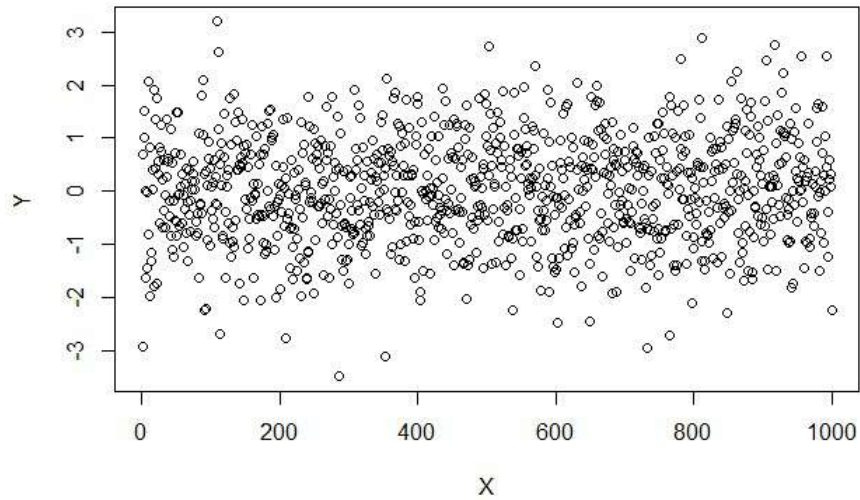
2.4 Gráficos

2.4.1 Gráfico de Dispersão

Um gráfico de dispersão, também conhecido como gráfico de dispersão de pontos ou gráfico de dispersão bidimensional, é uma representação visual que mostra a relação entre duas variáveis quantitativas. Ele é composto por um eixo horizontal (eixo x) que representa uma variável e um eixo vertical (eixo y) que representa a outra variável. Os pontos no gráfico são posicionados de acordo com os valores das duas variáveis em cada observação.

Os gráficos de dispersão são úteis para identificar padrões, tendências e relações entre variáveis (GELMAN et al., 2013). Eles podem ser utilizados para determinar se existe uma relação linear, não-linear, positiva, negativa ou nenhuma relação entre as variáveis. Além disso, os gráficos de dispersão também podem ajudar a identificar valores discrepantes (outliers) e observar a dispersão geral dos dados.

Figura 1: Exemplo de Gráfico de Dispersão

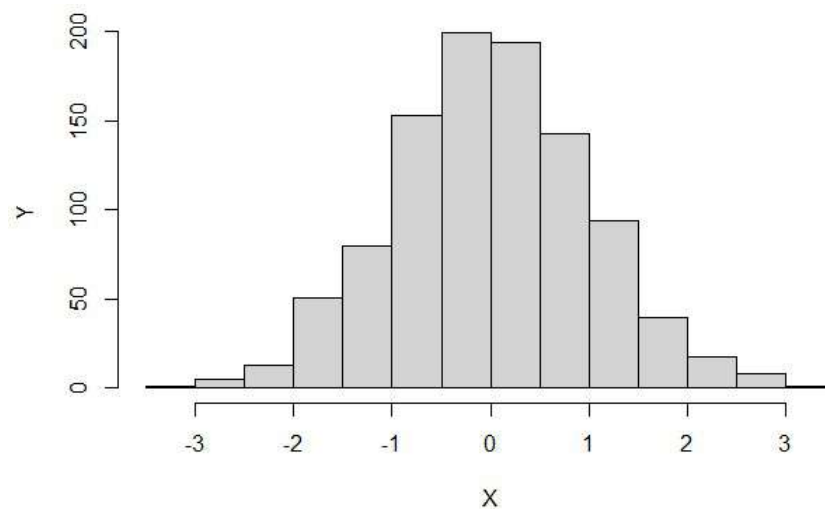


2.4.2 Histograma

Um histograma é um tipo de gráfico utilizado na análise de dados para visualizar a distribuição de uma variável quantitativa contínua. Ele consiste em barras adjacentes, onde cada barra representa uma faixa ou intervalo de valores e a altura da barra representa a frequência ou a proporção de observações que se enquadram nesse intervalo (WICKHAM; GROLEMUND, 2017).

Os histogramas permitem identificar a forma da distribuição dos dados, incluindo a presença de picos, assimetrias e dispersão. Eles são particularmente úteis para identificar a concentração dos dados em certas faixas ou intervalos e também para detectar possíveis valores discrepantes (outliers).

Figura 2: Exemplo de Histograma



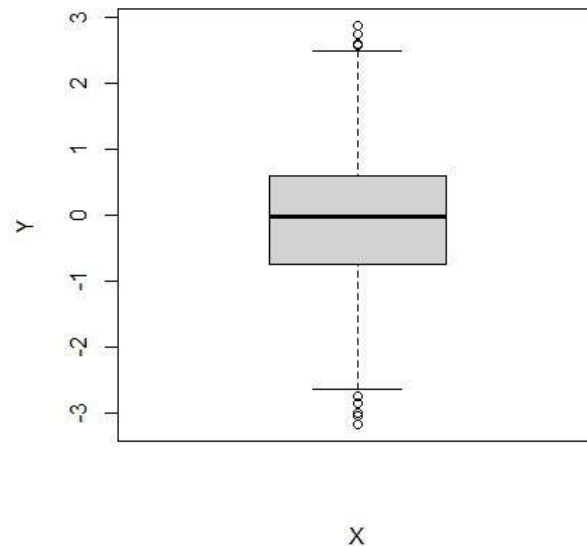
2.4.3 Boxplot

Boxplots, também conhecidos como diagramas de caixa, são gráficos estatísticos que resumem a distribuição de uma variável quantitativa. Eles fornecem uma representação visual das principais estatísticas descritivas, como mediana, quartis, valores mínimo e máximo, além de identificar possíveis valores discrepantes (outliers).

Um boxplot é construído a partir de cinco estatísticas resumidas da variável em questão: o valor mínimo, o primeiro quartil (25%), a mediana (50%), o terceiro quartil (75%) e o valor máximo. Essas estatísticas são representadas por um retângulo (caixa) e duas linhas verticais (hastes) que se estendem a partir da caixa. As hastes são geralmente estendidas até 1,5 vezes o intervalo interquartil (IQR), e qualquer valor fora dessa faixa é considerado um possível valor discrepante (VANDERPLAS, 2016).

Boxplots são úteis para identificar a simetria da distribuição, bem como a presença de valores discrepantes e a dispersão dos dados. Eles são especialmente úteis para comparar várias distribuições lado a lado e identificar diferenças nas medianas e na variabilidade.

Figura 3: Exemplo de Boxplot



2.4.4 Gráfico QQ

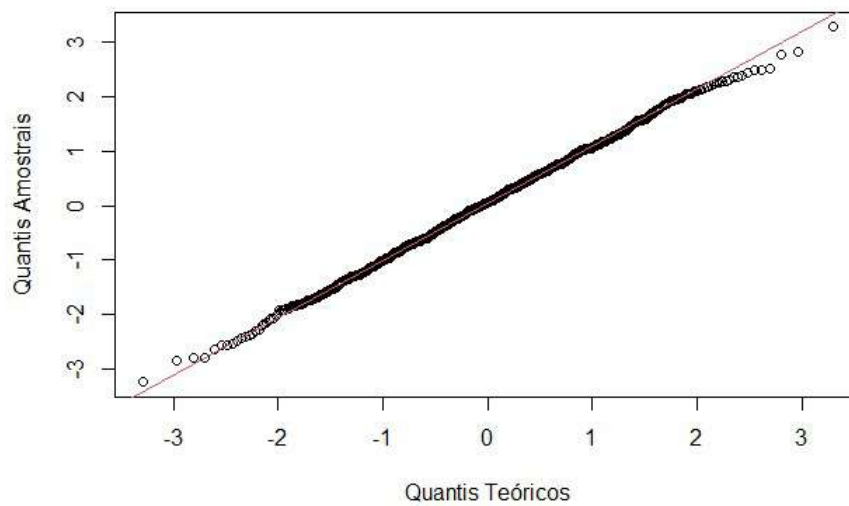
O gráfico QQ (Quantile-Quantile) é uma ferramenta gráfica utilizada na análise de dados para comparar a distribuição dos dados observados com uma distribuição teórica. Ele é especialmente útil para verificar se uma determinada distribuição teórica é adequada para descrever os dados empíricos (WILCOX, 2017).

No gráfico QQ, os quantis da distribuição teórica são plotados no eixo x, enquanto

os quantis correspondentes dos dados observados são plotados no eixo y. Se os dados observados seguirem exatamente a distribuição teórica, os pontos no gráfico QQ seguirão aproximadamente uma linha diagonal. Desvios dessa linha indicam desvios da distribuição teórica.

O gráfico QQ é frequentemente usado para verificar a normalidade dos dados. Nesse caso, os quantis teóricos são obtidos a partir da distribuição normal. Se os pontos no gráfico QQ seguirem aproximadamente uma linha diagonal, isso indica que os dados seguem uma distribuição normal. No entanto, se os pontos desviarem da linha diagonal, isso indica uma falta de ajuste da distribuição normal aos dados.

Figura 4: Exemplo de Gráfico QQ



3 Metodologia

3.1 Material

O material utilizado será o conjunto de dados do processo de torrefação de Eucalipto obtido do trabalho (SILVEIRA et al., 2021). A torrefação foi realizada em um analisador termogravimétrico (SDT Q600 TA). O tratamento foi realizado em duplicatas para amostras de 15 mg, com temperatura de torrefação de 210, 230, 250, 270 e 290 °C, todas as vezes, do minuto 0 ao minuto 70 dentro do forno, sempre em uma atmosfera inerte (livre de oxigênio)(SILVEIRA et al., 2021). Durante o tratamento, o equipamento caracterizava de minuto à minuto o produto torrefado através das análises de rendimento, de composição elementar (carbono/hidrogênio/oxigênio) e calorífica (PCS).

O conjunto de dados é composto por 355 observações, oriundas das permutações entre tempos e temperaturas citadas acima e por 5 variáveis, que são:

- X_1 : Tempo em minutos;
- X_2 : Temperatura em °C;
- X_3 : Temperatura desejada em °C;
- Y_1 : Porcentagem de Rendimento sólido;
- Y_2 : Poder Calorífico Superior em $KJ \cdot Kg^{-1}$ (Quilojoule por Quilograma) da biomassa.

As variáveis X_i , sendo $i = 1, 2, 3$, são as independentes, e as Y_j , sendo $j = 1, 2$, são as dependentes, que se deseja prever, pois são medidas retiradas de cada amostra após o processo de torrefação nelas aplicado.

3.2 Análise dos Dados

Para a análise dos dados, foi utilizado o software estatístico R (R Core Team, 2021) na versão 4.2.1, em que serão feitas as análises descritivas, as reamostragens e as modelagens estatísticas. Será utilizada a reamostragem computacional para criar diferentes amostras do banco de dados citado na subseção anterior e à elas serão ajustados modelos de regressão linear múltipla para estudar os efeitos de X_1 , X_2 em Y_1 e Y_2 , e avaliar a consistência de seus resultados.

Para avaliar o desempenho dos métodos de reamostragem que serão consideradas neste trabalho, propomos ajustar $R^* = 1000$ modelos para as amostras obtidas usando

o método Bootstrap, e $R^* = 355$ modelos serão considerados para o caso do método de Jackknife.

Seja β_j^{or} os coeficientes originais, isto é, coeficientes de regressão obtidos no ajuste do modelo original (quando não leva em conta os métodos de reamostragem computacional), com $j = 0, 1, 2$ o contador indexado aos coeficientes de regressão. Desta forma, as principais medidas de desempenho a serem consideradas são:

- (i) viés ordinário: $\text{viés}(\hat{\beta}_j) = \frac{1}{R^*} \sum_{i=1}^{R^*} (\hat{\beta}_{ij} - \beta_j^{or})$;
- (ii) probabilidade de cobertura (CP): que mede a capacidade dos intervalos de confiança “capturarem” o valor original, cuja expressão é dada por:

$$\text{CP (\%)} = \frac{1}{R^*} \sum_{i=1}^{R^*} 1_{\{\beta_j^{or} \in \text{IC}_{1-\alpha}(\hat{\beta}_j)\}}$$

- (iii) quadrado médio do erro (QME): sua expressão é dada por:

$$\text{QME}_{\hat{\beta}_j} = \frac{1}{R^*} \sum_{i=1}^{R^*} (\hat{\beta}_{ij} - \beta_j^{or})^2$$

- (iv) erro-padrão das estimativas (ep): que é uma medida de variabilidade, equivalente ao QME, quando as estimativas são não-viesadas. Sua expressão é dada por: $\text{ep}(\hat{\beta}_j) = \frac{1}{R^*} \sum_{i=1}^{R^*} \text{ep}(\hat{\beta}_{ij})$;
- (v) coeficiente de determinação (R^2): pode ser visto como sendo uma medida de qualidade de ajuste, fornecendo a contribuição das variações devidos à regressão nas variações totais. Aqui, será calculada a média dos valores do (R^2) obtido para cada modelo ajustado. Assim, $\bar{R}^2 = \frac{1}{R^*} \sum_{i=1}^{R^*} R^2$, sendo R^* o número de reamostras e R^2 o coeficiente de determinação;
- Outras medidas de desempenho incluem: as probabilidades de significância (vulgo valor-p); entre outras.

4 Resultados

Nessa seção, serão apresentados os resultados descritivos obtidos do conjunto de dados descrito na subseção 4.1, assim como os modelos de regressão linear múltiplos para as duas variáveis dependentes do estudo.

4.1 Análises Descritivas

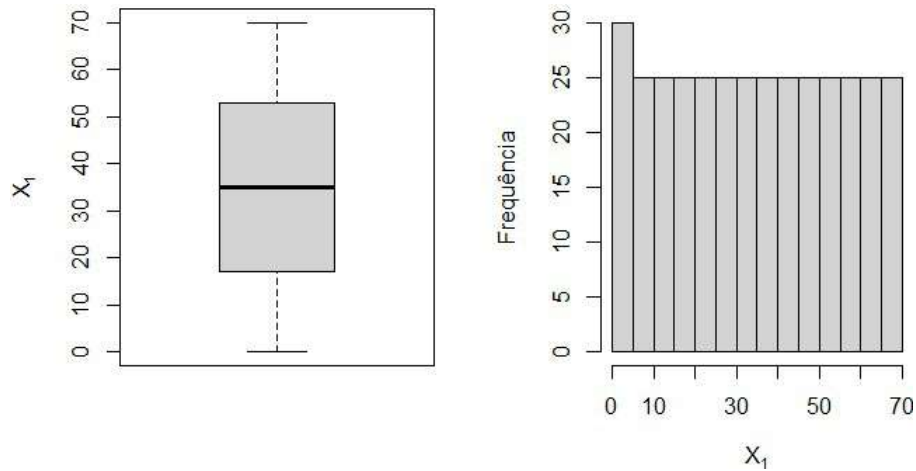
4.1.1 Variável X_1

A variável X_1 , a qual representa o tempo em minutos do processo de torrefação, é quantitativa discreta. Esta, por sua vez, vai de 0 à 70 minutos aonde cada minuto se repete 5 vezes, dado que foram realizados experimentos em 5 temperaturas diferentes. Através da Tabela 1, pode-se analisar suas medidas resumo.

Tabela 1: Medidas descritivas de X_1

Medida	Resultado
Mínimo	0.000
1º Quartil	17.000
Mediana	35.000
Média	35.000
3º Quartil	53.000
Máximo	70.000
Desvio Padrão	20.523

Além disso, a fim de visualizar melhor a maneira com que essa variável está distribuída, analisa-se a Figura 5.

Figura 5: Boxplot e Histograma de X_1 

A distribuição de X_1 é praticamente uniforme e tem maior densidade entre 0 e 5 minutos, tal características acontece em razão da repetição de 6 valores nesse intervalo, sendo que nos intervalos seguintes 5 valores se repetem 5 vezes cada. Assim, nele há uma frequência igual a 30 e nos outros igual a 25.

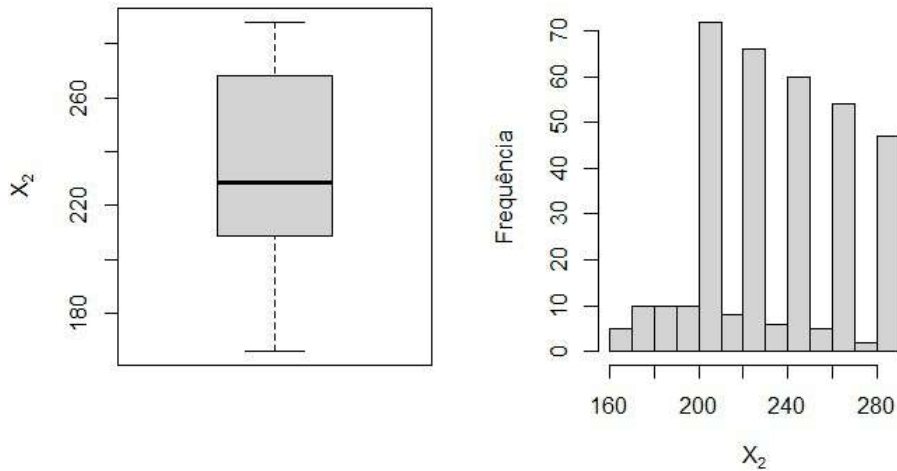
4.1.2 Variável X_2

A variável X_2 , a qual representa a temperatura em graus célcus do processo de torrefação, é quantitativa contínua. Esta, por sua vez, vai de 165.954 à 288.204 °C. Através da Tabela 2, pode-se analisar suas medidas resumo.

Tabela 2: Medidas descritivas de X_2

Medida	Resultado
Mínimo	165.954
1º Quartil	208.923
Mediana	228.704
Média	237.225
3º Quartil	268.140
Máximo	288.204
Desvio Padrão	32.108

Além disso, a fim de visualizar melhor a maneira com que essa variável está distribuída, analisa-se a Figura 6.

Figura 6: Boxplot e Histograma de X_2 

A distribuição de X_2 é assimétrica e possui picos de densidade em torno das 5 temperaturas desejadas dos experimentos realizados. Tal fato ocorre, pois, quando a biomassa é posta ao forno, ela ainda não se encontra na temperatura ideal, porém, quando esta é aproximadamente atingida, o forno estabiliza nela até o fim do tempo de experimento (70 minutos). Os picos de frequência decrescentes se dão por conta da rapidez que se atinge as menores temperaturas, com isso é natural que, conforme as temperaturas aumentem, existam menos observações nas estabilizações.

4.1.3 Variável X_3

A variável X_3 , a qual representa a temperatura desejada em graus célcus do processo de torrefação, é qualitativa ordinal. Esta, por sua vez, possui as temperaturas 210°C, 230°C, 250°C, 270°C e 290°C como categorias. Através da Tabela 3, pode-se analisar suas frequências.

Tabela 3: Tabela de frequência de X_3

Temperatura (°C)	Frequência Absoluta	Frequência Relativa
210	71	20%
230	71	20%
250	71	20%
270	71	20%
290	71	20%

Cada temperatura se repete 71 vezes, pois são as observações de cada experimento.

4.1.4 Variável Y_1

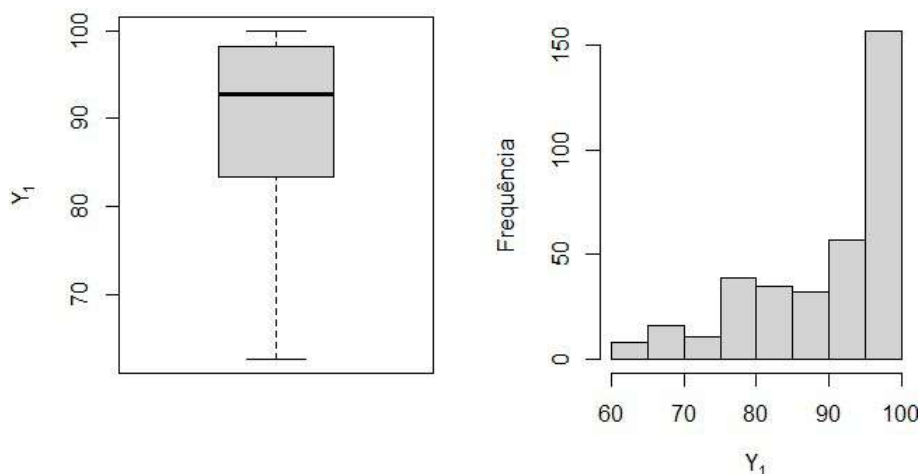
A variável Y_1 , a qual representa a porcentagem de rendimento de massa da madeira durante o processo de torrefação, é quantitativa contínua. Esta, por sua vez, vai de 62.636% à 100.000% e, através da Tabela 4, pode-se analisar suas medidas resumo.

Tabela 4: Medidas descritivas de Y_1

Medida	Resultado
Mínimo	62.636
1º Quartil	83.395
Mediana	92.726
Média	89.564
3º Quartil	98.137
Máximo	100.000
Desvio Padrão	10.230

Além disso, a fim de visualizar melhor a maneira com que essa variável está distribuída, analisa-se a Figura 7.

Figura 7: Boxplot e Histograma de Y_1



A distribuição de Y_1 é assimétrica à direita e possui um pico de densidade entre 95% e 100%.

4.1.5 Variável Y_2

A variável Y_2 , a qual representa o poder calorífico superior (capacidade energética) da madeira que está sendo torrefada, é quantitativa contínua. Esta, por sua vez, vai de

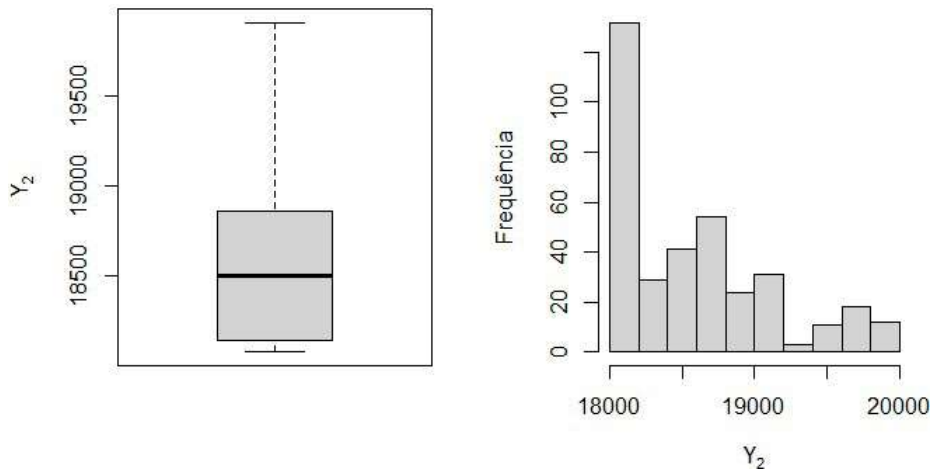
18075.447 $KJ \cdot Kg^{-1}$ à 19909.728 $KJ \cdot Kg^{-1}$ (Quilojoule por Quilograma) e, através da Tabela 5, pode-se analisar suas medidas resumo.

Tabela 5: Medidas descritivas de Y_2

Medida	Resultado
Mínimo	18075.447
1º Quartil	18139.598
Mediana	18496.816
Média	18595.720
3º Quartil	18859.864
Máximo	19909.728
Desvio Padrão	520.787

Além disso, a fim de visualizar melhor a maneira com que essa variável está distribuída, analisa-se a Figura 8.

Figura 8: Boxplot e Histograma de Y_2

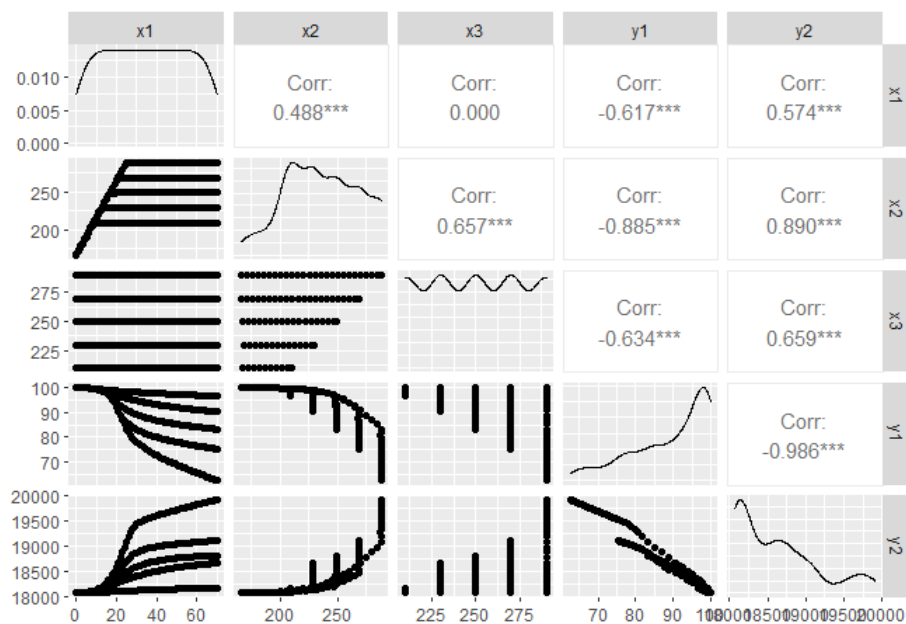


A distribuição de Y_2 é assimétrica à esquerda e possui um pico de densidade entre 18000 $KJ \cdot Kg^{-1}$ e 18250 $KJ \cdot Kg^{-1}$.

4.1.6 Correlações

Nesse momento, serão analisadas, por meio da matriz de correlações expressa na Figura 9, as correlações entre as variáveis estudadas. É importante ressaltar que, para essa análise, utilizarei a variável qualitativa X_3 como quantitativa discreta, sendo que terá suas próprias categorias de temperatura convertidas em números.

Figura 9: Matriz de correlações



Com base na Figura 9, é possível observar que as variáveis dependentes Y_1 e Y_2 estão fortemente e negativamente correlacionadas entre si, com uma correlação praticamente perfeita. Nesse sentido, interpreta-se que, conforme o rendimento físico da biomassa cai, o poder energético dela aumenta, ou vice-versa. Quanto a relação entre X_1 e as variáveis dependentes, nota-se que há uma correlação negativa e moderada entre ela e Y_1 , e positiva e moderada entre ela e Y_2 .

A respeito de X_2 , é visível uma correlação positiva moderada/alta com X_3 , o que era esperado pelo fato de ambas se tratarem da temperatura por diferentes pontos de vista. Porém, ao analisar a relação de X_2 com as variáveis dependentes, percebe-se que há uma correlação negativa e forte entre ela e Y_1 , e positiva e forte entre ela e Y_2 . Outrossim, observando os gráficos da diagonal da Figura 9, é importante ressaltar que nenhuma das variáveis se aproxima à distribuição normal, o que pode deixar as análises futuras mais complexas.

Na tentativa de verificar a relação de causa e efeito entre as variáveis independentes X_i , sendo $i = 1, 2$, e as dependentes Y_j , sendo $j = 1, 2$, nas próximas subseções serão apresentados os resultados do procedimento de reamostragem aplicados para a estimação dos seguintes modelos de regressão múltipla:

- Modelo I: Regressão entre Y_1 vs. X_1 e X_2 ;
- Modelo II: Regressão entre Y_2 vs. X_1 e X_2 .

4.2 Modelo I

4.2.1 Modelo Original

Para a variável Y_1 , que se trata do rendimento sólido em porcentagem da biomassa, considerou-se que sua variabilidade poderia estar em função do tempo e da temperatura do processo, dito isso, considerando a totalidade dos dados, foram estimados, por mínimos quadrados, os seguintes parâmetros:

Tabela 6: Avaliação das estimativas originais do Modelo I

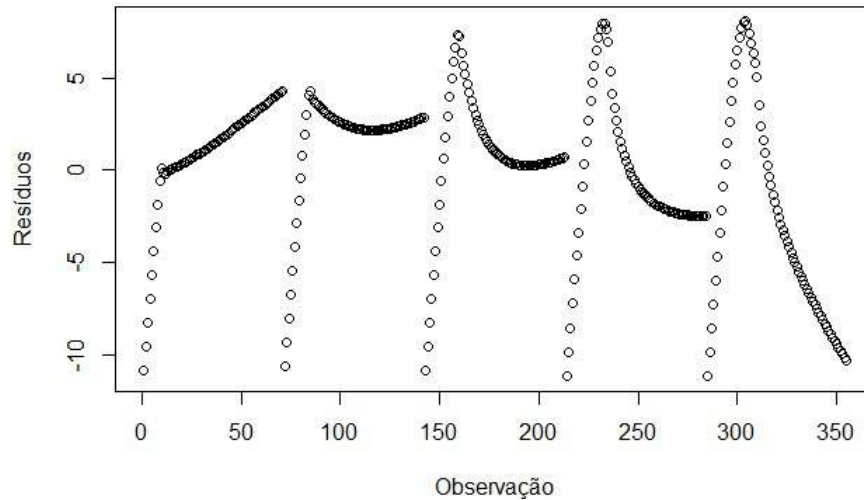
Coefficientes	Estimativa	Erro-padrão	t	p-valor
β_0	151.716	1.759	86.228	< 0.001
β_1	-0.121	0.013	-9.557	< 0.001
β_2	-0.244	0.008	-30.204	< 0.001

Dado que os p-valores para os testes t de significância de cada parâmetro do modelo resultaram em aproximadamente 0.00, pode-se dizer que, à um nível de significância de 5%, a hipótese de que os 3 parâmetros estimados são significantes não é rejeitada. Além disso, o R^2 ajustado resultou em 82.742%, o que indica que o modelo explica mais que 80% da variabilidade dos dados. Dessa maneira, pode-se interpretar os parâmetros estimados da seguinte forma:

- O rendimento sólido da biomassa começa em média em 151.716 unidades, quando, na teoria, o tempo de forno é de 0 minutos e a temperatura é de 0 °C, mesmo que na prática isso não aconteça.
- A cada minuto no forno, mantendo fixa a temperatura, o rendimento cai, em média, aproximadamente 0.121 unidades.
- E a cada grau célcium que a temperatura se eleva no forno, mantendo fixo o tempo, o rendimento cai, em média, aproximadamente 0.244 unidades.

Tendo em vista que o modelo parece explicar bem os dados, ainda é necessária a análise de seus resíduos para verificar se ele está de acordo com os pressupostos da regressão linear múltipla por mínimos quadrados.

Figura 10: Gráfico de dispersão dos Erros do Modelo I



É notável, por meio da Figura 10, que os resíduos não estão aleatoriamente distribuídos. Como para o cálculo do modelo as observações estavam ordenadas pela temperatura desejada (X_3) e depois pelo tempo (X_1), percebe-se que para cada temperatura desejada, os erros possuem uma tendência de acordo com o tempo.

Tabela 7: Resultado dos Testes para o Modelo I

Pressupostos	Testes	p-valor
Normalidade	Shapiro-Wilk	< 0.001
Independência dos erros	Durbin-Watson	< 0.001
Homogeneidade da variância	Breusch-Pagan	< 0.001

Para analisar se as suspeitas geradas pela Figura 10 são reais, analisa-se a Tabela 7. Tendo em vista os p-valores pequenos dos testes de hipóteses, à um nível de 5% de significância, pode-se rejeitar a normalidade, a independência e a homogeneidade da variância dos erros.

4.2.2 Métodos Bootstrap e Jackknife

Na subseção anterior, o modelo aparentou ser bem ajustado dados os seus parâmetros estimados serem significativos e também por seu alto valor de R^2 . Porém os resíduos não se comportaram de acordo com os pressupostos de um modelo de regressão linear, dado que foram rejeitadas a normalidade, a independência e a homogeneidade da variância deles. No entanto, na tentativa de contornar este problema, propõe-se um estudo de reamostragem para a estimação dos parâmetros por mínimos quadrados a fim de entender de maneira empírica suas características. Serão utilizados dois métodos de reamostragem,

o de Bootstrap e o de Jackknife.

Para aplicar a metodologia de Bootstrap foram sorteadas 1000 amostras aleatórias simples com 50% dos dados (178 observações), sem reposição, em seguida foram ajustados os modelos de regressão linear múltipla para cada uma delas. Nesse caso foram selecionadas reamostras sem reposição pois se trata de uma base de dados de um experimento controlado, como não desejamos que constem nas amostras resultados repetidos do experimento, faz-se o Bootstrap de reamostragem balanceada (sem reposição), Já para aplicar a reamostragem de Jackknife, foram feitas 355 amostras do banco de dados, cada uma dessas amostras com 354 dados, sempre retirando uma das observações, em seguida foram ajustados os modelos de regressão linear múltipla para cada uma delas, assim como feito no Bootstrap.

Tabela 8: Avaliação das estimativas de Bootstrap e Jackknife do Modelo I, onde β_j^{or} é o β_j original, $\hat{\beta}_j^{Btr}$ é a média das estimativas dos $\hat{\beta}_j$'s obtida usando o método de Bootstrap e $\hat{\beta}_j^{Jk}$ a média das estimativas dos $\hat{\beta}_j$'s obtida usando o método de Jackknife.

Método Bootstrap								
Coef.	β_j^{or}	$\hat{\beta}_j^{Btr}$	$ep(\hat{\beta}_j)$	t	$p - valor$	$CP(\%)$	Viés	QME
β_0	151.717	151.807	2.484	61.195	< 0.001	94.300%	0.090	6.041
β_1	-0.121	-0.121	0.018	-6.778	< 0.001	99.300%	0.000	0.000
β_2	-0.244	-0.245	0.011	-21.467	< 0.001	96.500%	0.000	0.000
Método Jacknife								
Coef.	β_j^{or}	$\hat{\beta}_j^{Jk}$	$ep(\hat{\beta}_j)$	t	$p - valor$	$CP(\%)$	Viés	QME
β_0	151.717	151.717	1.762	86.106	< 0.001	100.000%	0.000	0.015
β_1	-0.121	-0.121	0.013	-9.544	< 0.001	100.000%	0.000	0.000
β_2	-0.244	-0.244	0.008	-30.162	< 0.001	100.000%	0.000	0.000

Com a finalidade de comparar as estimativas de Bootstrap com as de Jackknife analisa-se a Tabela 8, e dessa maneira pode-se dizer que as estimativas dos parâmetros para ambos os métodos foram bem similares e significativas de acordo com o teste t , a um nível de significância de 5%. Na mesma tabela pode-se também visualizar que o único parâmetro que possui média de viés diferente de zero é o $\hat{\beta}_0$ no método de Bootstrap, mas mesmo assim essa média ainda é muito próxima a zero. É possível também observar que, as porcentagens de cobertura do intervalo de confiança deveriam, teoricamente, rondar em torno de 95%, já que o intervalo de confiança dos parâmetros dos modelos de Bootstrap são de 95%, o que acontece com clareza para o método de Bootstrap para $\hat{\beta}_0$ e $\hat{\beta}_2$, já para $\hat{\beta}_1$, a cobertura é superior à 99%. No caso do método Jackknife, como em cada reamostra é retirada apenas uma observação, não houve nenhuma delas influente o suficiente para

fazer com que o intervalo de confiança não englobasse os parâmetros originais, e dessa maneira, as coberturas dos ICs foram de 100%. Além disso, ainda nota-se que os Erros Quadráticos Médios estimados (QME) por Bootstrap e Jackknife de $\hat{\beta}_0$ são o únicos maiores que zero, com destaque para o estimado por Bootstrap que foi superior a 11.

A fim de comparar as distribuições de $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$, \hat{R}^2 e \bar{e} , analisa-se as Figuras 11, 12, 13, 14 e 15.

Figura 11: Histogramas e gráficos QQ das estimativas de $\hat{\beta}_0$ para o Modelo I

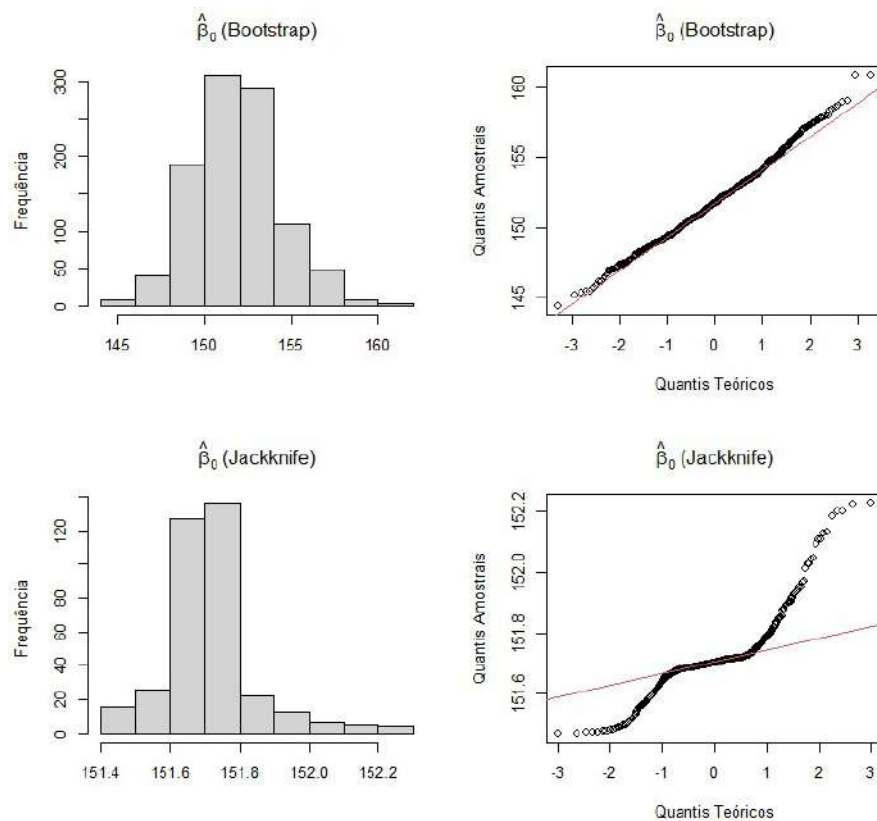


Figura 12: Histogramas e gráficos QQ das estimativas de $\hat{\beta}_1$ para o Modelo I

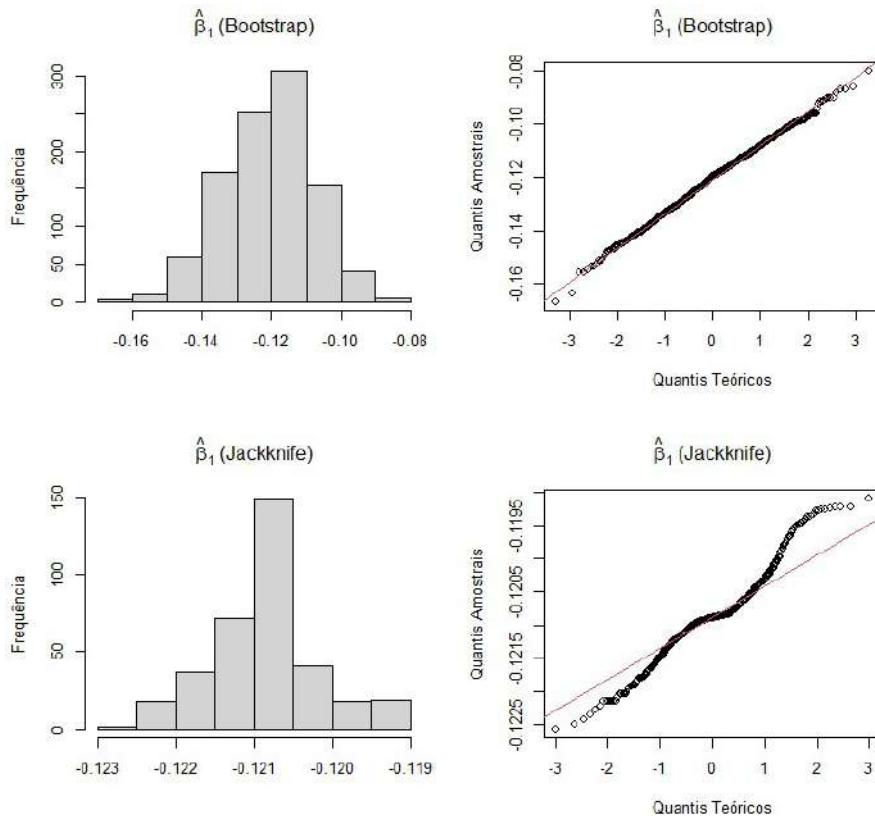


Figura 13: Histogramas e gráficos QQ das estimativas de $\hat{\beta}_2$ para o Modelo I

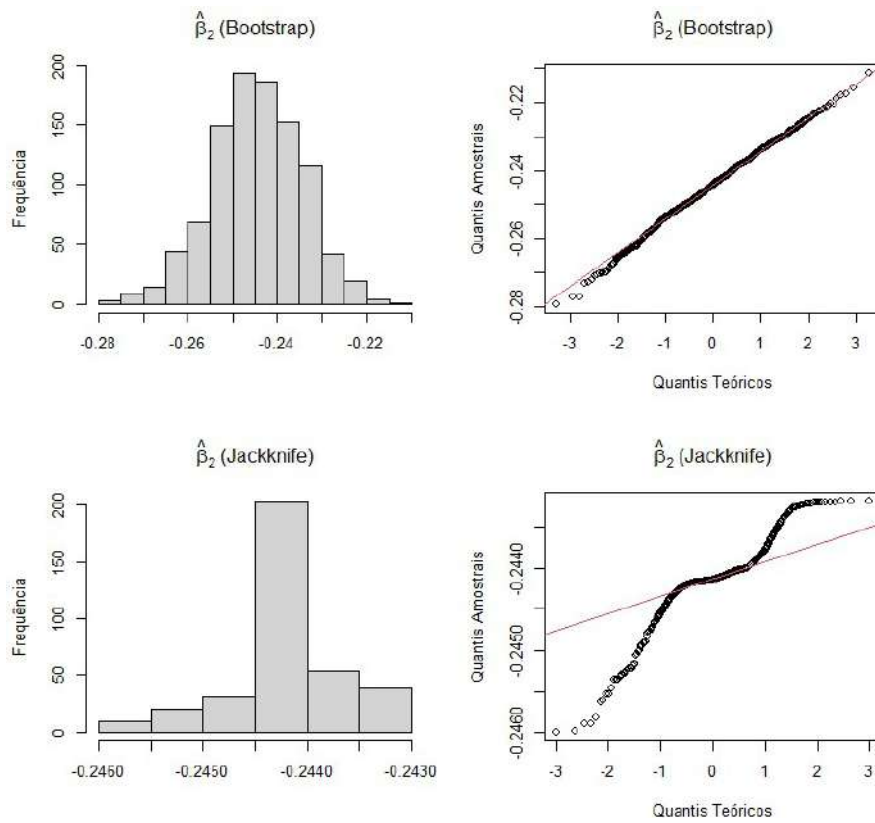


Figura 14: Histogramas e gráficos QQ das estimativas de \hat{R}^2 para o Modelo I

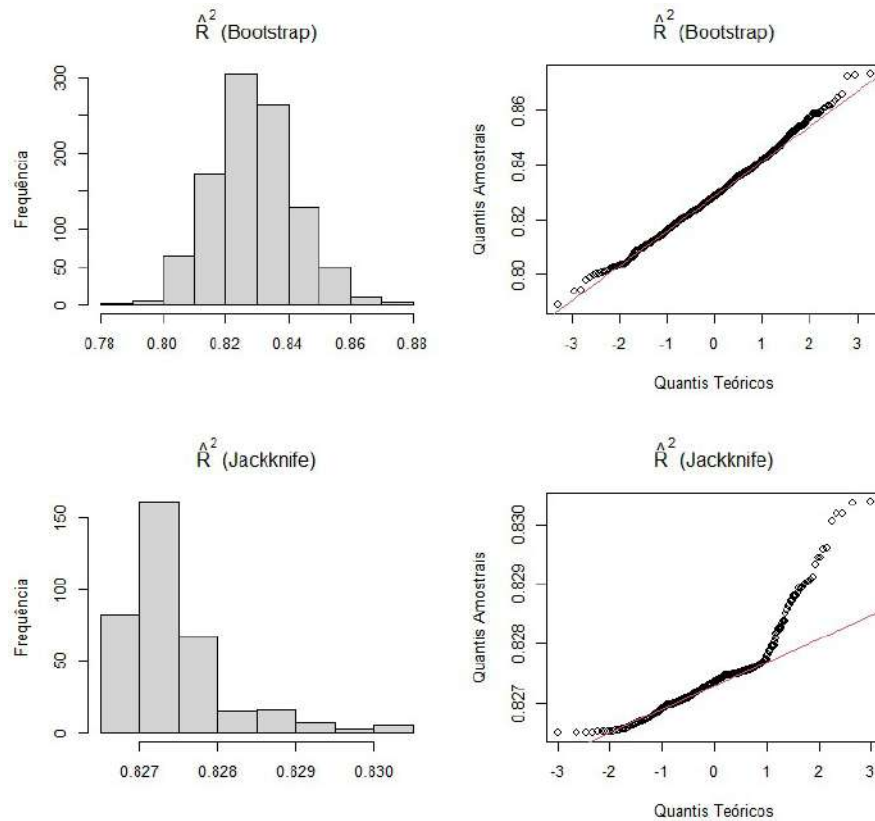
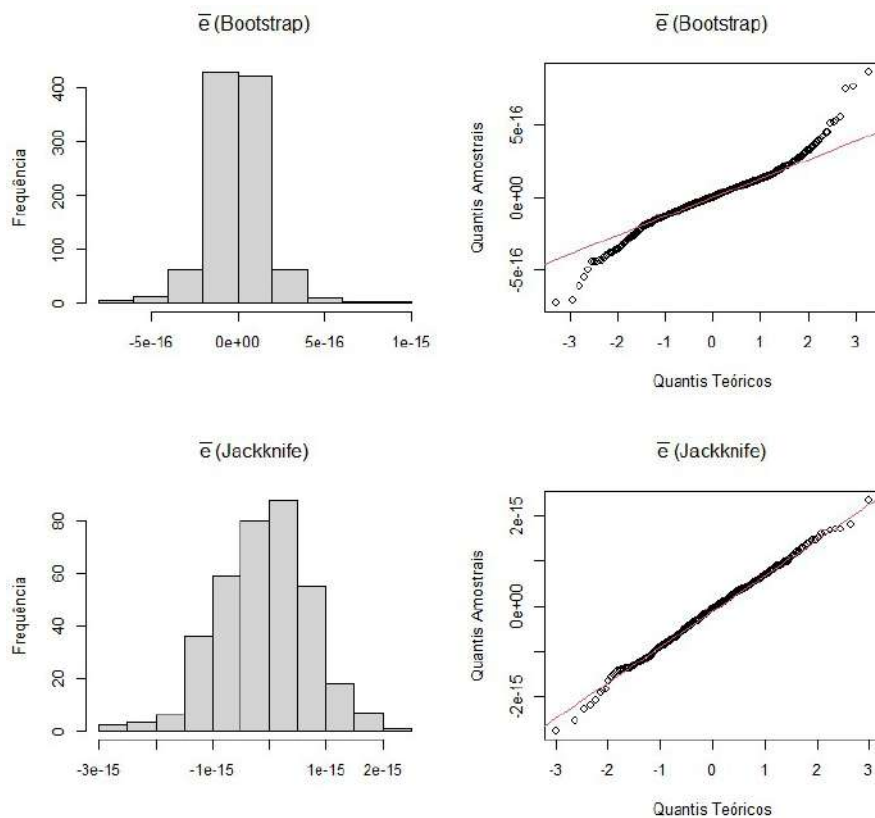


Figura 15: Histogramas e gráficos QQ das estimativas de \bar{e} para o Modelo I



Dados os histogramas e os gráficos QQ de normalidade das estimativas dos dois métodos de reamostragem expostos nas Figuras 11, 12, 13, 14 e 15, pode-se dizer que as médias dos resíduos estão bem distribuídas em torno do zero, porém no método de Bootstrap, esta é a única estimativa calculada que aparentemente foge bastante da distribuição normal, o que condiz com a análise da normalidade dos resíduos do modelo original. Já no caso de Jackknife, as médias dos resíduos aparentam estar distribuídas normalmente, tanto no histograma quanto no gráfico QQ, o que contraria a análise da normalidade do modelo original.

Acerca dos $\hat{\beta}_j$ e do \hat{R}^2 , no método de Bootstrap, ambos aparentam, graficamente, seguir uma distribuição normal, uma vez que nos gráficos QQ a maioria dos pontos estão sobre a reta, mesmo que na extremidade. Além disso, os histogramas estão aparentemente simétricos e concentrados em torno de um valor central. Por outro lado, para esses estimadores, no método de Jackknife, ocorre exatamente o contrário, de acordo com os histogramas, estão muito concentrados em um só intervalo, sendo também assimétricos.

O curioso da comparação entre os métodos é que a análise se inverte, o estimador que aparenta ser normal em um método, não aparenta no outro, e vice-versa. A fim de analisar melhor a normalidade dos resíduos, na Tabela 9 são expostos os p-valores resultantes dos testes de shapiro-wilk para \bar{e} dos dois métodos. Como era de se esperar, o teste rejeitou a normalidade das médias dos resíduos para o método de Bootstrap e não rejeitou para o método de Jackknife, à um nível de 5% de significância, concordando com as análises gráficas.

Tabela 9: P-valores dos testes de shapiro-wilk para \bar{e} das reamostragens do modelo I

Método	p-valor
Bootstrap	0.000
Jackknife	0.787

Tendo em vista as análises acima, é bem importante também, analisar o quanto as estimativas das reamostragens erraram, em relação aos parâmetros estimados no modelo original, e para isso, analisa-se as Figuras 16 e 17.

Figura 16: Boxplots dos Vieses das estimativas de Bootstrap do modelo I

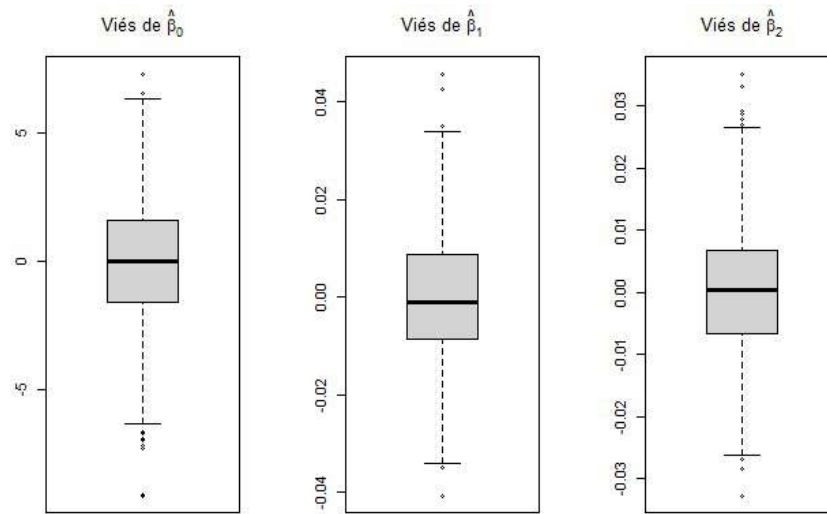
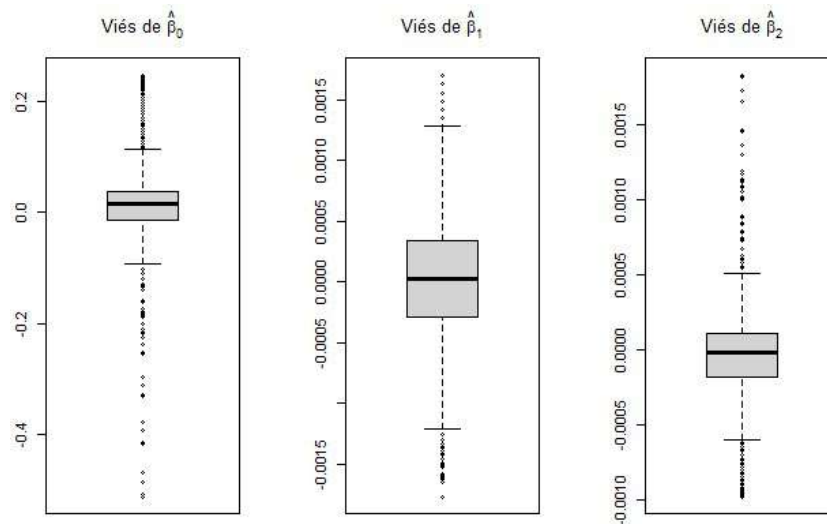


Figura 17: Boxplots dos Vieses das estimativas de Jackknife do modelo I



É possível notar, por meio dos boxplots das Figura 16 e 17 que $\hat{\beta}_0$ possui vieses em maior escala, mas que ainda assim são aparentemente simétricos e estão em torno de zero, assim como os vieses de $\hat{\beta}_1$ e de $\hat{\beta}_2$. A maior escala dos vieses de $\hat{\beta}_0$ acontece por ele se tratar do intercepto do modelo, sendo assim, não é um parâmetro estimado com base em uma variável, muitas vezes o que sobra da explicação da variabilidade da variável resposta, cai para ele, já que está "sozinho". Os outros parâmetros são estimados com base no efeito de uma variável independente sobre a dependente, o que gera menor escala de viés para eles.

O ocorrido citado acima é uma realidade para ambos os métodos, porém, o que os diferencia, é que no caso de Jackknife, existem diversos outliers para os vieses. Apesar de os boxplots estarem simétricos e em torno do zero, os outliers não estão simétricos para $\hat{\beta}_0$ e

$\hat{\beta}_2$, isso pode ocorrer devido ao cálculo do modelo ser feito sem observações influentes para a estimação destes parâmetros. Como no método de Bootstrap são sorteadas amostras aleatórias, sem retirar observações específicas, a ocorrência desses outliers de vieses é mais rara.

4.3 Modelo II

4.3.1 Modelo Original

Para a variável Y_2 , que representa o poder calorífico superior em $KJ \cdot kg^{-1}$ da biomassa em função do tempo e da temperatura do processo, foram estimados, por mínimos quadrados, os seguintes parâmetros:

Para a variável Y_2 , que representa o poder calorífico superior em $KJ \cdot kg^{-1}$ da biomassa, considerou-se que sua variabilidade poderia estar em função do tempo e da temperatura do processo, dito isso, considerando a totalidade dos dados, foram estimados, por mínimos quadrados, os seguintes parâmetros:

Tabela 10: Avaliação das estimativas originais do Modelo II

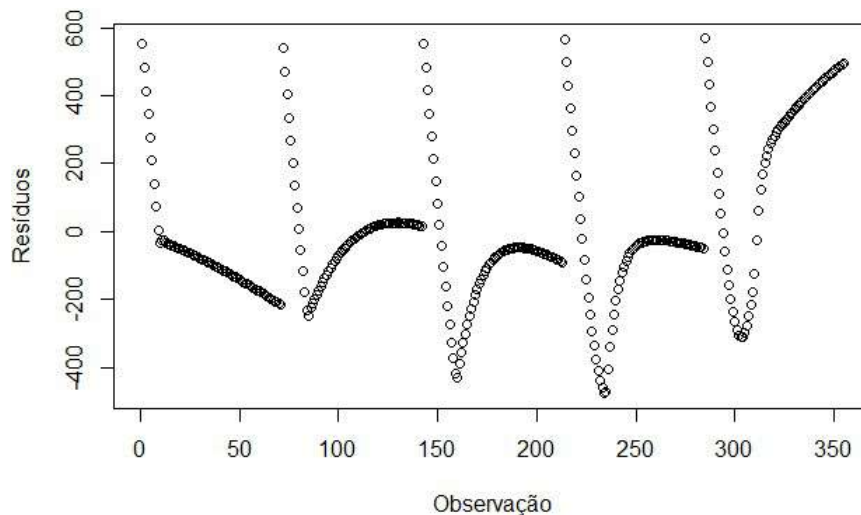
Coefficientes	Estimativa	Erro-padrão	t	valor-p
β_0	15354.262	92.189	166.552	< 0.001
β_1	4.659	0.663	7.030	< 0.001
β_2	12.977	0.424	30.638	< 0.001

Dado que os p-valores para os testes t de significância de cada parâmetro do modelo resultaram em aproximadamente 0.000, pode-se dizer que, à um nível de significância de 5%, a hipótese de que os 3 parâmetros estimados são significantes não é rejeitada. Além disso, o R^2 ajustado resultou em 81.718%, o que indica que o modelo explica mais que 80% da variabilidade dos dados. Dessa maneira, pode-se interpretar os parâmetros estimados da seguinte forma:

- O poder calorífico superior da biomassa começa, em média, em 15354.262 unidades, quando, na teoria, o tempo de forno é de 0 minutos e a temperatura é de 0 °C, mesmo que na prática isso não aconteça.
- A cada minuto no forno, mantendo fixa a temperatura, o poder calorífico superior aumenta, em média, aproximadamente 4.659 unidades.
- E cada grau célcio que a temperatura se eleva no forno, mantendo fixo o tempo, o poder calorífico superior aumenta, em média, aproximadamente 12.977 unidades.

Tendo em vista que o modelo parece explicar bem os dados, ainda é necessária a análise de seus resíduos para verificar se ele está de acordo com os pressupostos da regressão linear múltipla por mínimos quadrados.

Figura 18: Gráfico de dispersão dos Erros do Modelo II



É notável, por meio da Figura 18, que os resíduos não estão aleatoriamente distribuídos. Como para o cálculo do modelo as observações estavam ordenadas pela temperatura desejada (X_3) e depois pelo tempo (X_1), percebe-se que para cada temperatura desejada, os erros possuem uma tendência de acordo com o tempo.

Tabela 11: Resultado dos Testes para o Modelo II

Pressupostos	Testes	P-valor
Normalidade	Shapiro-Wilk	< 0,001
Independência dos erros	Durbin-Watson	< 0,001
Homogeneidade da variância	Breusch-Pagan	< 0,001

Para analisar se as suspeitas geradas pela Figura 18 são reais, analisa-se a Tabela 11. Tendo em vista os P-valores pequenos dos testes de hipóteses, à um nível de 5% de significância, pode-se rejeitar a normalidade, a independência e a homogeneidade da variância dos erros.

4.3.2 Métodos Bootstrap e Jackknife

O mesmo que aconteceu para o modelo I, voltou a ocorrer no modelo II, o que não gera espanto, dada a correlação quase perfeita entre Y_1 e Y_2 , dessa maneira, foi realizado o mesmo processo de reamostragem, com as mesmas 1000 amostras de Bootstrap e 355 de Jackknife sorteadas no Bootstrap do modelo I, mas agora utilizando Y_2 como variável

dependente.

Tabela 12: Avaliação das estimativas de Bootstrap e Jackknife do Modelo II, onde β_j^{or} é o β_j original, $\hat{\beta}_j^{Btr}$ é a média das estimativas dos $\hat{\beta}_j$'s obtida usando o método de Bootstrap e $\hat{\beta}_j^{Jk}$ a média das estimativas dos $\hat{\beta}_j$'s obtida usando o método de Jackknife.

Método Bootstrap								
Coef.	β_j^{or}	$\hat{\beta}_j^{Btr}$	$ep(\hat{\beta}_j)$	t	$p - valor$	$CP(\%)$	Viés	QME
β_0	15354.262	15352.331	130.101	118.143	< 0.001	94.900%	1.932	16355.140
β_1	4.659	4.667	0.934	5.002	< 0.001	99.400%	-0.008	0.398
β_2	12.977	12.981	0.598	21.757	< 0.001	96.300%	-0.004	0.294
Método Jackknife								
Coef.	β_j^{or}	$\hat{\beta}_j^{Jk}$	$ep(\hat{\beta}_j)$	t	$p - valor$	$CP(\%)$	Viés	QME
β_0	15354.262	15354.259	92.320	166.317	< 0.001	100.000%	0.003	43.725
β_1	4.659	4.659	0.664	7.021	< 0.001	100.000%	0.000	0.001
β_2	12.977	12.977	0.424	30.594	< 0.001	100.000%	0.000	0.001

Com a finalidade de comparar as estimativas de Bootstrap com as de Jackknife analisa-se a Tabela 12, e dessa maneira pode-se dizer que as estimativas dos parâmetros para ambos os métodos foram bem similares e significativas de acordo com o teste t , a um nível de significância de 5%. Na mesma tabela pode-se também visualizar que o único parâmetro que possui média de viés maior que 0.01 é o $\hat{\beta}_0$ no método de Bootstrap, mas mesmo assim essa média ainda é muito baixa, dada a grandeza dos dados. É possível também observar que, as porcentagens de cobertura do intervalo de confiança deveriam, teoricamente, rondar em torno de 95%, já que o intervalo de confiança dos parâmetros dos modelos de Bootstrap são de 95%, o que acontece com clareza para o método de Bootstrap para $\hat{\beta}_0$ e $\hat{\beta}_2$, já para $\hat{\beta}_1$, a cobertura é superior à 99%. No método de Jackknife, como em cada reamostra é retirada apenas uma observação, não houve nenhuma delas influente o suficiente para fazer com que o intervalo de confiança não englobasse os parâmetros originais, e dessa maneira, as coberturas dos ICs foram de 100%. Além disso, ainda nota-se que os Erros Quadráticos Médios estimados (QME) por Bootstrap e Jackknife de $\hat{\beta}_0$ se destacam perante os outros, com destaque maior ainda para o estimado por Bootstrap que foi superior a 16 mil.

A fim de comparar as distribuições de $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$, \hat{R}^2 e \bar{e} , analisa-se as Figuras 19, 20, 21, 22 e 23.

Figura 19: Histogramas e gráficos QQ das estimativas de $\hat{\beta}_0$ para o Modelo II

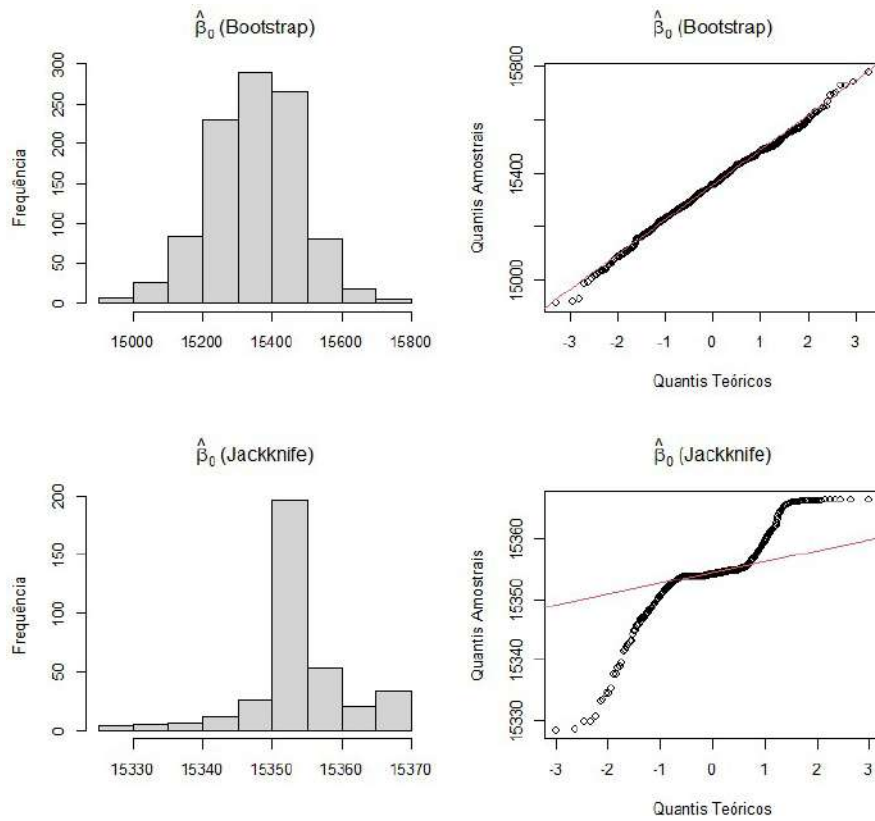


Figura 20: Histogramas e gráficos QQ das estimativas de $\hat{\beta}_1$ para o Modelo II

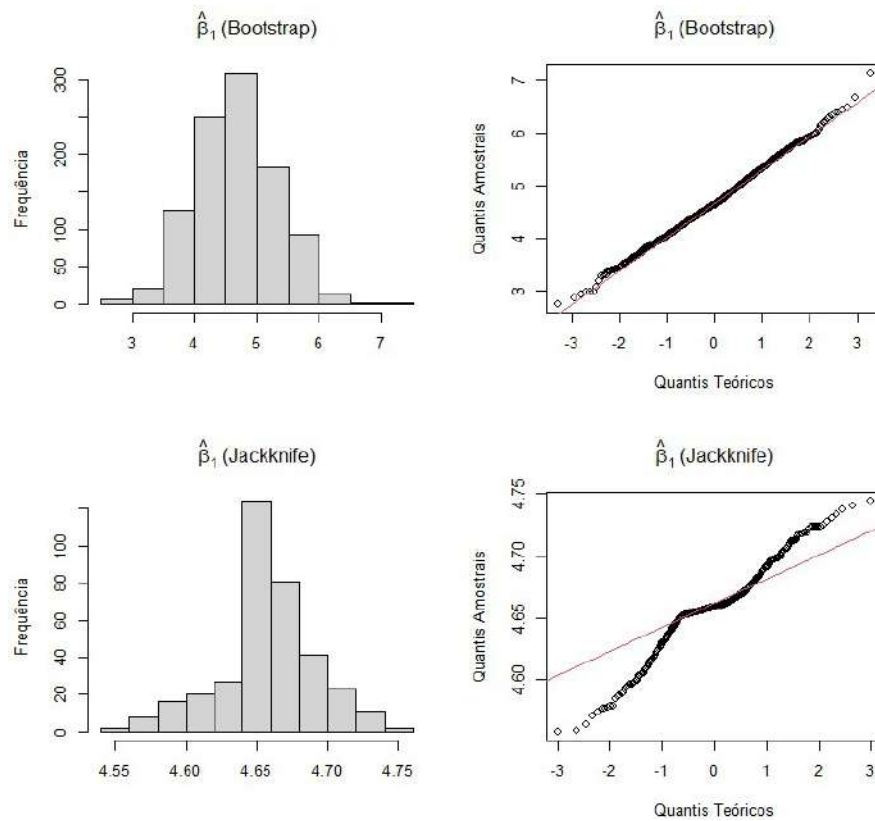


Figura 21: Histogramas e gráficos QQ das estimativas de $\hat{\beta}_2$ para o Modelo II

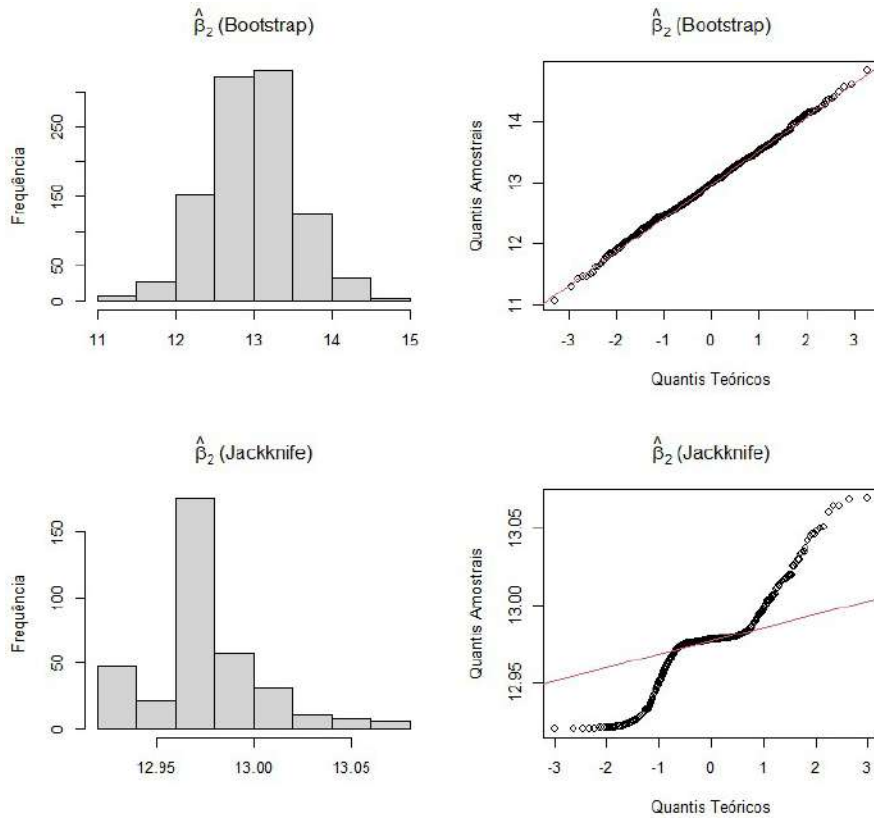


Figura 22: Histogramas e gráficos QQ das estimativas de \hat{R}^2 para o Modelo II

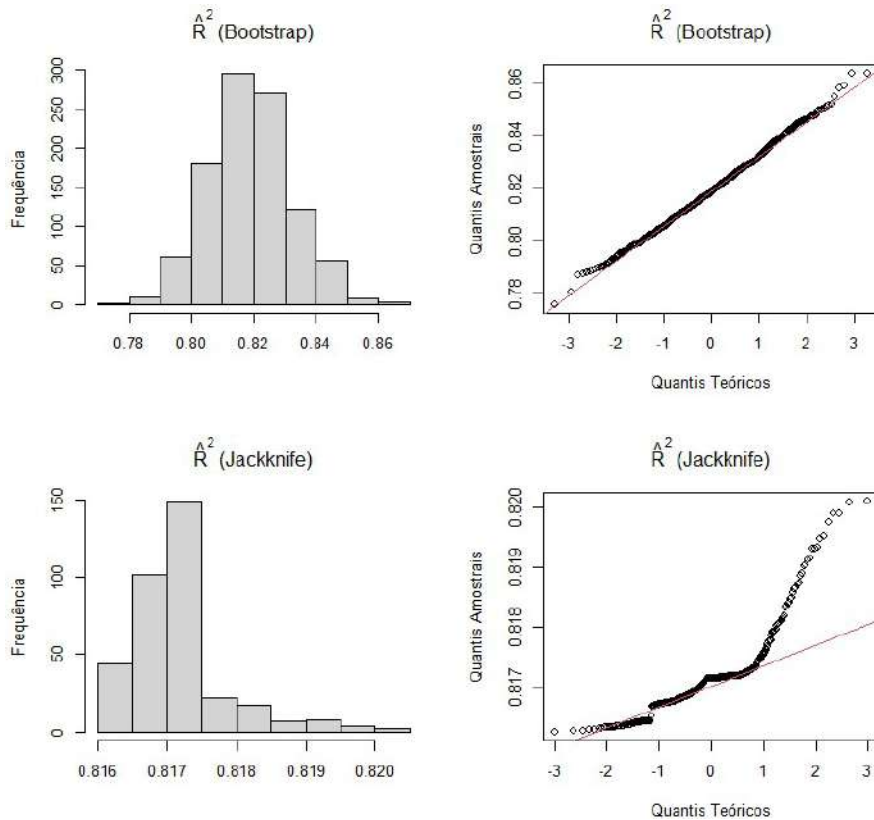
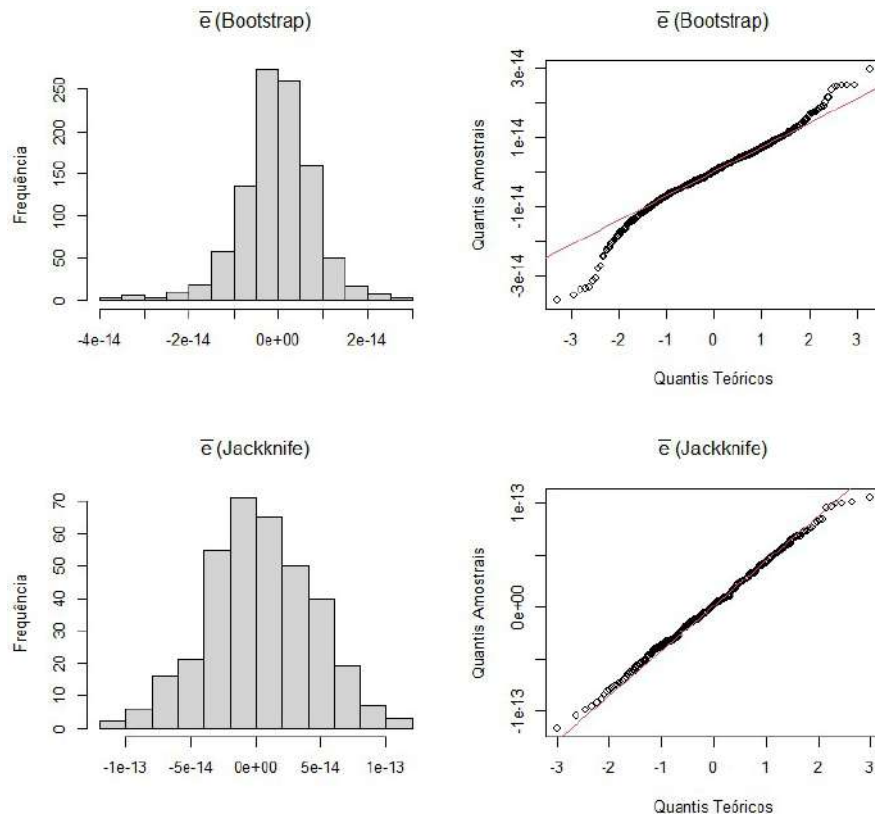


Figura 23: Histogramas e gráficos QQ das estimativas de \bar{e} para o Modelo II

Dados os Histogramas e gráficos QQ de normalidade das estimativas dos dois métodos de Bootstrap expostos nas Figuras 19, 20, 21, 22 e 23, pode-se dizer que as médias dos resíduos estão bem distribuídas em torno do zero, porém no método de Bootstrap, seu histograma pode dar a falsa impressão de que sua distribuição se aproxima de uma normal, mas seu gráfico QQ revela que nas extremidades, vários pontos se distanciam bem da linha ideal de normalidade, e dessa maneira, ela se torna, assim como no modelo I, a única estimativa calculada que aparentemente foge da distribuição normal, o que condiz com a análise da normalidade dos resíduos do modelo original. Já no método de Jackknife, as médias dos resíduos aparentam estar distribuídas normalmente, tanto no histograma quanto no gráfico QQ, o que contraria a análise da normalidade do modelo original.

Assim como no Modelo I, acerca dos $\hat{\beta}_j$ e do \hat{R}^2 , no método de Bootstrap, ambos aparentam, graficamente, seguir uma distribuição normal, uma vez que nos gráficos QQ a maioria dos pontos estão sobre a reta, mesmo que na extremidade. Além disso, os histogramas estão aparentemente simétricos e concentrados em torno de um valor central. Por outro lado, para esses estimadores, no método de Jackknife, ocorre exatamente o contrário, de acordo com os histogramas, estão muito concentrados em um só intervalo, sendo também assimétricos.

O curioso da comparação entre os métodos é que a análise se inverte, o estimador que aparenta ser normal em um método, não aparenta no outro, e vice-versa, exatamente

como ocorreu para o Modelo I. A fim de analisar melhor a normalidade dos resíduos, na Tabela 9 são expostos os p-valores resultantes dos testes de shapiro-wilk para \bar{e} dos dois métodos. Como era de se esperar, o teste rejeitou a normalidade das médias dos resíduos para o método de Bootstrap e não rejeitou para o método de Jackknife, à um nível de 5% de significância, concordando com as análises gráficas.

Tabela 13: P-valores dos testes de shapiro-wilk para \bar{e} das reamostragens do modelo II

Método	p-valor
Bootstrap	0.000
Jackknife	0.897

Tendo em vista as análises acima, é bem importante também, analisar o quanto as estimativas das reamostragens erraram, em relação aos parâmetros estimados no modelo original, e para isso, analisa-se as Figuras 24 e 25.

Figura 24: Boxplots dos Vieses das estimativas de Bootstrap do modelo II

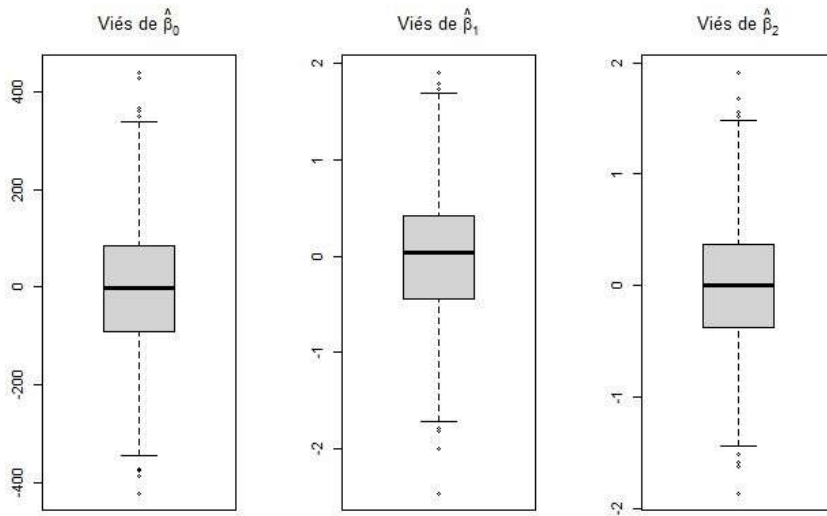
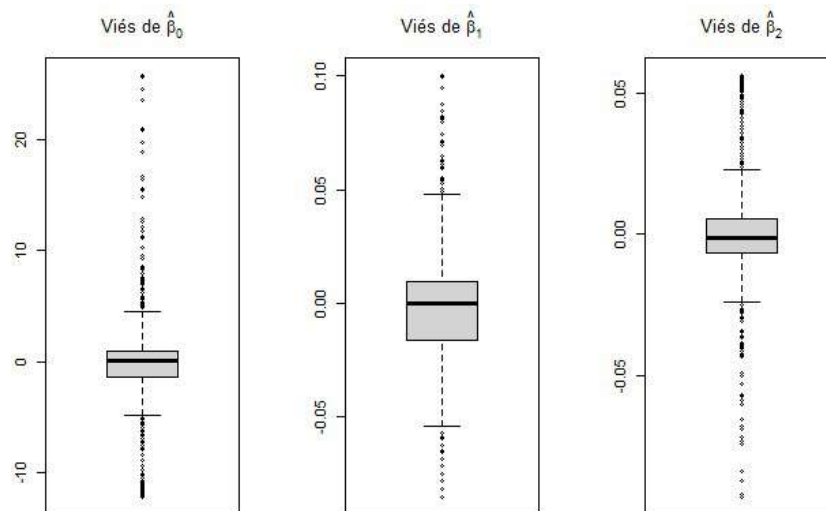


Figura 25: Boxplots dos Vieses das estimativas de Jackknife do modelo II



É possível notar, por meio dos boxplots das Figuras 24 e 25 que $\hat{\beta}_0$ possui vieses em maior escala, mas que ainda assim são aparentemente simétricos e estão em torno de zero, assim como os vieses de $\hat{\beta}_1$ e de $\hat{\beta}_2$. A maior escala dos vieses de $\hat{\beta}_0$ acontece por ele se tratar do intercepto do modelo, sendo assim, não é um parâmetro estimado com base em uma variável, muitas vezes o que sobra da explicação da variabilidade da variável resposta, cai para ele, já que está "sozinho". Os outros parâmetros são estimados com base no efeito de uma variável independente sobre a dependente, o que gera menor escala de viés para eles.

O ocorrido citado acima é uma realidade para ambos os métodos, porém, o que os diferencia, é que no caso de Jackknife, existem diversos outliers para os vieses. Apesar de os boxplots estarem simétricos e em torno do zero, os outliers não estão simétricos para $\hat{\beta}_0$ e $\hat{\beta}_2$, isso pode ocorrer devido ao cálculo do modelo ser feito sem observações influentes para a estimação destes parâmetros. Como no método de Bootstrap são sorteadas amostras aleatórias, sem retirar observações específicas, a ocorrência desses outliers de vieses é mais rara. O mesmo ocorreu no Modelo I, porém o que mudou foram os sinais da maioria dos outliers, o que era esperado, dada a correlação forte e negativa entre Y_1 e Y_2 .

5 Conclusão

Com base nos resultados obtidos através da aplicação dos métodos de reamostragem computacional, Bootstrap e Jackknife, para cálculo de modelos de regressão múltipla, avaliamos a consistência e o viés das estatísticas dos modelos, bem como o efeito dos parâmetros operacionais do processo de torrefação nos resultados de rendimento de massa e poder energético.

Conclui com o método de Bootstrap que as estimativas da influência do tempo no rendimento sólido e no poder energético da biomassa de Eucalipto são as mais consistentes, visto que independente da amostra selecionada, o tempo sempre influencia nesses quesitos de maneira muito parecida, com coberturas de intervalo de confiança acima de 99%. Os estimadores de intercepto e de temperatura também são consistentes, e geraram cobertura do intervalo de confiança entre 94% e 97%, o que é bem positivo também.

Por sua vez, o método de Jackknife proporcionou insights valiosos ao mostrar que não existiam observações específicas no banco de dados capazes de alterar significativamente as estimativas originais. Isso sugere uma certa estabilidade nos resultados, reforçando a confiabilidade das conclusões obtidas.

Como sugestão para trabalhos futuros, recomendo a separação dos experimentos com base em temperaturas específicas e a exploração da construção de modelos de série temporal. Os resíduos dos modelos desenvolvidos neste trabalho sugerem a presença de autorregressividade nos dados, o que indica a possibilidade de aprofundar a análise nesse sentido.

Em resumo, com base nos métodos de Bootstrap e Jackknife, concluímos que as estimativas da influência do tempo e da temperatura no rendimento sólido e no poder energético da biomassa de Eucalipto são consistentes, independentemente da amostra selecionada. Os resultados obtidos fornecem suporte para a compreensão do processo de torrefação e oferecem insights para a otimização dos parâmetros operacionais. Sugeriu-se ainda como trabalho futuro a exploração de modelos de série temporal para investigar a presença de autorregressividade nos dados.

Referências

- BUSSAB, W. O.; MORETTIN, P. A. *Estatística Básica*. [S.l.]: Saraiva, 2018.
- DA VEIGA, J. E. *O antropoceno e a Ciência do Sistema Terra*. [S.l.]: Editora 34, 2019.
- DAVISON, A. C.; HINKLEY, D. V. *Bootstrap Methods and their Application*. [S.l.]: Cambridge University Press, 1997.
- DEMÉTRIO, C. G. B.; ZOCCHI, S. S. Modelos de regressão. *Piracicaba: ESALQ*, 2006.
- EFRON, B. Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, Institute of Mathematical Statistics, v. 7, n. 1, p. 1 – 26, 1979. Disponível em: <https://doi.org/10.1214/aos/1176344552>.
- FISHER, R. A. *An Introduction to the Theory of Statistics*. [S.l.]: Oliver and Boyd, 1925. ISBN 978-0486600687.
- GELMAN, A. et al. *Bayesian Data Analysis*. 3rd. ed. [S.l.]: CRC Press, 2013.
- MATOS, R. A. D. S.; SOARES, F. K. *Balanço Energético Nacional*. [S.l.], 2022.
- NETER, J. et al. *Applied linear statistical models*. Irwin Chicago, 1996.
- R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2021. Disponível em: <https://www.R-project.org/>.
- SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete samples). *Biometrika*, Oxford University Press, v. 52, n. 3/4, p. 591–611, 1965.
- SILVEIRA, E. A. et al. Numerical modeling and experimental assessment of sustainable woody biomass torrefaction via coupled tg-ftir. *ELSEVIER*, 2021.
- VANDERPLAS, J. *Python Data Science Handbook: Essential Tools for Working with Data*. [S.l.]: O’Reilly Media, 2016.
- WICKHAM, H.; GROLEMUND, G. *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. [S.l.]: O’Reilly Media, 2017.
- WILCOX, R. R. *Introduction to Robust Estimation and Hypothesis Testing*. [S.l.]: Academic Press, 2017.

6 Apêndice

6.1 Banco de Dados

Tabela 14: Banco de Dados

	X_1	X_2	X_3	Y_1	Y_2
1	0.00	167.24	210.00	100.00	18075.45
2	1.00	172.20	210.00	99.98	18075.66
3	2.00	177.15	210.00	99.96	18075.93
4	3.00	182.09	210.00	99.94	18076.29
5	4.00	187.04	210.00	99.90	18076.74

351	66.00	287.87	290.00	63.83	19874.11
352	67.00	287.87	290.00	63.53	19883.21
353	68.00	287.87	290.00	63.23	19892.17
354	69.00	287.87	290.00	62.93	19901.02
355	70.00	287.87	290.00	62.64	19909.73

6.2 Códigos R

```
library(readxl)
```

```
library(latex2exp)
```

```
library(lmtest)
```

```
library(GGally)
```

```
library(xtable)
```

```
dados = read_excel("dados_tcc_joao.xlsx", sheet = 1)
xtable::xtable(dados)
```

```
# Análises descritivas
```

```
descritiva = function(x1, numvar){
  mr = round(c(min(x1),
               quantile(x1, 0.25),
               median(x1), mean(x1),
               quantile(x1, 0.75),
               max(x1),
               DP = sd(x1))),
```

```

4)
medidas = c("M nimo",
            "1 ¼Quartil",
            "Mediana",
            "M dia",
            "3 ¼Quartil",
            "M ximo",
            "Desvio ¼Padr o")
mr=data.frame(medidas, mr)
rownames(mr) = seq(1:7)
names(mr)=c("Medida", "Resultado")
print(xtable(mr, digits = c(0,0,3)),
      include.rownames=FALSE)
par(mfrow=c(1,2))
boxplot(x1, ylab = TeX(paste("$", numvar, "$")))
hist(x1, main=NULL, ylab = "Frequ ncia",
     xlab = TeX(paste("$", numvar, "$")))
}

descritiva(dados$x1, "X_1")

descritiva(dados$x2, "X_2")

descritiva(dados$y1, "Y_1")

descritiva(dados$y2, "Y_2")

matrizCorrelacao <- dados %>%
  ggpairs(columns = 1:ncol(dados),
          diag = list(continuous = "densityDiag",
                    discrete = "barDiag",
                    na = "naDiag"))
matrizCorrelacao

```

```
# Modelos Originais
```

```
m.ori_1 = lm(data=dados , y1~x1+x2)  
sm.ori_1= summary(m.ori_1)
```

```
b0.ori_1 = m.ori_1$coefficients [1]  
b1.ori_1 = m.ori_1$coefficients [2]  
b2.ori_1 = m.ori_1$coefficients [3]  
r2.ori_1 = sm.ori_1$r.squared  
res.ori_1 = mean(m.ori_1$residuals)
```

```
xtable::xtable(sm.ori_1$coefficients)
```

```
ci_1 = confint(m.ori_1)
```

```
shapiro.test(m.ori_1$residuals)  
dwtest(m.ori_1)  
bptest(m.ori_1)
```

```
par(mfrow=c(1,1))  
plot(m.ori_1$residuals ,  
      xlab = "Observação",  
      ylab = "Resíduos")
```

```
m.ori_2 = lm(data=dados , y2~x1+x2)  
sm.ori_2 = summary(m.ori_2)
```

```
b0.ori_2 = m.ori_2$coefficients [1]  
b1.ori_2 = m.ori_2$coefficients [2]  
b2.ori_2 = m.ori_2$coefficients [3]  
r2.ori_2 = sm.ori_2$r.squared  
res.ori_2 = mean(m.ori_2$residuals)
```

```
xtable::xtable(sm.ori_2$coefficients , digits = c(3,3,3,3,3))
```

```
ci_2 = confint(m.ori_2)
```

```
shapiro.test(m.ori_2$residuals)
```

```
dwtest(m.ori_2)
```

```
bptest(m.ori_2)
```

```
par(mfrow=c(1,1))
```

```
plot(m.ori_2$residuals ,  
      xlab = " Observa o" ,  
      ylab = " Res duos")
```

```
##### MC #####
```

```
## Modelos por Bootstrap Monte Carlo
```

```
n= 1000 # numero de reamostras
```

```
pct=0.5 # % do banco de dados para as reamostras
```

```
b0.hat_1=numeric()
```

```
b1.hat_1=numeric()
```

```
b2.hat_1=numeric()
```

```
r2.hat_1=numeric()
```

```
res.hat_1=numeric()
```

```
b0.ci.inf_1 = c()
```

```
b0.ci.sup_1 = c()
```

```
b1.ci.inf_1 = c()
```

```
b1.ci.sup_1 = c()
```

```
b2.ci.inf_1 = c()
```

```
b2.ci.sup_1 = c()
```

```
b0.vies_1 = c()
```

```
b1.vies_1 = c()
```

```
b2.vies_1 = c()
```

```
b0.eqm_1 = c()
```

```
b1.eqm_1 = c()
```

```
b2.eqm_1 = c()
```

```
b0.ep_1 = c()
```

```
b1.ep_1 = c()
```

```
b2.ep_1 = c()
```

```
b0.t_1 = c()
```

```
b1.t_1 = c()
```

```
b2.t_1 = c()
```

```
b0.pv_1 = c()
```

```
b1.pv_1 = c()
b2.pv_1 = c()

b0.hat_2=numeric()
b1.hat_2=numeric()
b2.hat_2=numeric()
r2.hat_2=numeric()
res.hat_2=numeric()
b0.ci.inf_2 = c()
b0.ci.sup_2 = c()
b1.ci.inf_2 = c()
b1.ci.sup_2 = c()
b2.ci.inf_2 = c()
b2.ci.sup_2 = c()
b0.vies_2 = c()
b1.vies_2 = c()
b2.vies_2 = c()
b0.eqm_2 = c()
b1.eqm_2 = c()
b2.eqm_2 = c()
b0.ep_2 = c()
b1.ep_2 = c()
b2.ep_2 = c()
b0.t_2 = c()
b1.t_2 = c()
b2.t_2 = c()
b0.pv_2 = c()
b1.pv_2 = c()
b2.pv_2 = c()

set.seed(180103229)
sementes = sample(1:10000, size = n, replace = F)

for(i in 1:n){
  set.seed(sementes[i])
  amostra<-sample(nrow(dados),
                 size = round(nrow(dados)*pct),
```

```
                                replace = F)
y1.<-dados$y1[amostra]
y2.<-dados$y2[amostra]
x1.<-dados$x1[amostra]
x2.<-dados$x2[amostra]

# Modelo 1
m_1 = lm(y1. ~ x1. + x2.)
sm_1 = summary(m_1)
b0.hat_1[i] <- m_1$coefficients [1]
b1.hat_1[i] <- m_1$coefficients [2]
b2.hat_1[i] <- m_1$coefficients [3]
res.hat_1[i] <- mean(m_1$residuals)
r2.hat_1[i] <- sm_1$r.squared

ci_1 = confint(m_1)
b0.ci.inf_1[i] = ci_1[1,1]
b0.ci.sup_1[i] = ci_1[1,2]
b1.ci.inf_1[i] = ci_1[2,1]
b1.ci.sup_1[i] = ci_1[2,2]
b2.ci.inf_1[i] = ci_1[3,1]
b2.ci.sup_1[i] = ci_1[3,2]

b0.vies_1[i] = m_1$coefficients [1] - b0.ori_1
b1.vies_1[i] = m_1$coefficients [2] - b1.ori_1
b2.vies_1[i] = m_1$coefficients [3] - b2.ori_1

b0.eqm_1[i] = (m_1$coefficients [1] - b0.ori_1)^2
b1.eqm_1[i] = (m_1$coefficients [2] - b1.ori_1)^2
b2.eqm_1[i] = (m_1$coefficients [3] - b2.ori_1)^2

b0.ep_1[i] = sm_1$coefficients [1,2]
b1.ep_1[i] = sm_1$coefficients [2,2]
b2.ep_1[i] = sm_1$coefficients [3,2]
b0.t_1[i] = sm_1$coefficients [1,3]
b1.t_1[i] = sm_1$coefficients [2,3]
b2.t_1[i] = sm_1$coefficients [3,3]
b0.pv_1[i] = sm_1$coefficients [1,4]
b1.pv_1[i] = sm_1$coefficients [2,4]
```

```

b2.pv_1[i] = sm_1$coefficients [3,4]

# Modelo 2
m_2 = lm(y2. ~ x1. + x2.)
sm_2 = summary(m_2)
b0.hat_2[i] <- m_2$coefficients [1]
b1.hat_2[i] <- m_2$coefficients [2]
b2.hat_2[i] <- m_2$coefficients [3]
res.hat_2[i] <- mean(m_2$residuals)
r2.hat_2[i] <- sm_2$r.squared

ci_2 = confint(m_2)
b0.ci.inf_2[i] = ci_2[1,1]
b0.ci.sup_2[i] = ci_2[1,2]
b1.ci.inf_2[i] = ci_2[2,1]
b1.ci.sup_2[i] = ci_2[2,2]
b2.ci.inf_2[i] = ci_2[3,1]
b2.ci.sup_2[i] = ci_2[3,2]

b0.vies_2[i] = b0.ori_2 - m_2$coefficients [1]
b1.vies_2[i] = b1.ori_2 - m_2$coefficients [2]
b2.vies_2[i] = b2.ori_2 - m_2$coefficients [3]

b0.eqm_2[i] = (m_2$coefficients [1] - b0.ori_2)^2
b1.eqm_2[i] = (m_2$coefficients [2] - b1.ori_2)^2
b2.eqm_2[i] = (m_2$coefficients [3] - b2.ori_2)^2

b0.ep_2[i] = sm_2$coefficients [1,2]
b1.ep_2[i] = sm_2$coefficients [2,2]
b2.ep_2[i] = sm_2$coefficients [3,2]
b0.t_2[i] = sm_2$coefficients [1,3]
b1.t_2[i] = sm_2$coefficients [2,3]
b2.t_2[i] = sm_2$coefficients [3,3]
b0.pv_2[i] = sm_2$coefficients [1,4]
b1.pv_2[i] = sm_2$coefficients [2,4]
b2.pv_2[i] = sm_2$coefficients [3,4]
}

```



```
##### JN #####  
## Modelos por Bootstrap Jackknife
```

```
b0.hat_1j=numeric()  
b1.hat_1j=numeric()  
b2.hat_1j=numeric()  
r2.hat_1j=numeric()  
res.hat_1j=numeric()  
b0.ci.inf_1j = c()  
b0.ci.sup_1j = c()  
b1.ci.inf_1j = c()  
b1.ci.sup_1j = c()  
b2.ci.inf_1j = c()  
b2.ci.sup_1j = c()  
b0.vies_1j = c()  
b1.vies_1j = c()  
b2.vies_1j = c()  
b0.eqm_1j = c()  
b1.eqm_1j = c()  
b2.eqm_1j = c()  
b0.ep_1j = c()  
b1.ep_1j = c()  
b2.ep_1j = c()  
b0.t_1j = c()  
b1.t_1j = c()  
b2.t_1j = c()  
b0.pv_1j = c()  
b1.pv_1j = c()  
b2.pv_1j = c()
```

```
b0.hat_2j=numeric()  
b1.hat_2j=numeric()  
b2.hat_2j=numeric()  
r2.hat_2j=numeric()  
res.hat_2j=numeric()  
b0.ci.inf_2j = c()  
b0.ci.sup_2j = c()  
b1.ci.inf_2j = c()
```

```

b1.ci.sup_2j = c()
b2.ci.inf_2j = c()
b2.ci.sup_2j = c()
b0.vies_2j = c()
b1.vies_2j = c()
b2.vies_2j = c()
b0.eqm_2j = c()
b1.eqm_2j = c()
b2.eqm_2j = c()
b0.ep_2j = c()
b1.ep_2j = c()
b2.ep_2j = c()
b0.t_2j = c()
b1.t_2j = c()
b2.t_2j = c()
b0.pv_2j = c()
b1.pv_2j = c()
b2.pv_2j = c()

for(i in 1:nrow(dados)){
  y1.<-dados$y1[-i]
  y2.<-dados$y2[-i]
  x1.<-dados$x1[-i]
  x2.<-dados$x2[-i]

  # Modelo 1
  m_1j = lm(y1. ~ x1. + x2.)
  sm_1j = summary(m_1j)
  b0.hat_1j[i] <- m_1j$coefficients[1]
  b1.hat_1j[i] <- m_1j$coefficients[2]
  b2.hat_1j[i] <- m_1j$coefficients[3]
  res.hat_1j[i] <- mean(m_1j$residuals)
  r2.hat_1j[i] <- sm_1j$r.squared

  ci_1j = confint(m_1j)
  b0.ci.inf_1j[i] = ci_1j[1,1]
  b0.ci.sup_1j[i] = ci_1j[1,2]
  b1.ci.inf_1j[i] = ci_1j[2,1]
  b1.ci.sup_1j[i] = ci_1j[2,2]

```

```

b2.ci.inf_1j[i] = ci_1j[3,1]
b2.ci.sup_1j[i] = ci_1j[3,2]

b0.vies_1j[i] = b0.ori_1 - m_1j$coefficients[1]
b1.vies_1j[i] = b1.ori_1 - m_1j$coefficients[2]
b2.vies_1j[i] = b2.ori_1 - m_1j$coefficients[3]

b0.eqm_1j[i] = (m_1j$coefficients[1] - b0.ori_1)^2
b1.eqm_1j[i] = (m_1j$coefficients[2] - b1.ori_1)^2
b2.eqm_1j[i] = (m_1j$coefficients[3] - b2.ori_1)^2

b0.ep_1j[i] = sm_1j$coefficients[1,2]
b1.ep_1j[i] = sm_1j$coefficients[2,2]
b2.ep_1j[i] = sm_1j$coefficients[3,2]
b0.t_1j[i] = sm_1j$coefficients[1,3]
b1.t_1j[i] = sm_1j$coefficients[2,3]
b2.t_1j[i] = sm_1j$coefficients[3,3]
b0.pv_1j[i] = sm_1j$coefficients[1,4]
b1.pv_1j[i] = sm_1j$coefficients[2,4]
b2.pv_1j[i] = sm_1j$coefficients[3,4]

# Modelo 2
m_2j = lm(y2. ~ x1. + x2.)
sm_2j = summary(m_2j)
b0.hat_2j[i] <- m_2j$coefficients[1]
b1.hat_2j[i] <- m_2j$coefficients[2]
b2.hat_2j[i] <- m_2j$coefficients[3]
res.hat_2j[i] <- mean(m_2j$residuals)
r2.hat_2j[i] <- sm_2j$r.squared

ci_2j = confint(m_2j)
b0.ci.inf_2j[i] = ci_2j[1,1]
b0.ci.sup_2j[i] = ci_2j[1,2]
b1.ci.inf_2j[i] = ci_2j[2,1]
b1.ci.sup_2j[i] = ci_2j[2,2]
b2.ci.inf_2j[i] = ci_2j[3,1]
b2.ci.sup_2j[i] = ci_2j[3,2]

```

```

b0.vies_2j[i] = b0.ori_2 - m_2j$coefficients[1]
b1.vies_2j[i] = b1.ori_2 - m_2j$coefficients[2]
b2.vies_2j[i] = b2.ori_2 - m_2j$coefficients[3]

b0.eqm_2j[i] = (m_2j$coefficients[1] - b0.ori_2)^2
b1.eqm_2j[i] = (m_2j$coefficients[2] - b1.ori_2)^2
b2.eqm_2j[i] = (m_2j$coefficients[3] - b2.ori_2)^2

b0.ep_2j[i] = sm_2j$coefficients[1,2]
b1.ep_2j[i] = sm_2j$coefficients[2,2]
b2.ep_2j[i] = sm_2j$coefficients[3,2]
b0.t_2j[i] = sm_2j$coefficients[1,3]
b1.t_2j[i] = sm_2j$coefficients[2,3]
b2.t_2j[i] = sm_2j$coefficients[3,3]
b0.pv_2j[i] = sm_2j$coefficients[1,4]
b1.pv_2j[i] = sm_2j$coefficients[2,4]
b2.pv_2j[i] = sm_2j$coefficients[3,4]
}

```

```
## Resultados Modelo com Y1 (Rendimento s lido de massa)
```

```
# an lise resumo bt
```

```

b0.ori.in.ci_1 = ifelse(b0.ori_1 >= b0.ci.inf_1 &
                        b0.ori_1 <= b0.ci.sup_1,
                        TRUE,FALSE)

b1.ori.in.ci_1 = ifelse(b1.ori_1 >= b1.ci.inf_1 &
                        b1.ori_1 <= b1.ci.sup_1,
                        TRUE,FALSE)

b2.ori.in.ci_1 = ifelse(b2.ori_1 >= b2.ci.inf_1 &
                        b2.ori_1 <= b2.ci.sup_1,
                        TRUE,FALSE)

b0.ori.in.ci_1j = ifelse(b0.ori_1 >= b0.ci.inf_1j &
                        b0.ori_1 <= b0.ci.sup_1j,
                        TRUE,FALSE)

```

```

b1.ori.in.ci_1j = ifelse (b1.ori_1 >= b1.ci.inf_1j &
                        b1.ori_1 <= b1.ci.sup_1j ,
                        TRUE,FALSE)
b2.ori.in.ci_1j = ifelse (b2.ori_1 >= b2.ci.inf_1j &
                        b2.ori_1 <= b2.ci.sup_1j ,
                        TRUE,FALSE)

resumobj = data.frame(
  M todo = c("Bootstrap", "Bootstrap", "Bootstrap",
            "Jackknife", "Jackknife", "Jackknife"),
  Coeficientes = c("$\\hat{\\beta}_0$", "$\\hat{\\beta}_1$",
                  "$\\hat{\\beta}_2$", "$\\hat{\\beta}_0$",
                  "$\\hat{\\beta}_1$", "$\\hat{\\beta}_2$"),
  originais = c(b0.ori_1, b1.ori_1,
                b2.ori_1, b0.ori_1,
                b1.ori_1, b2.ori_1),
  Estimativa = c(mean(b0.hat_1), mean(b1.hat_1),
                  mean(b2.hat_1), mean(b0.hat_1j),
                  mean(b1.hat_1j), mean(b2.hat_1j)),
  'Erro-padr o ' = c(mean(b0.ep_1), mean(b1.ep_1),
                      mean(b2.ep_1), mean(b0.ep_1j),
                      mean(b1.ep_1j), mean(b2.ep_1j)),
  '$t$' = c(mean(b0.t_1), mean(b1.t_1),
             mean(b2.t_1), mean(b0.t_1j),
             mean(b1.t_1j), mean(b2.t_1j)),
  'valor-p' = c(mean(b0.pv_1), mean(b1.pv_1),
                 mean(b2.pv_1), mean(b0.pv_1j),
                 mean(b1.pv_1j), mean(b2.pv_1j)),
  CP = c(mean(b0.ori.in.ci_1)*100, mean(b1.ori.in.ci_1)*100,
          mean(b2.ori.in.ci_1)*100, mean(b0.ori.in.ci_1j)*100,
          mean(b1.ori.in.ci_1j)*100, mean(b2.ori.in.ci_1j)*100),
  vies=c(mean(b0.vies_1), mean(b1.vies_1),
          mean(b2.vies_1), mean(b0.vies_1j),
          mean(b1.vies_1j), mean(b2.vies_1j)),
  qme = c(mean(b0.eqm_1), mean(b1.eqm_1),
           mean(b2.eqm_1), mean(b0.eqm_1j),
           mean(b1.eqm_1j), mean(b2.eqm_1j))
)

```

```
xtable(resumobj, digits = c(3,3,3,3,3,3,3,3,3,3,3))
```

```
### Boxplots dos Vi ses
```

```
par(mfrow=c(1,3))
boxplot(b0.vies_1,main= TeX("Vi s_de_\\hat{\\beta}_0$"))
boxplot(b1.vies_1,main= TeX("Vi s_de_\\hat{\\beta}_1$"))
boxplot(b2.vies_1,main= TeX("Vi s_de_\\hat{\\beta}_2$"))
boxplot(b0.vies_1j,main= TeX("Vi s_de_\\hat{\\beta}_0$"))
boxplot(b1.vies_1j,main= TeX("Vi s_de_\\hat{\\beta}_1$"))
boxplot(b2.vies_1j,main= TeX("Vi s_de_\\hat{\\beta}_2$"))
```

```
### Histogramas e QQplots das estimativas
```

```
par(mfrow=c(2,2))
hist(b0.hat_1,main = TeX("$\\hat{\\beta}_0$(Bootstrap)"),
      xlab = "", ylab = "Frequ ncia")
qqnorm(b0.hat_1,xlab = "Quantis_Te ricos",
        ylab = "Quantis_Amostrais",
        main = TeX("$\\hat{\\beta}_0$(Bootstrap)"))
qqline(b0.hat_1,col=2)

hist(b0.hat_1j,main = TeX("$\\hat{\\beta}_0$(Jackknife)"),
      xlab = "", ylab = "Frequ ncia")
qqnorm(b0.hat_1j,xlab = "Quantis_Te ricos",
        ylab = "Quantis_Amostrais",
        main = TeX("$\\hat{\\beta}_0$(Jackknife)"))
qqline(b0.hat_1j,col=2)
```

```
hist(b1.hat_1,main = TeX("$\\hat{\\beta}_1$(Bootstrap)"),
      xlab = "", ylab = "Frequ ncia")
qqnorm(b1.hat_1,xlab = "Quantis_Te ricos",
        ylab = "Quantis_Amostrais",
        main = TeX("$\\hat{\\beta}_1$(Bootstrap)"))
qqline(b1.hat_1,col=2)
```

```
hist(b1.hat_1j,main = TeX("$\\hat{\\beta}_1$(Jackknife)"),
      xlab = "", ylab = "Frequ ncia")
qqnorm(b1.hat_1j,xlab = "Quantis_Te ricos",
```

```

      ylab = "Quantis_Amostrais" ,
      main = TeX("$\\hat{\\beta}_1$( Jackknife)"))
qqline(b1.hat_1j , col=2)

```

```

hist(b2.hat_1 , main = TeX("$\\hat{\\beta}_2$( Bootstrap)")) ,
      xlab = "" , ylab = "Frequ ncia")
qqnorm(b2.hat_1 , xlab = "Quantis_Te ricos" ,
        ylab = "Quantis_Amostrais" ,
        main = TeX("$\\hat{\\beta}_2$( Bootstrap)"))
qqline(b2.hat_1 , col=2)

```

```

hist(b2.hat_1j , main = TeX("$\\hat{\\beta}_2$( Jackknife)")) ,
      xlab = "" , ylab = "Frequ ncia")
qqnorm(b2.hat_1j , xlab = "Quantis_Te ricos" ,
        ylab = "Quantis_Amostrais" ,
        main = TeX("$\\hat{\\beta}_2$( Jackknife)"))
qqline(b2.hat_1j , col=2)

```

```

hist(res.hat_1 , main = TeX("$\\bar{e}$( Bootstrap)")) ,
      xlab = "" , ylab = "Frequ ncia")
qqnorm(res.hat_1 , xlab = "Quantis_Te ricos" ,
        ylab = "Quantis_Amostrais" ,
        main = TeX("$\\bar{e}$( Bootstrap)"))
qqline(res.hat_1 , col=2)

```

```

hist(res.hat_1j , main = TeX("$\\bar{e}$( Jackknife)")) ,
      xlab = "" , ylab = "Frequ ncia")
qqnorm(res.hat_1j , xlab = "Quantis_Te ricos" ,
        ylab = "Quantis_Amostrais" ,
        main = TeX("$\\bar{e}$( Jackknife)"))
qqline(res.hat_1j , col=2)

```

```

hist(r2.hat_1 , main = TeX("$\\hat{R}^2$( Bootstrap)")) ,
      xlab = "" , ylab = "Frequ ncia")

```

```

qqnorm(r2.hat_1,xlab = "Quantis Te ricos",
       ylab = "Quantis Amostrais",
       main = TeX("$\\hat{R}^2$(Bootstrap)"))
qqline(r2.hat_1,col=2)

hist(r2.hat_1j,main = TeX("$\\hat{R}^2$(Jackknife)"),
     xlab = "", ylab = "Frequ ncia")
qqnorm(r2.hat_1j,xlab = "Quantis Te ricos",
       ylab = "Quantis Amostrais",
       main = TeX("$\\hat{R}^2$(Jackknife)"))
qqline(r2.hat_1j,col=2)

pvst = data.frame(round(c(
  shapiro.test(res.hat_1)$p.value,
  shapiro.test(res.hat_1j)$p.value
),3))

xtable::xtable(pvst)

## Resultados Modelo com Y2 (Poder calor fico)

# an lise resumo bt

b0.ori.in.ci_2 = ifelse(b0.ori_2 >= b0.ci.inf_2 &
                       b0.ori_2 <= b0.ci.sup_2,
                       TRUE,FALSE)
b1.ori.in.ci_2 = ifelse(b1.ori_2 >= b1.ci.inf_2 &
                       b1.ori_2 <= b1.ci.sup_2,
                       TRUE,FALSE)
b2.ori.in.ci_2 = ifelse(b2.ori_2 >= b2.ci.inf_2 &
                       b2.ori_2 <= b2.ci.sup_2,
                       TRUE,FALSE)

b0.ori.in.ci_2j = ifelse(b0.ori_2 >= b0.ci.inf_2j &
                       b0.ori_2 <= b0.ci.sup_2j,

```



```

                                TRUE,FALSE)
b1.ori.in.ci_2j = ifelse (b1.ori_2 >= b1.ci.inf_2j &
                        b1.ori_2 <= b1.ci.sup_2j ,
                        TRUE,FALSE)
b2.ori.in.ci_2j = ifelse (b2.ori_2 >= b2.ci.inf_2j &
                        b2.ori_2 <= b2.ci.sup_2j ,
                        TRUE,FALSE)

resumobj = data.frame(
  M todo = c("Bootstrap" ,"Bootstrap" ,"Bootstrap" ,
            "Jackknife" ,"Jackknife" ,"Jackknife" ),
  Coeficientes = c("$\\hat{\\beta}_0$" ,"$\\hat{\\beta}_1$" ,
                  "$\\hat{\\beta}_2$" ,"$\\hat{\\beta}_0$" ,
                  "$\\hat{\\beta}_1$" ,"$\\hat{\\beta}_2$" ),
  originais = c(b0.ori_2,
               b1.ori_2,
               b2.ori_2,
               b0.ori_2,
               b1.ori_2,
               b2.ori_2),
  Estimativa = c(mean(b0.hat_2),mean(b1.hat_2),mean(b2.hat_2),
                 mean(b0.hat_2j),mean(b1.hat_2j),mean(b2.hat_2j)),
  'Erro-padr o' = c(mean(b0.ep_2),mean(b1.ep_2),mean(b2.ep_2),
                    mean(b0.ep_2j),mean(b1.ep_2j),mean(b2.ep_2j)),
  '$t$' = c(mean(b0.t_2),mean(b1.t_2),mean(b2.t_2),mean(b0.t_2j),
            mean(b1.t_2j),mean(b2.t_2j)),
  'valor-p' = c(mean(b0.pv_2),mean(b1.pv_2),mean(b2.pv_2),
                mean(b0.pv_2j),mean(b1.pv_2j),mean(b2.pv_2j)),
  CP = c(mean(b0.ori.in.ci_2)*100,
         mean(b1.ori.in.ci_2)*100,
         mean(b2.ori.in.ci_2)*100,
         mean(b0.ori.in.ci_2j)*100,
         mean(b1.ori.in.ci_2j)*100,
         mean(b2.ori.in.ci_2j)*100),
  vies=c(mean(b0.vies_2),
         mean(b1.vies_2),
         mean(b2.vies_2),
         mean(b0.vies_2j),

```

```

        mean(b1.vies_2j),
        mean(b2.vies_2j)),
qme = c(mean(b0.eqm_2),
        mean(b1.eqm_2),
        mean(b2.eqm_2),
        mean(b0.eqm_2j),
        mean(b1.eqm_2j),
        mean(b2.eqm_2j))
)
xtable(resumobj, digits = c(3,3,3,3,3,3,3,3,3,3,3,10))

### Boxplots dos Vi ses
par(mfrow=c(1,3))
boxplot(b0.vies_2,main= TeX(" Vi s_de_\\hat{\\beta}_0$"))
boxplot(b1.vies_2,main= TeX(" Vi s_de_\\hat{\\beta}_1$"))
boxplot(b2.vies_2,main= TeX(" Vi s_de_\\hat{\\beta}_2$"))
boxplot(b0.vies_2j,main= TeX(" Vi s_de_\\hat{\\beta}_0$"))
boxplot(b1.vies_2j,main= TeX(" Vi s_de_\\hat{\\beta}_1$"))
boxplot(b2.vies_2j,main= TeX(" Vi s_de_\\hat{\\beta}_2$"))

### Histogramas e QQplots das estimativas
par(mfrow=c(2,2))
hist(b0.hat_2,main = TeX("$\\hat{\\beta}_0$(Bootstrap)"),
     xlab = "", ylab = "Frequ ncia")
qqnorm(b0.hat_2,xlab = "Quantis_Te ricos",
       ylab = "Quantis_Amostrais",
       main = TeX("$\\hat{\\beta}_0$(Bootstrap)"))
qqline(b0.hat_2,col=2)

hist(b0.hat_2j,main = TeX("$\\hat{\\beta}_0$(Jackknife)"),
     xlab = "", ylab = "Frequ ncia")
qqnorm(b0.hat_2j,xlab = "Quantis_Te ricos",
       ylab = "Quantis_Amostrais",
       main = TeX("$\\hat{\\beta}_0$(Jackknife)"))
qqline(b0.hat_2j,col=2)

hist(b1.hat_2,main = TeX("$\\hat{\\beta}_1$(Bootstrap)"),

```

```

      xlab = "", ylab = "Frequência")
qqnorm(b1.hat_2, xlab = "Quantis Teóricos",
       ylab = "Quantis Amostrais",
       main = TeX("$\\hat{\\beta}_1$(Bootstrap)"))
qqline(b1.hat_2, col=2)

hist(b1.hat_2j, main = TeX("$\\hat{\\beta}_1$(Jackknife)"),
     xlab = "", ylab = "Frequência")
qqnorm(b1.hat_2j, xlab = "Quantis Teóricos",
       ylab = "Quantis Amostrais",
       main = TeX("$\\hat{\\beta}_1$(Jackknife)"))
qqline(b1.hat_2j, col=2)

hist(b2.hat_2, main = TeX("$\\hat{\\beta}_2$(Bootstrap)"),
     xlab = "", ylab = "Frequência")
qqnorm(b2.hat_2, xlab = "Quantis Teóricos",
       ylab = "Quantis Amostrais",
       main = TeX("$\\hat{\\beta}_2$(Bootstrap)"))
qqline(b2.hat_2, col=2)

hist(b2.hat_2j, main = TeX("$\\hat{\\beta}_2$(Jackknife)"),
     xlab = "", ylab = "Frequência")
qqnorm(b2.hat_2j, xlab = "Quantis Teóricos",
       ylab = "Quantis Amostrais",
       main = TeX("$\\hat{\\beta}_2$(Jackknife)"))
qqline(b2.hat_2j, col=2)

hist(res.hat_2, main = TeX("$\\bar{e}$(Bootstrap)"),
     xlab = "", ylab = "Frequência")
qqnorm(res.hat_2, xlab = "Quantis Teóricos",
       ylab = "Quantis Amostrais",
       main = TeX("$\\bar{e}$(Bootstrap)"))
qqline(res.hat_2, col=2)

hist(res.hat_2j, main = TeX("$\\bar{e}$(Jackknife)"),

```

```

      xlab = "", ylab = "Frequência")
qqnorm(res.hat_2j, xlab = "Quantis Teóricos",
       ylab = "Quantis Amostrais",
       main = TeX("$\\bar{e}$ (Jackknife)"))
qqline(res.hat_2j, col=2)

hist(r2.hat_2, main = TeX("$\\hat{R}^2$ (Bootstrap)"),
     xlab = "", ylab = "Frequência")
qqnorm(r2.hat_2, xlab = "Quantis Teóricos",
       ylab = "Quantis Amostrais",
       main = TeX("$\\hat{R}^2$ (Bootstrap)"))
qqline(r2.hat_2, col=2)

hist(r2.hat_2j, main = TeX("$\\hat{R}^2$ (Jackknife)"),
     xlab = "", ylab = "Frequência")
qqnorm(r2.hat_2j, xlab = "Quantis Teóricos",
       ylab = "Quantis Amostrais",
       main = TeX("$\\hat{R}^2$ (Jackknife)"))
qqline(r2.hat_2j, col=2)

pvst = data.frame(round(c(
  shapiro.test(res.hat_2)$p.value,
  shapiro.test(res.hat_2j)$p.value
), 3))

xtable::xtable(pvst, digits = c(3,3))

```