

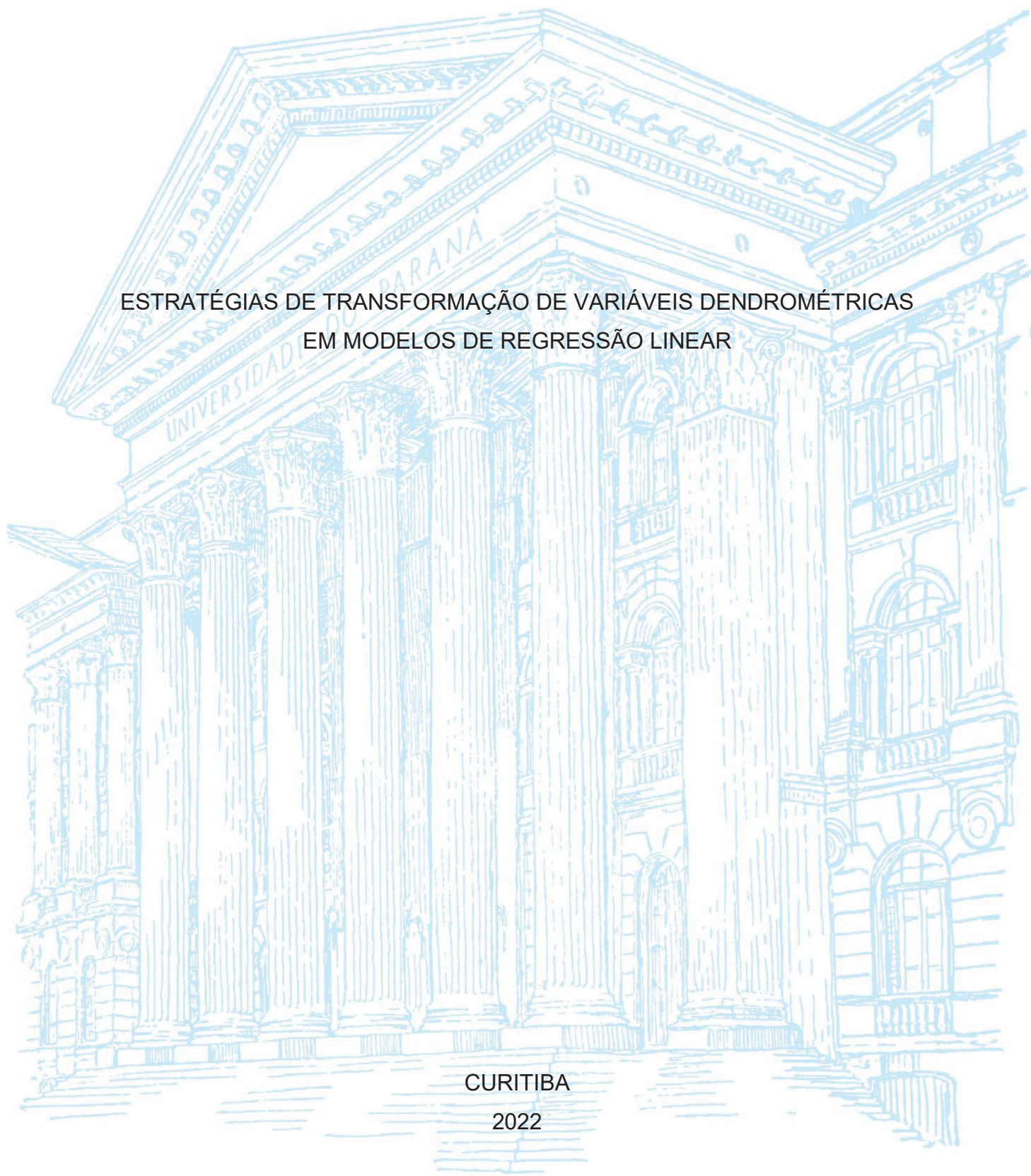
UNIVERSIDADE FEDERAL DO PARANÁ

JOBERT SILVA DA ROCHA

ESTRATÉGIAS DE TRANSFORMAÇÃO DE VARIÁVEIS DENDROMÉTRICAS  
EM MODELOS DE REGRESSÃO LINEAR

CURITIBA

2022



JOBERT SILVA DA ROCHA

ESTRATÉGIAS DE TRANSFORMAÇÃO DE VARIÁVEIS DENDROMÉTRICAS  
EM MODELOS DE REGRESSÃO LINEAR

Dissertação apresentada ao programa de Pós-Graduação em Engenharia Florestal, Área de concentração em Manejo Florestal, Setor de Ciências Agrárias, da Universidade Federal do Paraná, como requisito parcial à obtenção do título de Mestre em Engenharia Florestal.

Orientador:

Prof. Dr. Allan Libanio Pelissari

Coorientadores:

Prof. Dr. Luan Demarco Fiorentin

Prof. Dr. Luciano Rodrigo Lanssanova

CURITIBA

2022

Ficha catalográfica elaborada pela  
Biblioteca de Ciências Florestais e da Madeira - UFPR

Rocha, Jobert Silva da  
Estratégias de transformação de variáveis dendrométricas em  
modelos de regressão linear / Jobert Silva da Rocha. – Curitiba, 2022.  
1 recurso on-line : PDF.

Orientador: Prof. Dr. Allan Libanio Pelissari  
Coorientadores: Prof. Dr. Luan Demarco Fiorentin  
Prof. Dr. Luciano Rodrigo Lanssanova

Dissertação (Mestrado) - Universidade Federal do Paraná, Setor de  
Ciências Agrárias, Programa de Pós-Graduação em Engenharia  
Florestal. Defesa: Curitiba, 09/03/2022.  
Área de concentração: Manejo Florestal.

1. Florestas - Medição - Métodos estatísticos. 2. Dendrometria.  
3. Florestas - Medição - Modelos matemáticos. 4. Biomassa florestal.  
5. Teses. I. Pelissari, Allan Libanio. II. Fiorentin, Luan Demarco.  
III. Lanssanova, Luciano Rodrigo. IV. Universidade Federal do Paraná,  
Setor de Ciências Agrárias. V. Título.

CDD – 634.9  
CDU – 634.0.51

Bibliotecária: Berenice Rodrigues Ferreira – CRB 9/1160



MINISTÉRIO DA EDUCAÇÃO  
SETOR DE CIÊNCIAS AGRÁRIAS  
UNIVERSIDADE FEDERAL DO PARANÁ  
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO  
PROGRAMA DE PÓS-GRADUAÇÃO ENGENHARIA  
FLORESTAL - 40001016015P0

## TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação ENGENHARIA FLORESTAL da Universidade Federal do Paraná foram convocados para realizar a arguição da Dissertação de Mestrado de **JOBERT SILVA DA ROCHA** intitulada: **ESTRATÉGIAS DE TRANSFORMAÇÃO DE VARIÁVEIS DENDROMÉTRICAS EM MODELOS DE REGRESSÃO LINEAR**, sob orientação do Prof. Dr. ALLAN LIBANIO PELISSARI, que após terem inquirido o aluno e realizada a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de mestre está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 09 de Março de 2022.

Assinatura Eletrônica

10/03/2022 10:38:44.0

ALLAN LIBANIO PELISSARI

Presidente da Banca Examinadora

Assinatura Eletrônica

10/03/2022 09:25:43.0

ALEXANDRE BEHLING

Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

Assinatura Eletrônica

14/03/2022 11:01:10.0

ANGELO AUGUSTO EBLING

Avaliador Externo (UNIVERSIDADE FEDERAL RURAL DA AMAZÔNIA)

Dedico aos meus amados pais, José e Socorro.

## **AGRADECIMENTOS**

Em primeiro lugar, agradeço ao meu Deus, pelo dom da vida e pelas incontáveis bênçãos que me concedeu.

Com muito carinho, agradeço a toda minha família, em especial a minha mãe, Dona Socorro, pelo exemplo de humildade, perseverança e fé de que sempre dias melhores virão. Sem dúvidas a minha maior incentivadora.

Agradeço a minha amada avó, Dona Olívia, pelas doses diárias de carinho, amor e orações. Você foi fundamental para mim nessa etapa.

A minha namorada, Bruna Braga, por compartilhar essa experiência do mestrado junto comigo, dividindo alegrias, incertezas, experiências e sonhos. Sobretudo, por sempre me apoiar quando foi necessário.

Ao meu orientador, Dr. Allan Libanio Pelissari, não apenas pelas orientações ao longo desses dois anos, mas também pela oportunidade, confiança, compreensão, respeito e inspiração. Não poderia ter um orientador melhor.

Agradeço também aos meus coorientadores, Dr. Luan Demarco Fiorentin e Dr. Luciano Rodrigo Lansanova, pelo incentivo, disponibilidade e valiosas sugestões.

E, por fim, ao Conselho de Desenvolvimento Científico e Tecnológico – CNPq, por me conceder bolsa de estudos para o desenvolvimento desse projeto.



Sobre conquistas

## RESUMO

Na ciência florestal, a modelagem de variáveis dendrométricas representa um procedimento estatístico fundamental para estudos relacionados a produção de atributos biométricos. Nessa perspectiva, a técnica de análise de regressão linear tem sido amplamente utilizada nas mais diversas áreas de conhecimento por conta de sua praticidade, fácil compreensão e difusão consolidada. No entanto, para a sua adequada aplicação, um conjunto de premissas deve ser estritamente atendido. À vista disso, devido as características relacionadas à natureza de variáveis dendrométricas, as suposições da regressão linear são frequentemente violadas, dentre as quais destaca-se a distribuição normal dos resíduos e homogeneidade das variâncias, que são fundamentais para a confiabilidade da equação gerada e validade estatística dos testes  $t$  e  $F$  empregados no ajuste. Diante desses casos, as técnicas de transformação de variáveis em modelos lineares representam ferramentas promissoras, pois, ao passo que promovem o atendimento simultâneo das suposições exigidas, oportunizam ao analista de dados a permanecer no ambiente paramétrico de modelagem. Contudo, o entendimento de como as diferentes técnicas de transformação se comportam diante de variáveis dendrométricas para a adequação dos pressupostos paramétricos, ainda representa uma lacuna científica no âmbito florestal. Essa limitação, em diversas vezes, contribui para o uso indiscriminado de transformações específicas, dentre as quais destaca-se o método logarítmico. Diante disso, buscou-se com este estudo aplicar diferentes estratégias de transformação de variáveis em modelos de regressão linear simples, relacionados as previsões de altura, volume e biomassa em plantios e florestas naturais. Dessa forma, objetivou-se determinar as melhores estratégias de transformação para cada classe de modelagem dendrométrica, de modo que os modelos ajustados com variáveis transformadas sejam apropriados para previsão, bem como estatisticamente confiáveis. Nessa perspectiva, foram testadas distintas estratégias de transformação de variáveis: Log, recíproca, Box-Cox, Manly, Bickel-Doksum, Yeo-Johnson, Glog, Dual power, G power, Log shift, raiz quadrada deslocada e raiz quadrada. As modelagens com variáveis transformadas que atenderam simultaneamente as premissas de normalidade dos resíduos e homocedasticidade foram avaliadas estatisticamente para a seleção das mais adequadas. Assim, orienta-se a modelagem com transformação em situações na qual pelo menos um dos pressupostos é violado. Nessas ocasiões, em estudos de relação altura-diâmetro, recomenda-se as estratégias de transformações potenciais de Manly ou Box-Cox. Por outro lado, em estudos volumétricos, encoraja-se a aplicação do método logaritmo ou de Box-Cox. Por fim, para estudos alométricos da biomassa, quando se visa a previsão da biomassa total e de galhos, sugere-se a transformação logarítmica; quando o objetivo é a previsão das biomassas do fuste e das raízes, recomenda-se a transformação potencial de Box-Cox; ao passo que, para a alometria da biomassa de folhas e de galhos, incentiva-se o uso dos métodos de Manly e Yeo-Johnson, respectivamente.

Palavras-chave: alometria; homocedasticidade; modelagem linear; normalidade dos resíduos.



## ABSTRACT

In forest science, modeling tree variables represents a fundamental statistical procedure for studies related to the production of biometric attributes. In this perspective, linear regression analysis technique has been widely used in the most diverse areas of knowledge because of its practicality, easy understanding, and consolidated dissemination. However, for its adequate application, a set of premises must be strictly followed. Therefore, due to the characteristics related to the nature of tree variables, assumptions of linear regression are often violated, among which we highlight the normal distribution of residuals and homogeneity of variances. In these cases, techniques of variables' transformation in linear models represent promising tools, because, while promoting the simultaneous fulfillment of required assumptions, they provide the data analyst the opportunity to remain in parametric modeling environment. However, understanding of how different transformation techniques behave in relation to tree variables for parametric assumptions still represents a scientific gap in forestry field. This limitation often contributes to the indiscriminate use of specific transformations, among which the log method stands out. In view of this, this study aims to apply different variable transformation strategies in simple linear regression models, related to height, volume, and biomass predictions in plantations and natural forests. Thus, the aim was to determine the best transformation strategies for each class of tree modeling, in order that the models fitted with transformed variables are appropriate for prediction, as well as statistically reliable. In this perspective, different variable transformation strategies were tested: Log, reciprocal, Box-Cox, Manly, Bickel-Doksum, Yeo-Johnson, Glog, Dual power, G power, Log shift, shifted square root, and square root. Modeling with transformed variables that simultaneously met the assumptions of normality of residuals and homoscedasticity were statistically evaluated to select the most appropriate ones. Thus, modeling with transformation is oriented in situations in which at least one of the assumptions is violated. On these occasions, in height-diameter relationship studies, potential transformation strategies of Manly or Box-Cox are recommended. On the other hand, in volume studies, application of Log or Box-Cox method is encouraged. Finally, for allometric biomass studies, when the aim is to predict the total and branch biomass, Log transformation is suggested; when the objective is to predict stem and roots biomasses, potential transformation of Box-Cox is recommended; while for the allometry of leaves and branches biomasses, the use of Manly and Yeo-Johnson methods, respectively, is encouraged.

Keywords: allometry; homoscedasticity; linear modeling; normality of residues.

## LISTA DE FIGURAS

FIGURA 1 - MODELAGEM DA RELAÇÃO ALTURA-DIÂMETRO PARA <i>Pinus oocarpa</i> AOS 19 ANOS. ....	73
FIGURA 2 - ALTURAS OBSERVADAS VS ESTIMADAS, GRÁFICOS Q-Q E DISTRIBUIÇÃO RESIDUAL DAS MELHORES ESTRATÉGIAS DE MODELAGEM DA RELAÇÃO ALTURA-DIÂMETRO PARA <i>Pinus oocarpa</i> AOS 5 ANOS. ....	74
FIGURA 3 - GRÁFICOS HNP (HALF-NORMAL PLOT) DAS MODELAGENS DA RELAÇÃO ALTURA-DIÂMETRO PARA <i>Pinus oocarpa</i> AOS 5 ANOS. ....	75
FIGURA 4 - VOLUMES OBSERVADOS VS ESTIMADOS, GRÁFICOS Q-Q E DISTRIBUIÇÃO RESIDUAL DAS MELHORES ESTRATÉGIAS DE MODELAGEM VOLUMÉTRICA PARA <i>Pinus oocarpa</i> AOS 5 ANOS. ....	78
FIGURA 5 - GRÁFICOS HNP (HALF-NORMAL PLOT) DAS MODELAGENS VOLUMÉTRICAS PARA <i>Pinus oocarpa</i> AOS 5 ANOS. ....	79
FIGURA 6 - VOLUMES OBSERVADOS VS ESTIMADOS, GRÁFICOS Q-Q E DISTRIBUIÇÃO RESIDUAL DAS MELHORES ESTRATÉGIAS DE MODELAGEM VOLUMÉTRICA PARA <i>Pinus oocarpa</i> AOS 19 ANOS. ....	82
FIGURA 7 - GRÁFICOS HNP (HALF-NORMAL PLOT) DAS MODELAGENS VOLUMÉTRICAS PARA <i>Pinus oocarpa</i> AOS 19 ANOS. ....	83
FIGURA 8 - VOLUMES OBSERVADOS VS ESTIMADOS, GRÁFICOS Q-Q E DISTRIBUIÇÃO RESIDUAL DAS MELHORES ESTRATÉGIAS DE MODELAGEM VOLUMÉTRICA DE UMA AMOSTRAGEM DE ÁRVORES NATURAIS DA MATA ATLÂNTICA. ....	86
FIGURA 9 - GRÁFICOS HNP (HALF-NORMAL PLOT) DAS MODELAGENS VOLUMÉTRICAS PARA A UMA AMOSTRAGEM DE ÁRVORES DA MATA ATLÂNTICA. ....	87
FIGURA 10 - VOLUMES OBSERVADOS VS ESTIMADOS, GRÁFICOS Q-Q E DISTRIBUIÇÃO RESIDUAL DAS MELHORES ESTRATÉGIAS DE MODELAGEM VOLUMÉTRICA DE UMA AMOSTRA DE ÁRVORES NATURAIS DA FLORESTA AMAZÔNICA. ....	90
FIGURA 11 - GRÁFICOS HNP (HALF-NORMAL PLOT) DAS MODELAGENS VOLUMÉTRICAS DE UMA AMOSTRA DE ÁRVORES NATURAIS DA FLORESTA AMAZÔNICA. ....	91
FIGURA 12 - BIOMASSA TOTAL OBSERVADA VS ESTIMADA, GRÁFICOS Q-Q E DISTRIBUIÇÃO RESIDUAL DAS MELHORES ESTRATÉGIAS DE MODELAGEM DE BIOMASSA TOTAL DE <i>Mimosa scabrella</i> EM FUNÇÃO DO DIÂMETRO. ....	97
FIGURA 13 - GRÁFICOS HNP (HALF-NORMAL PLOT) DAS MODELAGENS DA BIOMASSA TOTAL DE ÁRVORES DE <i>Mimosa scabrella</i> EM FUNÇÃO DO DIÂMETRO. ....	98
FIGURA 14 - BIOMASSA DO FUSTE OBSERVADA VS ESTIMADA, GRÁFICOS Q-Q E DISTRIBUIÇÃO RESIDUAL E DAS MELHORES ESTRATÉGIAS DE	

MODELAGEM DE BIOMASSA DO FUSTE DE <i>Mimosa scrabella</i> EM FUNÇÃO DA ALTURA. ....	101
FIGURA 15 - GRÁFICOS HNP (HALF-NORMAL PLOT) DAS MODELAGENS DA BIOMASSA DO FUSTE DE ÁRVORES DE <i>Mimosa scrabella</i> EM FUNÇÃO DA ALTURA. ....	102
FIGURA 16 - BIOMASSA DO FUSTE OBSERVADA VS ESTIMADA, GRÁFICOS Q-Q E DISTRIBUIÇÃO RESIDUAL DAS MELHORES ESTRATÉGIAS DE MODELAGEM DE BIOMASSA DO FUSTE DE <i>Mimosa scrabella</i> EM FUNÇÃO DO DIÂMETRO. ....	103
FIGURA 17 - GRÁFICOS HNP (HALF-NORMAL PLOT) DAS MODELAGENS DA BIOMASSA DO FUSTE DE ÁRVORES DE <i>Mimosa scrabella</i> EM FUNÇÃO DO DIÂMETRO. ....	104
FIGURA 18 - BIOMASSA DOS GALHOS OBSERVADA VS ESTIMADA, GRÁFICOS Q-Q E DISTRIBUIÇÃO RESIDUAL DAS MELHORES ESTRATÉGIAS DE MODELAGEM DE BIOMASSA DOS GALHOS DE <i>MIMOSA SCRABELLA</i> EM FUNÇÃO DA ALTURA. ....	108
FIGURA 19 - GRÁFICOS HNP (HALF-NORMAL PLOT) DAS MODELAGENS DA BIOMASSA DOS GALHOS DE ÁRVORES DE <i>Mimosa scrabella</i> EM FUNÇÃO DA ALTURA. ....	109
FIGURA 20 - BIOMASSA DOS GALHOS OBSERVADA VS ESTIMADA, GRÁFICOS Q-Q E DISTRIBUIÇÃO RESIDUAL DAS MELHORES ESTRATÉGIAS DE MODELAGEM DE BIOMASSA DOS GALHOS DE <i>MIMOSA SCRABELLA</i> EM FUNÇÃO DO DIÂMETRO. ....	110
FIGURA 21 - GRÁFICOS HNP (HALF-NORMAL PLOT) DAS MODELAGENS DA BIOMASSA DOS GALHOS DE ÁRVORES DE <i>Mimosa scrabella</i> EM FUNÇÃO DO DIÂMETRO. ....	111
FIGURA 22 - BIOMASSA DE FOLHAS OBSERVADA VS ESTIMADA, GRÁFICOS Q-Q E DISTRIBUIÇÃO RESIDUAL DAS MELHORES ESTRATÉGIAS DE MODELAGEM DE BIOMASSA DAS FOLHAS DE <i>Mimosa scrabella</i> EM FUNÇÃO DA ALTURA. ....	114
FIGURA 23 - GRÁFICOS HNP (HALF-NORMAL PLOT) DAS MODELAGENS DA BIOMASSA DAS FOLHAS DE ÁRVORES DE <i>Mimosa scrabella</i> EM FUNÇÃO DA ALTURA. ....	115
FIGURA 24 - BIOMASSA DE FOLHAS OBSERVADA VS ESTIMADA, GRÁFICOS Q-Q E DISTRIBUIÇÃO RESIDUAL DAS MELHORES ESTRATÉGIAS DE MODELAGEM DE BIOMASSA DE FOLHAS DE <i>Mimosa scrabella</i> EM FUNÇÃO DO DIÂMETRO. ....	116
FIGURA 25 - GRÁFICOS HNP (HALF-NORMAL PLOT) DAS MODELAGENS DA BIOMASSA DAS FOLHAS DE ÁRVORES DE <i>Mimosa scrabella</i> EM FUNÇÃO DA DIÂMETRO. ....	117
FIGURA 26 - BIOMASSA DA CASCA OBSERVADA VS ESTIMADA, GRÁFICOS Q-Q E DISTRIBUIÇÃO RESIDUAL E DAS MELHORES ESTRATÉGIAS DE MODELAGEM DE BIOMASSA DA CASCA DE <i>Mimosa scrabella</i> EM FUNÇÃO DA ALTURA. ....	120

FIGURA 27 - GRÁFICOS HNP (HALF-NORMAL PLOT) DAS MODELAGENS DA BIOMASSA DA CASCA DE ÁRVORES DE <i>Mimosa scrabella</i> EM FUNÇÃO DA ALTURA. ....	121
FIGURA 28 - BIOMASSA DA CASCA OBSERVADA VS ESTIMADA, GRÁFICOS Q-Q E DISTRIBUIÇÃO RESIDUAL DAS MELHORES ESTRATÉGIAS DE MODELAGEM DE BIOMASSA DA CASCA DE <i>Mimosa scrabella</i> EM FUNÇÃO DO DIÂMETRO. ....	122
FIGURA 29 - GRÁFICOS HNP (HALF-NORMAL PLOT) DAS MODELAGENS DA BIOMASSA DA CASCA DE ÁRVORES DE <i>Mimosa scrabella</i> EM FUNÇÃO DO DIÂMETRO. ....	123
FIGURA 30 - BIOMASSA DA RAIZ OBSERVADA VS ESTIMADA, GRÁFICOS Q-Q E DISTRIBUIÇÃO RESIDUAL DAS MELHORES ESTRATÉGIAS DE MODELAGEM DE BIOMASSA DA RAIZ DE <i>Mimosa scrabella</i> EM FUNÇÃO DA ALTURA. ....	126
FIGURA 31 - GRÁFICOS HNP (HALF-NORMAL PLOT) DAS MODELAGENS DA BIOMASSA DAS RAÍZES DE ÁRVORES DE <i>Mimosa scrabella</i> EM FUNÇÃO DA ALTURA. ....	127
FIGURA 32 - BIOMASSA DA RAIZ OBSERVADA VS ESTIMADA, GRÁFICOS Q-Q E DISTRIBUIÇÃO RESIDUAL DAS MELHORES ESTRATÉGIAS DE MODELAGEM DE BIOMASSA DA RAIZ DE <i>Mimosa scrabella</i> EM FUNÇÃO DO DIÂMETRO. ....	128
FIGURA 33 - GRÁFICOS HNP (HALF-NORMAL PLOT) DAS MODELAGENS DA BIOMASSA DAS RAÍZES DE ÁRVORES DE <i>Mimosa scrabella</i> EM FUNÇÃO DO DIÂMETRO. ....	129

## LISTA DE TABELAS

TABELA 1 - ESTATÍSTICA DESCRITIVA DAS VARIÁVEIS DENDROMÉTRICAS DE POVOAMENTOS EQUIÂNEOS DE <i>Pinus oocarpa</i> UTILIZADAS NAS MODELAGENS VOLUMÉTRICAS E HIPSOMÉTRICAS. ....	41
TABELA 2 - ESTATÍSTICA DESCRITIVA DAS VARIÁVEIS DENDROMÉTRICAS DAP E VOLUME DE UMA FLORESTA NATURAL DA AMAZÔNIA.....	42
TABELA 3 - ESTATÍSTICA DESCRITIVA DAS VARIÁVEIS DENDROMÉTRICAS DAP E VOLUME DE UMA FLORESTA NATURAL DA MATA ATLÂNTICA.....	43
TABELA 4 - ESTATÍSTICA DESCRITIVA DAS VARIÁVEIS DENDROMÉTRICAS DAP E ALTURA DE UMA AMOSTRAGEM DE ÁRVORES DE <i>Mimosa scrabella</i> USADAS PARA MODELAGENS DE BIOMASSA TOTAL. ....	44
TABELA 5 - ESTATÍSTICA DESCRITIVA DA BIOMASSA DOS COMPONENTES FUSTE, GALHOS, FOLHAS, CASCA E RAÍZES DE ÁRVORES DE <i>Mimosa scrabella</i> .....	46
TABELA 6 - ESTRATÉGIAS DE MODELAGEM LINEAR SIMPLES ENVOLVENDO TRANSFORMAÇÕES DE VARIÁVEIS.....	47
TABELA 7 - ESTATÍSTICAS DE REGRESSÃO DAS MELHORES ESTRATÉGIAS DE MODELAGEM DA RELAÇÃO ALTURA-DIÂMETRO PARA <i>Pinus oocarpa</i> AOS 5 ANOS.....	72
TABELA 8 - COEFICIENTES DE REGRESSÃO E PARÂMETROS DE TRANSFORMAÇÃO DAS MELHORES ESTRATÉGIAS DE MODELAGEM HIPSOMÉTRICA PARA <i>Pinus oocarpa</i> AOS 5 ANOS.....	74
TABELA 9 - ESTATÍSTICAS DE REGRESSÃO DAS MELHORES ESTRATÉGIAS DE MODELAGEM VOLUMÉTRICA PARA <i>Pinus oocarpa</i> AOS 5 ANOS.....	76
TABELA 10 - COEFICIENTES DE REGRESSÃO E PARÂMETROS DE TRANSFORMAÇÃO DAS MELHORES ESTRATÉGIAS DE MODELAGEM VOLUMÉTRICA PARA <i>Pinus oocarpa</i> AOS 5 ANOS. ....	77
TABELA 11 - ESTATÍSTICAS DE REGRESSÃO DAS MELHORES ESTRATÉGIAS DE MODELAGEM VOLUMÉTRICA PARA <i>Pinus oocarpa</i> AOS 19 ANOS.....	80
TABELA 12 - COEFICIENTES DE REGRESSÃO E PARÂMETROS DE TRANSFORMAÇÃO DAS MELHORES ESTRATÉGIAS DE MODELAGEM VOLUMÉTRICA PARA <i>Pinus oocarpa</i> AOS 19 ANOS. ....	81
TABELA 13 - ESTATÍSTICAS DE REGRESSÃO DAS MELHORES ESTRATÉGIAS DE MODELAGEM VOLUMÉTRICA DE UMA AMOSTRAGEM DE ÁRVORES NATURAIS DA MATA ATLÂNTICA.....	84

TABELA 14 - COEFICIENTES DE REGRESSÃO E PARÂMETROS DE TRANSFORMAÇÃO DAS MELHORES ESTRATÉGIAS DE MODELAGEM VOLUMÉTRICA DE UMA AMOSTRAGEM DE ÁRVORES NATURAIS DA MATA ATLÂNTICA.....	85
TABELA 15 - ESTATÍSTICAS DE REGRESSÃO DAS MELHORES ESTRATÉGIAS DE MODELAGEM VOLUMÉTRICA DE UMA AMOSTRA DE ÁRVORES NATURAIS DA FLORESTA AMAZÔNICA.....	88
TABELA 16 - COEFICIENTES DE REGRESSÃO E PARÂMETROS DE TRANSFORMAÇÃO DAS MELHORES ESTRATÉGIAS DE MODELAGEM VOLUMÉTRICA DE UMA AMOSTRA DE ÁRVORES NATURAIS DA FLORESTA AMAZÔNICA.....	89
TABELA 17 - ESTATÍSTICAS DE REGRESSÃO DAS MELHORES ESTRATÉGIAS DE MODELAGEM DA BIOMASSA TOTAL DE <i>Mimosa scrabella</i> EM FUNÇÃO DA ALTURA E DO DIÂMETRO.....	92
TABELA 18 - COEFICIENTES DE REGRESSÃO DAS MELHORES ESTRATÉGIAS DE MODELAGEM DA BIOMASSA TOTAL DE <i>Mimosa scrabella</i> EM FUNÇÃO DA ALTURA E DO DIÂMETRO.....	99
TABELA 19 - ESTATÍSTICAS DE REGRESSÃO DAS MELHORES ESTRATÉGIAS DE MODELAGEM DA BIOMASSA DO FUSTE DE <i>Mimosa scrabella</i> EM FUNÇÃO DA ALTURA E DO DIÂMETRO.....	100
TABELA 20 - COEFICIENTES DE REGRESSÃO E PARÂMETROS DE TRANSFORMAÇÃO DAS MELHORES ESTRATÉGIAS DE MODELAGEM DA BIOMASSA DO FUSTE DE <i>Mimosa scrabella</i> EM FUNÇÃO DA ALTURA E DO DIÂMETRO.....	105
TABELA 21 - ESTATÍSTICAS DE REGRESSÃO DAS MELHORES ESTRATÉGIAS DE MODELAGEM DA BIOMASSA DE GALHOS DE <i>Mimosa scrabella</i> EM FUNÇÃO DA ALTURA E DO DIÂMETRO.....	106
TABELA 22 - COEFICIENTES DE REGRESSÃO E PARÂMETROS DE TRANSFORMAÇÃO DAS MELHORES ESTRATÉGIAS DE MODELAGEM DA BIOMASSA DE GALHOS DE <i>Mimosa scrabella</i> EM FUNÇÃO DA ALTURA E DO DIÂMETRO.....	112
TABELA 23 - ESTATÍSTICAS DE REGRESSÃO DAS MELHORES ESTRATÉGIAS DE MODELAGEM DA BIOMASSA DE FOLHAS DE <i>Mimosa scrabella</i> EM FUNÇÃO DA ALTURA E DO DIÂMETRO.....	113
TABELA 24 - COEFICIENTES DE REGRESSÃO E PARÂMETROS DE TRANSFORMAÇÃO DAS MELHORES ESTRATÉGIAS DE MODELAGEM DA BIOMASSA DAS FOLHAS DE <i>Mimosa scrabella</i> EM FUNÇÃO DA ALTURA E DO DIÂMETRO.....	117
TABELA 25 - ESTATÍSTICAS DE REGRESSÃO DAS MELHORES ESTRATÉGIAS DE MODELAGEM DA BIOMASSA DE CASCA DE <i>Mimosa scrabella</i> EM FUNÇÃO DA ALTURA E DO DIÂMETRO.....	119
TABELA 26 - COEFICIENTES DE REGRESSÃO E PARÂMETROS DE TRANSFORMAÇÃO DAS MELHORES ESTRATÉGIAS DE MODELAGEM DA BIOMASSA DA CASCA DE <i>Mimosa scrabella</i> EM FUNÇÃO DA ALTURA E DO DIÂMETRO.....	124

TABELA 27 - ESTATÍSTICAS DE REGRESSÃO DAS MELHORES ESTRATÉGIAS DE MODELAGEM DA BIOMASSA DAS RAÍZES DE <i>Mimosa scrabella</i> EM FUNÇÃO DA ALTURA E DO DIÂMETRO.....	125
TABELA 28 - COEFICIENTES DE REGRESSÃO E PARÂMETROS DE TRANSFORMAÇÃO DAS MELHORES ESTRATÉGIAS DE MODELAGEM DA BIOMASSA DAS RAÍZES DE <i>Mimosa scrabella</i> EM FUNÇÃO DA ALTURA E DO DIÂMETRO.....	130

## LISTA DE SÍMBOLOS

$\mathcal{Y}$  – variável dependente/resposta

$\mathcal{X}$  – variável independente/explicativa

$\mathcal{Y}'$  – variável dependente/resposta transformada

$\mathcal{X}'$  – variável independente/explicativa transformada

$L$  – função de verossimilhança

$\lambda$  – parâmetro de transformação lambda

$\beta$  – coeficiente beta de regressão

$\Sigma$  – somatório de números

$\neq$  – diferente de

$=$  – igual a

$>$  – maior que

$<$  – menor que

$\geq$  – maior ou igual que

$\leq$  – menor ou igual que

$\varepsilon$  – erro estatístico



## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	<b>17</b>
<b>2 OBJETIVOS</b> .....	<b>20</b>
2.1 Objetivo geral.....	20
2.2 Objetivos específicos.....	20
<b>3 REFERENCIAL TEÓRICO</b> .....	<b>21</b>
3.1 TÓPICO 1: REGRESSÃO LINEAR PARA ESTIMATIVA DE VARIÁVEIS DENDROMÉTRICAS.....	21
3.2 TÓPICO 2: MÉTODOS DE TRANSFORMAÇÃO DE VARIÁVEIS PARA REGRESSÃO LINEAR .....	27
3.2.1 Transformação logarítmica .....	28
3.2.2 Transformação raiz quadrada.....	29
3.2.3 Transformação recíproca ou inversa.....	30
3.2.4 Transformação Box-Cox.....	30
3.2.5 Transformação Manly .....	32
3.2.6 Transformação Bickel-Doksum.....	33
3.2.7 Transformação Yeo-Johnson .....	33
3.2.8 Transformação Glog .....	34
3.2.9 Transformação Dual power .....	35
3.2.10 Transformação G power .....	36
3.2.11 Transformação Log shift .....	37
3.2.12 Transformação raiz quadrada deslocada.....	37
<b>4 MATERIAL E MÉTODOS</b> .....	<b>39</b>
4.1 BASES DE DADOS .....	39
4.1.2 Florestas plantadas de <i>Pinus oocarpa</i> .....	40
4.1.3 Floresta natural amazônica .....	41
4.1.4 Floresta natural atlântica.....	42
4.1.5 Floresta natural de <i>Mimosa scrabella</i> Benth.....	43
4.2 MÉTODOS DE TRANSFORMAÇÃO DE VARIÁVEIS.....	46
4.2.1 Transformação Logarítmica (Log e Log-Log).....	49
4.2.2 Transformação recíproca.....	50
4.2.3 Transformação de Box-Cox.....	50
4.2.4 Transformação de Manly .....	52
4.2.5 Transformação de Bickel-Doksum .....	53
4.2.6 Transformação de Yeo-Johnson .....	54
4.2.7 Transformação Glog .....	55

4.2.8	Transformação Dual power .....	56
4.2.9	Transformação G power .....	57
4.2.10	Transformação Log shift .....	58
4.2.11	Transformação Raiz quadrada deslocada .....	59
4.2.12	Transformação Raiz - Raiz .....	60
4.3	AVALIAÇÃO ESTATÍSTICA DOS MODELOS DE REGRESSÃO .....	60
4.3.1	Métricas de Qualidade de Ajuste.....	61
4.3.1.1	Coeficiente de determinação ajustado ( $R^2_{aj}$ ).....	61
4.3.1.2	Erro padrão da estimativa (SEE).....	62
4.3.1.3	Critério de informação Bayesiano (BIC).....	63
4.3.2	Análise residual.....	63
<b>5</b>	<b>RESULTADOS.....</b>	<b>65</b>
5.1	ANÁLISE EXPLORATÓRIA DOS DADOS .....	65
5.1.1	Base de dados de florestas plantadas .....	65
5.1.2	Base de dados de florestas naturais da Amazônia.....	66
5.1.3	Base de dados de florestas naturais da Mata Atlântica .....	66
5.1.4	Base de dados de biomassa de povoamentos naturais de <i>Mimosa scrabella</i> .....	67
5.1.4.1	Biomassa total.....	67
5.1.4.2	Biomassa do fuste .....	68
5.1.4.3	Biomassa dos galhos.....	69
5.1.4.4	Biomassa das folhas.....	69
5.1.4.5	Biomassa das cascas .....	70
5.1.4.6	Biomassa das raízes.....	71
5.2	Estratégias de modelagem DA RELAÇÃO ALTURA-DIÂMETRO para florestas plantadas.....	71
5.3	Estratégias de modelagem volumétrica para florestas plantadas.....	76
5.4	Estratégias de modelagem volumétrica para florestas naturais.....	83
5.5	Estratégias de modelagem da biomassa total de árvores de <i>Mimosa scrabella</i>	91
5.6	Estratégias de modelagem da biomassa do fuste de árvores de <i>Mimosa scrabella</i> .....	99
5.7	Estratégias de modelagem da biomassa dos galhos de árvores de <i>Mimosa scrabella</i> .....	105
5.8	Estratégias de modelagem da biomassa das folhas de árvores de <i>Mimosa scrabella</i> .....	112
5.9	Estratégias de modelagem da biomassa da casca de árvores de <i>Mimosa scrabella</i> .....	118
5.10	Estratégias de modelagem da biomassa da raiz de árvores de <i>Mimosa scrabella</i> .....	124

<b>6 DISCUSSÃO</b> .....	<b>130</b>
<b>7 CONCLUSÕES</b> .....	<b>135</b>
<b>REFERÊNCIAS</b> .....	<b>136</b>
<b>Apêndice 1 - HISTOGRAMAS DE FREQUÊNCIA E RELAÇÕES DE DISPERSÃO ENTRE AS VARIÁVEIS DENDROMÉTRICAS VOLUME, DIÂMETRO E ALTURA EM ESCALA ORIGINAL E SOB EFEITO DA TRANSFORMAÇÃO LOGARÍTMICA (Log).</b> .....	<b>148</b>
<b>Apêndice 2 – HISTOGRAMAS DE FREQUÊNCIA E RELAÇÕES DE DISPERSÃO ENTRE AS VARIÁVEIS DENDROMÉTRICAS VOLUME E DIÂMETRO EM ESCALA ORIGINAL E SOB EFEITO DA TRANSFORMAÇÃO LOGARÍTMICA (LOG). BASE DE DADOS ADVINDA DE UMA FLORESTA NATURAL DA AMAZÔNIA.</b> .....	<b>150</b>
<b>Apêndice 3 – HISTOGRAMAS DE FREQUÊNCIA E RELAÇÕES DE DISPERSÃO ENTRE AS VARIÁVEIS DEDROMÉTRICAS DIÂMETRO E VOLUME EM ESCALA ORIGINAL E SOB EFEITO DA TRANSFORMAÇÃO LOGARÍTMICA (Log). BASE DE DADOS ADVINDA DE UMA FLORESTA NATURAL DA MATA ATLÂNTICA.</b> .....	<b>151</b>
<b>Apêndice 4 – GRÁFICOS DE DISPERSÃO E HISTOGRAMAS DE FREQUÊNCIA DAS VARIÁVEIS DENDROMÉTRICAS DIÂMETRO E ALTURA EM ESCALA ORIGINAL E SOB EFEITO DA TRANSFORAMAÇÃO LOGARÍTMICA (Log). BASE DE DADOS ADVINDA DE UM POVOAMENTO DE <i>Mimosa scrabella</i>.</b> .....	<b>152</b>
<b>Apêndice 5 – HISTOGRAMA DE FREQUÊNCIA DA VARIÁVEL RESPOSTA BIOMASSA TOTAL E SUA DISPERSÃO EM FUNÇÃO DAS COVARIÁVEIS DIÂMETRO E ALTURA. VARIÁVEIS REPRESENTADAS EM SUA ESCALA ORIGINAL E SOB EFEITO DA TRANSFORMAÇÃO LOGARÍTMICA (Log). BASE DE DADOS ADVINDA DE UM POVOAMENTO DE <i>Mimosa scrabella</i>.</b> .....	<b>153</b>
<b>Apêndice 6 – HISTOGRAMA DE FREQUÊNCIA DA VARIÁVEL RESPOSTA BIOMASSA DOS FUSTES E SUA DISPERSÃO EM FUNÇÃO DAS COVARIÁVEIS DIÂMETRO E ALTURA. VARIÁVEIS REPRESENTADAS EM SUA ESCALA ORIGINAL E SOB EFEITO DA TRANSFORMAÇÃO LOGARÍTMICA (Log). BASE DE DADOS ADVINDA DE UM POVOAMENTO DE <i>Mimosa scrabella</i>.</b> .....	<b>154</b>
<b>Apêndice 7 – HISTOGRAMA DE FREQUÊNCIA DA VARIÁVEL RESPOSTA BIOMASSA DOS GALHOS E SUA DISPERSÃO EM FUNÇÃO DAS COVARIÁVEIS DIÂMETRO E ALTURA. VARIÁVEIS REPRESENTADAS EM SUA ESCALA ORIGINAL E SOB EFEITO DA TRANSFORMAÇÃO LOGARÍTMICA (Log). BASE DE DADOS ADVINDA DE UM POVOAMENTO DE <i>Mimosa scrabella</i>.</b> .....	<b>155</b>
<b>Apêndice 8 – HISTOGRAMA DE FREQUÊNCIA DA VARIÁVEL RESPOSTA BIOMASSA DAS FOLHAS E SUA DISPERSÃO EM FUNÇÃO DAS COVARIÁVEIS DIÂMETRO E ALTURA. VARIÁVEIS REPRESENTADAS EM SUA ESCALA ORIGINAL E SOB EFEITO DA TRANSFORMAÇÃO LOGARÍTMICA (Log). BASE DE DADOS ADVINDA DE UM POVOAMENTO DE <i>Mimosa scrabella</i>.</b> .....	<b>156</b>
<b>Apêndice 9 – HISTOGRAMA DE FREQUÊNCIA DA VARIÁVEL RESPOSTA BIOMASSA DAS cascas E SUA DISPERSÃO EM FUNÇÃO DAS COVARIÁVEIS DIÂMETRO E ALTURA. VARIÁVEIS REPRESENTADAS EM SUA ESCALA ORIGINAL E SOB EFEITO DA TRANSFORMAÇÃO LOGARÍTMICA (Log). BASE DE DADOS ADVINDA DE UM POVOAMENTO DE <i>Mimosa scrabella</i>.</b> .....	<b>157</b>

<b>Apêndice 10 – HISTOGRAMA DE FREQUÊNCIA DA VARIÁVEL RESPOSTA BIOMASSA DAS RAÍZES E SUA DISPERSÃO EM FUNÇÃO DAS COVARIÁVEIS DIÂMETRO E ALTURA. VARIÁVEIS REPRESENTADAS EM SUA ESCALA ORIGINAL E SOB EFEITO DA TRANSFORMAÇÃO LOGARÍTMICA (Log). BASE DE DADOS ADVINDA DE UM POVOAMENTO DE <i>Mimosa scrabella</i> .....</b>	<b>158</b>
<b>Apêndice 11 – DEMONSTRAÇÃO DOS PERFIS DE MÁXIMA VEROSSIMILHANÇA DE LAMBDA (<math>\lambda</math>) PARA OS CASOS ESPECIAIS DE TRANSFORMAÇÕES POTENCIAIS DE BOX-COX.....</b>	<b>159</b>

## 1 INTRODUÇÃO

Uma das questões mais importantes para o manejo adequado de florestas diz respeito às estimativas precisas de variáveis dendrométricas (VIBRANS et al., 2015; OLIVEIRA et al., 2018; SANTOS et al., 2020). Assim, compreender as relações entre os atributos biométricos de florestas nativas e plantadas são fundamentais para o manejo da produção florestal (NASCIMENTO et al., 2020). A técnica de regressão linear aplicada em variáveis dendrométricas configura-se como uma ferramenta imprescindível na ciência florestal (MOSEER; OLIVEIRA, 2017), pois pode ser empregada para prever o comportamento e a relação de determinados atributos de interesse a partir de uma ou mais variáveis preditoras (DRAPER; SMITH, 1998; MONTGOMERY, et al., 2012).

A regressão linear é um dos métodos estatísticos mais empregados para a compreensão e estimativa de variáveis dendrométricas, sendo amplamente utilizada em estudos florestais para modelagem de altura, volume e biomassa de povoamentos florestais (COLE; EWEL, 2006; VIBRANS et al., 2015; OLIVEIRA et al., 2018; CYSNEIROS et al., 2020a; NASCIMENTO et al., 2020; SANTOS et al., 2020). Partindo dessa conjectura, os modelos de regressão são essenciais para o manejo eficiente dos recursos florestais, pois fornecem informações consistentes para a tomada de decisões e alternativas silviculturais, sobretudo, no que diz respeito à produção e ao uso sustentável das florestas.

No entanto, para o uso adequado de modelos lineares para fins preditivos, um conjunto de premissas precisa ser rigorosamente atendido. Contudo, apesar de fundamentais, algumas dessas premissas são comumente negligenciadas na dendrometria, das quais destaca-se a linearidade, normalidade residual e homogeneidade das variâncias. O não atendimento dessas suposições ocorrem devido ao comportamento natural das variáveis biológicas usadas em modelagens dendrométricas (BELHING et al., 2019; FIORENTIN et al., 2020).

Nessa perspectiva, em estudos dendrométricos que envolvem a aplicação alométrica, a exemplo da relação altura-diâmetro, volumétrica e de biomassa, frequentemente são violados e ignorados alguns importantes pressupostos da regressão linear, como a normalidade dos resíduos e a

homogeneidade das variâncias (homocedasticidade) (VIBRANS et al., 2015; CYSNEIROS et al., 2020a). Entretanto, o atendimento desses dois critérios é essencial para se obter equações cujos coeficientes não sejam estimados com perda de eficiência ou que forneçam estimativas e estatísticas duvidosas relacionadas a testes paramétricos, tais como  $t$  e  $F$  (DRAPER; SMITH, 1998; SAKIA, 1992; OSBORNE, 2010; GUJARATI; PORTER, 2011; MEDINA et al., 2019). Deste modo, como forma de corrigir a violação desses pressupostos, destacam-se as técnicas de transformação de variáveis, as quais, em muitas ocasiões, apresentam promissora capacidade de adequação dos modelos à normalidade dos resíduos e homocedasticidade das variâncias (COLE; EWEL, 2006; NG`ANDWE et al., 2019).

Nesse segmento, a transformação de variáveis em regressão linear representa uma prática importante de tratamento e preparo dos dados para as análises estatísticas posteriores, necessária principalmente quando os resíduos gerados não são aderentes às suposições da regressão linear (FOX, 2016; ROJAS-PERILLA, 2018; MEDINA et al., 2019; CYSNEIROS et al., 2020a). Contudo, especialmente em estudos dendrométricos, verifica-se uma notável resistência quanto as adoções dessa técnica, pois, com exceção da transformação logarítmica, raros são os trabalhos descritos na literatura florestal abordando essa temática (FISCHER, 2016; ZELL, 2018).

Não obstante, a transformação logarítmica é um método comumente aplicado em estudos alométricos (COLE; EWEL, 2006; VIBRANS et al., 2015; FADINI et al., 2019; CYSNEIROS et al., 2020a; TERRA et al., 2021). Todavia, em muitas ocasiões na qual é utilizado, observa-se que os problemas de ausência de normalidade dos resíduos e de homogeneidade das variâncias persistem, ou, em outros casos, corrige-se a inadequação de apenas um dos pressupostos (SAKIA, 1992, OSBORNE, 2010).

Dessa forma, essa problemática nos leva a necessidade de explorar novas técnicas de transformações estatísticas em variáveis dendrométricas, especialmente naquelas advindas de bases de dados de inventários florestais, de onde se derivam as variáveis mais utilizadas em modelos lineares. No entanto, ressalta-se, que esse tipo de estudo raramente foi desenvolvido sob a ótica da ciência florestal, o que gera escassez de informações na literatura sobre aplicações de transformações de variáveis em modelagens alométricas nessa

área de conhecimento (ex.: EASTAUGH; HASENAUER, 2011; MONNESS, 2011; SERINALDI et al., 2012; FISCHER, 2016; FAYOLLE et al., 2018).

Diante desse contexto, esta dissertação visa determinar as estratégias de transformação de variáveis mais adequadas para estudos com modelagem linear de variáveis dendrométricas, de modo que o atendimento dos pressupostos de normalidade dos resíduos e homocedasticidade das variâncias sejam simultaneamente contemplados, tendo em vista a minimização de esforços operacionais em estudos dendrométricos envolvendo a técnica de regressão linear. Para isso, considerou-se que os ajustes de modelos lineares com transformação de variáveis podem resultar em equações estatisticamente mais apropriadas para a predição da relação altura-diâmetro, do volume comercial e de biomassas em inventários de plantios florestais e florestas naturais.

## 2 OBJETIVOS

### 2.1 OBJETIVO GERAL

Aplicar diferentes estratégias de transformação de variáveis em modelos de regressão linear e determinar as melhores para o uso em estudos alométricos, visando a adequação simultânea dos pressupostos clássicos da análise de regressão linear, de modo que as equações geradas sejam hábeis à predição e estatisticamente confiáveis.

### 2.2 OBJETIVOS ESPECÍFICOS

- a) Recomendar estratégias de transformação em modelos de regressão linear ajustados para estimativas hipsométricas em florestas plantadas em diferentes idades.
- b) Propor estratégias de transformação que atendam simultaneamente as suposições de normalidade residual e homogeneidade das variâncias em estudos de modelagem volumétrica em florestas plantadas e naturais.
- c) Ajustar modelos lineares com variáveis transformadas que sejam estatisticamente confiáveis e apropriados à estimativa de biomassa total de árvores, bem como para cada um dos seus componentes: fuste, galhos, folhas e raízes.



### 3 REFERENCIAL TEÓRICO

#### 3.1 TÓPICO 1: REGRESSÃO LINEAR PARA ESTIMATIVA DE VARIÁVEIS DENDROMÉTRICAS

Em diversas áreas de conhecimento, é fundamental conhecer os efeitos de relação entre variáveis. Essa abordagem analítica, na qual se pretende conhecer o grau de influência entre variáveis, especificamente, a previsão de uma variável dependente baseada nos valores de uma ou mais variáveis preditoras é uma técnica estatística denominada de regressão (GOTTELL; ELLISON, 2011; GUJARATI; PORTER, 2011; DINIZ; THIELE, 2021). Por assim dizer, enfatiza-se que a regressão é uma das técnicas estatísticas mais importantes na ciência de dados voltada à tomada de decisões (BRUCE; BRUCE, 2019). Diante disso, a regressão linear se destaca como a mais amplamente usada na maioria dos ramos do conhecimento, em particular nas ciências sociais e naturais, devido a sua praticidade de aplicação e fácil entendimento (ROJAS-PERILLA, 2018).

Historicamente, o termo “regressão” foi proposto pela primeira vez no século XIX por Sir Francis Galton (1885), em um estudo sobre tendências genéticas, no qual demonstrou que a altura dos filhos não reflete a altura dos pais, mas sim tende a regredir para a média da população. Desde então, a análise de regressão se consolidou ao longo dos anos para definir e quantificar a relação entre as variáveis em um conjunto de dados, possibilitando prever o valor de uma variável dependente com base no valor de pelo menos uma variável independente (DUNN; SMYTH, 2018; KUMARI; YADAV, 2018; BRUCE; BRUCE, 2019). Nessa conjectura, o modelo de regressão linear configura-se como o mais simples e popular na análise estatística de previsão, especificidades que o colocam em uma posição de prioridade de aplicação nas mais diversas áreas de conhecimento (FADINI et al., 2019; YANG et al., 2019; NASCIMENTO et al., 2020; ROJAS-PERILLA, 2021).

À vista disso, a técnica de regressão linear pode ser categorizada como simples ou múltipla de acordo com sua estrutura. Assim, os modelos lineares simples são aqueles estruturados com uma variável dependente contínua, designada por  $y$ , associada a apenas uma variável independente ou resposta,

designada por  $X$ . Em contrapartida, os modelos lineares múltiplos são os que detêm mais de uma variável explicativa associada à variável resposta (CASSIN; FARMER, 2014; KUMARI; RAMESH, 2020).

Na estatística, diz-se que a regressão linear simples modela o relacionamento entre a magnitude de uma variável resposta com relação a uma variável independente ou preditora (CHARNET et al., 2008; BRUCE; BRUCE, 2019). Em outras palavras, os modelos de regressão linear simples estimam exatamente o quanto  $Y$  mudará quando  $X$  é alterado em uma dada proporção. Dessa forma, com a análise de regressão, a predição de uma variável de interesse ocorre a partir do relacionamento linear desta com outra, denominada preditora.

A regressão linear múltipla, por sua vez, é uma técnica multivariada cuja finalidade principal é obter relações matemáticas entre uma variável resposta e outras preditoras que descrevem o sistema (WEISBERG, 2014; MOSER; OLIVEIRA, 2017). Nessa perspectiva, por haver múltiplas preditoras, o modelo simplesmente se estende para poder acomodá-las. Sobretudo, a regressão linear múltipla é a mais amplamente utilizada na ciência florestal para predição de variáveis dendrométricas, haja vista que, em muitas ocasiões, a variável de interesse apresenta, naturalmente, mais de uma variável preditora em potencial.

Assim, a incorporação de mais covariáveis ao modelo linear pode ser realizada com o propósito de melhorar a capacidade explicativa do modelo de regressão (NOGUEIRA et al., 2008; TEMESGEN et al., 2015; CAMPOS; LEITE, 2017). No entanto, na literatura florestal, alguns pesquisadores defendem o uso de modelos lineares simples para predição de variáveis dendrométricas, especialmente o volume, pois priorizam a maior parcimônia nas modelagens e a redução de custos operacionais na obtenção de dados e nas estimativas (SANTOS et al., 2020; SANTOS et al., 2021b).

Nessa perspectiva, a regressão linear é um pilar estatístico fundamental em atividades alométricas envolvendo variáveis dendrométricas, tais como: o diâmetro, a altura, a forma, o volume e a biomassa (FINGER, 2006; LIMA et al., 2012; VIBRANS et al., 2015; FISHER, 2016; CYSNEIROS et al., 2020a). Nesse sentido, enfatiza-se que a modelagem por análise de regressão é uma ferramenta necessária à ciência florestal, sendo recorrida para diversas finalidades, nas quais envolvem: o manejo e produção florestal, a ecologia

numérica, o sensoriamento remoto e diversas outras (MOSER; OLIVEIRA, 2017).

Entretanto, os manejadores florestais têm convivido desde sempre com dificuldades de medição de variáveis de interesse, principalmente em ecossistemas de florestas nativas (NASCIMENTO et al., 2020; SANTOS et al., 2020). Dessa forma, modelos estatísticos destinados a relação entre variáveis dendrométricas são frequentemente ajustados a partir de amostras, com o propósito de estimar variáveis de interesse a partir daquelas medidas com maior facilidade e menor custo (NOGUEIRA et al., 2008; PICARD et al., 2012).

No Brasil, a técnica de regressão é empregada desde as primeiras iniciativas de inventariar os recursos florestais do país (MOSER; OLIVEIRA, 2017). Na região Sul, por exemplo, modelos de volume foram desenvolvidos na década de 1980 para estimar os estoques de madeira de araucária e angiospermas em florestas naturais (PÉLLICO NETTO, 1984). Em estudos mais recentes, também no Sul do Brasil, Vibrans et al. (2015) dedicaram esforços para ajustar modelos volumétricos para diferentes espécies e tipos de floresta a partir de dados coletados pelo Inventário Florístico Florestal de Santa Catarina (IFFSC).

No cenário nacional, destacam-se também investigações volumétricas desenvolvidas por Scolforo et al. (2008) através do Inventário Florestal de Minas Gerais, a partir do qual desenvolveram equações por espécie e tipologia florestal no estado de Minas Gerais. Assim como estudos de Rufini et al. (2010) e Santos et al. (2021a) que modelaram equações de volume por espécies em fragmento de cerrado no Centro-Oeste e Sudeste brasileiro. Outro estudo importante envolvendo regressão linear na alometria do volume de árvores foi o desenvolvido por Cysneiros et al. (2020a), que modelaram equações genéricas e específicas por espécie e tipologia florestal para fragmentos florestais da Mata Atlântica.

Não obstante, modelagens genéricas e específicas também foram desenvolvidas para a relação altura-diâmetro em fragmentos de Mata Atlântica situados no estado do Rio de Janeiro (CYSNEIROS et al., 2020b). Esforços semelhantes foram dedicados às florestas tropicais amazônicas, nas quais diversas pesquisas florestais estiveram direcionadas a quantificações dendrométricas (ex.: NOGUEIRA et al., 2008; COLPINI et al., 2009; LIMA et al.,

2012; CYSNEIROS et al., 2017; GIMENEZ et al., 2017; GOMES et al., 2018) das quais destaca-se o recente estudo desenvolvido por Santos et al. (2020), sobre modelagens volumétricas por espécies comerciais através de dados agrupados e específicos em uma floresta amazônica de produção.

Assim, observa-se que são inúmeros os estudos dendrométricos desenvolvidos no Brasil que envolvem o uso da análise de regressão linear, os quais demonstram a sua basilar relevância à ciência florestal. Destaca-se, porém, que, como toda metodologia estatística, a modelagem por regressão linear compreende numerosos aspectos para a sua correta aplicação (ACZEL; SOUNDERPANDIAN, 2009; ALEXOPOULOS, 2010; MEDINA et al., 2019). Nesse sentido, cuidados básicos precisam ser tomados durante a sua execução, pois o uso inadequado dessa ferramenta pode gerar consequências não desejadas às predições (FISCHER, 2016; ZELL, 2018). Diante disso, o atendimento das suposições de normalidade residual e homocedasticidade nos procedimentos de modelagem é uma etapa necessária para se evitar problemas no ajuste de equações lineares; entretanto, essa é uma prática comumente ignorada pela comunidade silvícola brasileira (VIBRANS et al., 2015; CYSNEIROS et al., 2020a).

Em dendrometria, os modelos de regressão linear são frequentemente ajustados usando a abordagem do método dos mínimos quadrados ordinários (*Ordinary least squares – OLS*) (RÉ et al., 2014; RIBEIRO et al., 2014; CYSNEIROS et al., 2020b; SANTOS et al., 2021b). Em suma, esse método visa ajustar os coeficientes de regressão das equações, denominados de coeficiente linear ( $\beta_0$ ) e angular ( $\beta_1$ ). Em vista disso, o método dos mínimos quadrados trata-se de uma técnica estatística direcionada a encontrar o melhor ajuste para um conjunto de dados observados, de modo que as estimativas de  $\hat{\beta}_0$  e  $\hat{\beta}_1$ , que representam os melhores estimadores lineares imparciais, são os valores que minimizam a soma dos quadrados dos erros:  $SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$ .

Os mínimos quadrados ordinários não é o único método factível em ajustes lineares, contudo, é o mais usado para estimar uma relação de regressão em modelos lineares, especialmente em estudos dendrométricos (MOSER; OLIVEIRA, 2017; SANTOS et al., 2020; ZEA-CAMAÑO et al., 2020). Cabe

destacar, porém, que para a aplicação correta desse método alguns pressupostos devem ser contemplados nos ajustes, dentre os quais destacam-se a normalidade residual e a homogeneidade das variâncias (LONG; ERVIN, 2000; GUJARATI; PORTER, 2011; KUMARI; RAMESH, 2020).

Enfatiza-se, porém, que a exigência de distribuição normal dos resíduos é o pressuposto que, geralmente, é menos marcante em regressão linear. Na verdade, para o propósito de estimar a linha de regressão, a suposição de normalidade não se apresenta necessária (SMITH; FINAN, 2018); todavia, é sobre esse pressuposto que se fundamenta a realização dos testes de validação do modelo (estatísticas  $t$ ,  $F$  e  $\chi^2$ ). Afirma-se, ainda, que é a distribuição normal associada a não correlação dos erros que garantem a independência dos resíduos (ACZEL; SOUNDERPANDIAN, 2009; BRUCE; BRUCE, 2019).

Apesar disso, há afirmações de que modelos de regressão linear com erros não normais, em via de regra, demonstram bons resultados quando as amostras apresentam tamanhos suficientes (SCHMIDT; FINAN, 2018; YANG et al., 2019). Por essa razão, há um direcionamento de raciocínio na literatura estatística pautado na afirmação de que, em amostras grandes, a suposição da normalidade residual pode ser desprezada. Entretanto, essa mesma corrente ideológica afirma que, na presença de amostras pequenas, a suposição da normalidade do termo de erro é imprescindível e deve ser rigorosamente atendida.

Por conseguinte, outro pressuposto clássico da regressão linear remete-se a homogeneidade das variâncias. A violação desse pressuposto, diferentemente da normalidade residual, acarreta consequências diretas sobre os estimadores de mínimos quadrados. Assim, diante de variâncias heterocedástica, a técnica de mínimos quadrados ordinários deixa de ser o melhor estimador linear de variâncias mínimas (DRAPER; SMITH, 1998; LONG; ERVIN, 2000; GUJARATI; PORTER, 2011; WEISBERG, 2014). Dessa forma, testes de hipóteses, tais como  $t$  e  $F$  tornam-se passíveis de descrédito estatístico devido a inconfiabilidade dos estimadores gerados.

Na presença da heterocedasticidade, o estimador gerado pelos mínimos quadrados ordinários permanece não tendencioso e consistente, porém, ao passo que não apresenta variância mínima, torna-se ineficiente (LONG; ERVIN, 2000; FOX, 2016). Diante disso, a heterocedasticidade não é um empecilho,

quando a finalidade da modelagem é apenas avaliar a contribuição das variáveis preditoras na explicação da variável resposta. No entanto, quando o objetivo principal da modelagem é a predição, a presença de heterocedasticidade pode ser adversa, devendo ser motivo de atenção especialmente quando presente em níveis elevados.

Essa cautela acerca da violação da homocedasticidade é necessária, pois promove consequências diretas na minimização dos quadrados dos resíduos. Em razão disso, o modelo pode não ter um poder preditivo satisfatório em decorrência da estimativa dos parâmetros mostrarem-se ineficientes. Outro fator limitante ocasionado pelas variâncias não homogêneas está relacionado ao estimador do erro padrão residual, que, neste caso, apresenta-se enviesado, o que compromete os resultados dos testes de significância e intervalos de confiança (MOSER; OLIVEIRA, 2017).

Em termos gerais, tanto a distribuição normal dos resíduos quanto a homocedasticidade são pressupostos comumente violados em análise de regressão linear voltada aos estudos dendrométricos. Diante disso, Vibrans et al. (2015) e Cysneiros et al. (2020a) chamaram atenção em suas pesquisas para o fato de que esse procedimento é frequentemente negligenciado em estudos alométricos no Brasil. Entretanto, convém citar que essas inconsistências se devem a natureza das variáveis e seus métodos de amostragens, visto que em inventários de diâmetro, volume e biomassa é recorrente amostragens nas quais maiores variabilidades e menores frequências de dados estão associadas, conjuntamente, às classes superiores (CUNIA; BRIGGS, 1984; BEHLING et al., 2019; FIORENTIN, 2020).

Em suma, para que a inferência formal acerca de uma modelagem linear seja totalmente válida, os resíduos terão que ser considerados como normalmente distribuídos e tenham a mesma variância (homocedásticos). Assim, a validação desses pressupostos pode ser conduzida através de análises gráficas ou testes estatísticos; no entanto, devido a questões relacionadas aos Erros do Tipo I e II associados aos usos desses testes de hipóteses, têm-se priorizado cada vez mais abordagens gráficas para inferências acerca dos pressupostos. Diante disso, para deduções a respeito da normalidade residual e da homocedasticidade das variâncias em modelagens lineares, comumente se

utilizam gráficos quantil-quantil (Q-Q) e de resíduos, respectivamente (FIFE, 2020; ROJAS-PERILLA et al., 2020).

Todavia, quando esses pressupostos clássicos da regressão linear não são satisfeitos, medidas corretivas devem ser tomadas (LONG; ERVIN, 2000; ALEXOPOULOS, 2010; PICARD et al., 2012; WEISBERG, 2014). De certo modo, quando há a violação das suposições de normalidade residual e homocedasticidade das variâncias, a aplicação de técnicas de transformação de dados é uma alternativa bastante recorrida em modelagens lineares cuja finalidade é a predição. Assim, nas literaturas estatística e florestal, diversos pesquisadores têm usado esse recurso como uma ferramenta essencial para auxiliar o trabalho científico com fins de modelagem (por ex.: COSTA et al., 2015; FISCHER, 2016; PODLASK, 2017; ZELL, 2018). Em vista disso, essa técnica assegura a ancoragem de esforços na esfera estatística da regressão linear paramétrica, que, de certo modo, é uma alternativa mais atrativa e prática. Além disso, não exige do analista de dados a adoção de métodos complexos, muito menos estender funções básicas de *softwares* estatísticos, haja vista que a aplicação de novas teorias de modelagens com conceitos mais aprofundados e laboriosos é evitada (MEDINA et al., 2019).

No entanto, em estudos dendrométricos, as transformações são frequentemente aplicadas de forma automática e rotineira, sem considerar diferentes aspectos de sua utilidade. Diante disso, esta dissertação aborda aplicações clássicas e atuais de transformações de variáveis em modelagens lineares. Assim, direcionou-se esforços, exclusivamente, às estimativas alométricas mais relevantes em atividades de produção florestal - altura, volume e biomassa - de modo a unificar conceitos sobre diferentes formas de transformação de dados para estas classes de estimativas.

### 3.2 TÓPICO 2: MÉTODOS DE TRANSFORMAÇÃO DE VARIÁVEIS PARA REGRESSÃO LINEAR

Diversas análises estatísticas, assim como a regressão linear, usam duas suposições relevantes: i) resíduos das variáveis normalmente distribuídos e ii) homocedasticidade ou homogeneidade da variância, que se refere a variância em torno da linha de regressão ser constante em todo o intervalo de

dados observado (DRAPER; SMITH, 1998; ACZEL; SOUNDERPANDIAN, 2009; OSBORNE, 2010; GUJARATI; PORTER, 2011). Assim, uma das formas de identificar o atendimento dessas premissas paramétricas é através de testes de hipóteses e análises gráficas (OSBORNE, 2010; MEDINA et al., 2019; FIFE, 2020; ROJAS-PERILLA et al., 2020).

Dessa forma, quando a violação de pelo menos uma das suposições requeridas é detectada, a transformação das variáveis surge como uma alternativa de correção dessas violações para a aplicação adequada da análise de regressão. Nesse procedimento, há uma modificação matemática dos valores da variável, que pode ocorrer de várias formas: adicionando constantes, elevando ao quadrado ou a uma potência, convertendo para escalas logarítmicas, invertendo a variável, obtendo a raiz quadrada, dentre outras (BOX; COX, 1964; SAKIA, 1992; OSBORNE, 2005; ROJAS-PERILLA, 2018; MEDINA et al., 2019). Diante disso, os subtópicos seguintes abordarão em detalhes as principais técnicas de transformação de dados descritas na literatura e que foram empregadas neste estudo.

### 3.2.1 Transformação logarítmica

A transformação logarítmica caracteriza-se por sua eficiência em tratar dados assimétricos (não normais) e não lineares (FENG et al., 2014). Assim, essa propriedade, aliada à sua fácil aplicação, contribuiu para sua ampla popularidade nas mais diversas áreas de conhecimento (MEDINA et al., 2019). Por essa razão, a estratégia logarítmica (Log) de transformação é comumente encontrada como uma função básica na maioria dos *softwares* estatísticos.

Para essa técnica de transformação, Feng et al. (2014) explicam que, caso os dados originais apresentem uma distribuição log-normal ou próxima, provavelmente quando transformados apresentarão uma distribuição normal ou quase normal. Segundo esses autores, nessas situações, a transformação logarítmica irá remover ou reduzir a assimetria, geralmente, de forma satisfatória. Por outro lado, eles relatam, ainda, que, para estudos cujos dados não se aproximam da distribuição log-normal, a transformação logarítmica não se mostra eficiente, podendo, inclusive, tornar a distribuição ainda mais assimétrica.



Ademais, na ciência florestal, a técnica de transformação logarítmica é, inquestionavelmente, a mais recorrida em modelagens de variáveis dendrométricas, estando inclusive incluída em modelos alométricos renomados na área (ex.: SCHUMACHER; HALL, 1933, HUSCH, 2003). Diante disso, devido aos seus promissores resultados quanto à adequação das suposições paramétricas, como normalidade residual e homocedasticidade das variâncias, a transformação logarítmica é amplamente utilizada em pesquisas florestais envolvendo modelagem linear (BASKERVILLE, 1972; PICARD et al., 2015; VIBRANS et al., 2015; CHAKRABORTY et al., 2016; FISCHER, 2016; CYSNEIROS et al., 2020a; SANTOS et al., 2021b).

### 3.2.2 Transformação raiz quadrada

Essa transformação representa um dos métodos mais conhecidos e difundidos na literatura estatística; consiste, basicamente, na obtenção da raiz quadrada de cada valor de uma variável (OSBORNE, 2002). No entanto, o fator limitante dessa estratégia é a impossibilidade de se obter a raiz quadrada quando a base de dados apresenta valores negativos, sendo necessário adicionar uma constante a cada número para torná-los positivos (BARTLETT, 1937; OSBORNE, 2002; MCDONALD, 2008).

Osborne (2002) relata que, ao se aplicar a transformação de raiz quadrada, verifica-se um comportamento diferente entre os números acima de 1 e os que variam entre 0 e 0,99. O autor explica que 0 e 1 permanecem constantes ao aplicar a raiz quadrada, mas os números acima de 1 sempre ficam menores, enquanto os que estão entre 0 e 1 ficam maiores. Dessa forma, ao aplicar a raiz quadrada em uma base de dados que envolva valores superiores e inferiores a 1, o pesquisador tratará alguns números de maneira diferente de outros, o que não é desejável.

No entanto, o diferencial é que essa transformação se torna útil para compactar valores altos de forma mais moderada que o método logaritmo (GELMAN; HILL, 2007). Porém, essas transformações, assim como a inversa, são apropriadas para transformar dados assimétricos à direita (PINO, 2014). Contudo, apesar de sua praticidade, essa estratégia de transformação não demonstra uso difundido em estudos alométricos aplicados para predições de

variáveis dendrométricas. Assim, a compreensão do efeito dessa estratégia na correção dos pressupostos da análise de regressão pode representar novas opções ao analista de dados.

### 3.2.3 Transformação recíproca ou inversa

A transformação recíproca caracteriza-se pela sua capacidade de tornar números muito pequenos em muito grandes e vice-versa, o que inverte a ordem dos valores (OSBORNE, 2002). Devido a essa propriedade de minimizar o efeito de possíveis valores muito altos da variável de interesse, essa estratégia de transformação é usada com a finalidade de estabilizar a variância em bases de dados com grandes dispersões (TUKEY, 1977, OSBORNE, 2010). A técnica recíproca de transformação é útil ainda em tentativas de normalização de distribuições de dados que apresentam assimetria negativa (PINO, 2014; MEDINA et al., 2019).

Essa transformação apresenta como particularidade um ajuste mais extremo dos dados em comparação aos métodos de transformação logarítmico e de raiz quadrada (SHESKIN, 2011). Dessa forma, em tentativas de normalização que essas transformações não sejam capazes de atender, o método recíproco de transformação pode se mostrar capaz devido as suas propriedades mais enérgicas.

No âmbito florestal, Costa et al. (2016) e Orellana e Figueiredo-Filho (2017) aplicaram a transformação recíproca buscando apenas aumentar a correlação entre variáveis dendrométricas para fins de seleção de variáveis via procedimento *stepwise*. Nesse sentido, em detrimento aos outros métodos tradicionais de transformação, a estratégia recíproca é menos difundida em estudos envolvendo regressão linear. Diante disso, na ciência florestal, essa estratégia de transformação apresenta lacunas de conhecimento a serem preenchidas sobre como ela poderia adequar as estruturas residuais de modelos lineares de predição aos pressupostos clássicos da análise de regressão linear.

### 3.2.4 Transformação Box-Cox

Em estudos de regressão, a transformação de Box-Cox geralmente é designada para solucionar conjuntamente o problema de não normalidade dos resíduos e não homocedasticidade das variâncias (GARSON, 2012; MEDINA et al., 2019; ROJAS-PERILLA et al., 2020). Dentre todas as transformações disponíveis na literatura, a potência Box-Cox (BOX; COX, 1964) é a mais conhecida por ser eficaz em obter distribuições aproximadamente normais.

Nesse caso, o método de transformação de Box-Cox representa uma família de transformações potenciais, levando-se em consideração diversas transformações disponíveis para ajustar a normalização e equalizar as variâncias (OSBORNE, 2010). Para isso, a transformação Box-Cox é fundamentada no método de máxima verossimilhança para determinar a transformação ideal para fins de normalização a partir de um parâmetro de transformação lambda ( $\lambda$ ).

Dessa forma, com base nas estimativas do parâmetro  $\lambda$ , pode-se chegar a uma transformação ideal dos dados, buscando-se o atendimento das suposições para análise estatística paramétrica (SAKIA, 1992; OSBORNE, 2010). A implementação desse método, bem como a estimativa de  $\lambda$ , são procedimentos relativamente simples de serem executados através de programas estatísticos (MEDINA et al., 2019; ROJAS-PERILLA et al., 2020).

Nesse contexto, os pacotes estatísticos aplicam um algoritmo de máxima verossimilhança iterativo para calcular  $\lambda$ , a fim de gerar com eficácia a melhor transformação (GARSON, 2012). Osborne (2010) explica que, como regra geral, o método de Box-Cox pode assumir, a partir do parâmetro de transformação  $\lambda$ , diferentes tipos de transformações, em que para:  $\lambda = 0,5$  aplica-se a transformação de raiz quadrada;  $\lambda = 0,33$  a transformação de raiz cúbica;  $\lambda = 0,25$  a raiz a quarta potência;  $\lambda = 0$  o logaritmo;  $\lambda = -0,5$  a transformação de raiz quadrada recíproca;  $\lambda = -1$  a transformação recíproca (inversa) e  $\lambda = 1$  nenhuma transformação será necessária. Os perfis de máxima verossimilhança dos principais parâmetros de transformação  $\lambda$  para o método Box-Cox estão ilustrados na Figura 30 do Apêndice 11.

Na literatura florestal, o uso dessa técnica geralmente demonstra bons resultados (COSTA et al., 2015; FISCHER, 2016; FAYOLLE et al., 2018; LHOTKA et al., 2018). Porém, cabe enfatizar que, apesar da sua versatilidade,

a estratégia de transformação proposta por Box-Cox (1964) apresenta limitações, pois é uma técnica restrita apenas aos números positivos, além disso, mostra-se ineficaz para correção de dados assimétricos à esquerda (MANLY, 1976; ZHU; MELNYKOV, 2018). Diante disso, percepções específicas a respeito da sua aplicação em modelagens lineares de variáveis dendrométricas ainda são incipientes na literatura florestal.

### 3.2.5 Transformação Manly

A transformação de Manly (MANLY, 1976) foi implementada, inicialmente, como uma alternativa ao método Box-Cox. Em seu esforço, a técnica propôs satisfazer distribuições que abrangessem valores negativos, buscando preencher as lacunas do método de transformação de Box-Cox (PERRY, 2018; ZHU; MELNYKOV, 2018). Assim, a transformação exponencial Manly foi projetada para ser aplicada à dados na faixa de  $-\infty$  para  $\infty$  (SAKIA, 1992).

Nesse caso, essa transformação é fundamentada no cálculo do exponencial natural da variável dependente relacionada com a estimativa de um parâmetro  $\lambda$  de transformação, semelhante ao método proposto por Box-Cox (1964). Todavia, o ponto forte desse método de transformação é a eficácia para adequar conjuntos de dados com distribuição unimodal à normalidade (SAKIA, 1992; CHORTIRAT et al., 2011). Apesar de ser considerada uma opção mais flexível em relação à transformação Box-Cox, a transformação de Manly é utilizada com menos frequência na literatura; isso se deve a maior popularidade e maior facilidade de interpretação da transformação Box-Cox (SAKIA, 1992; PERRY, 2018; MEDINA et al., 2019).

Diante disso, quanto às interpretações dos parâmetros de transformação  $\lambda$  estimados para a transformação de Manly, assume-se que valores de  $\lambda < 0$  são suficientes para transformar dados com assimetria positiva, enquanto valores de  $\lambda > 0$  são suficientes para transformar dados com assimetria negativa (PERRY, 2018). Em contrapartida, quando a estimativa de  $\lambda = 0$ , assume-se que nenhuma operação de transformação é necessária (MANLY, 1976). Contudo, apesar da sua versatilidade de aplicação, relativa popularidade e eficácia na adequação de dados assimétricos, são escassos os conhecimentos a respeito

desse método para adequação da normalidade residual e homocedasticidade das variâncias em modelagens lineares, sobretudo com relação a estimativas de variáveis dendrométricas.

### 3.2.6 Transformação Bickel-Doksum

Essa transformação foi desenvolvida por Bickel e Doksum (1981) como uma tentativa de substituir a de Box-Cox. Diante disso, os autores propuseram um método que, ao contrário do proposto por Box e Cox (1964), assume conjuntos de valores negativos para a adequação das suposições estatísticas paramétricas. Dessa forma, devido a sua maior flexibilidade, Zeuehlke (2020) referiu-se a esse método como transformação de potência sinalizada por apresentar capacidade de aplicação para todos os números reais.

Contudo, o método de transformação Bickel-Doksum é recomendável quando se pretende corrigir distribuições quase simétricas, haja vista que não se apresenta muito eficiente para adequação de dados muito assimétricos (FRASER et al., 2016). Assim, o uso dessa transformação é preconizado especialmente para lidar com curtose em vez de assimetria, particularmente para distribuições leptocúrticas e platicúrticas (MEDINA et al., 2019).

Além disso, afirma-se que essa transformação altera os valores transformados de convexos para côncavos à medida que a variável  $y$  altera seus valores de negativo para positivo. Por essa razão, torna-se difícil prever seu efeito em dados assimétricos, ao menos que a variável resposta seja inteiramente positiva ou negativa (CARROLL; RUPPERT, 1988). Todavia, essa limitação não representa impedimento para sua aplicação em variáveis dendrométricas, visto que, geralmente, a natureza das variáveis é contínua.

### 3.2.7 Transformação Yeo-Johnson

Yeo e Johnson (2000) propuseram uma família alternativa de transformações que pode lidar com dados negativos e positivos. Essa proposta foi embasada na limitação da conhecida transformação Box-Cox (BOX; COX, 1964), em não assumir valores negativos. Aliás, destaca-se que essa limitação da estratégia Box-Cox, foi extensivamente debatida na literatura por outros

autores (MANLY, 1976; BICKEL; DOKSUM, 1981; RAYMAEKERS; ROUSSEEUW, 2021). Diante disso, a transformação Yeo-Johnson usa uma condição de suavidade para atender tantos resultados positivos como negativos (ATKINSON, 2021).

Dessa forma, a transformação é normalmente realizada por meio do método de máxima verossimilhança, seguindo os mesmos procedimentos do método Box-Cox. Assim, a partir das estimativas de  $\lambda$ , a transformação de Yeo-Johnson fica sujeita a interpretações, de modo que, diante de  $\lambda < 1$ , a transformação comprime o lado direito da cauda de distribuição e expande a cauda esquerda, o que demonstra um uso adequado para transformar distribuições assimétricas à direita. Enquanto isso, quando  $\lambda > 1$ , o procedimento é o inverso (BEAN et al., 2016).

Nessa conjectura, a família de transformações Yeo-Johnson, por demonstrar, em termos práticos, maior abrangência que a família de transformações Box-Cox, mostra-se um método mais completo para tratamento de dados (SOLEYMANI, 2018). Entretanto, apesar da popularidade da família de transformações Yeo-Johnson para adequação de dados às suposições gaussianas (RAYMAEKERS; ROUSSEEUW, 2021), conhecimentos sobre seus efeitos em estudos envolvendo variáveis dendrométricas ainda são inexistentes.

Diante disso, para estudos florestais, a problemática acerca das características negativas e positivas das variáveis não é um fator limitante, pois geralmente a natureza das variáveis dendrométricas são positivas. Nesse sentido, o enfoque principal da aplicação da transformação Yeo-Johnson estaria relacionado sobre a sua performance no atendimento das suposições da modelagem de regressão linear em estudos alométricos.

### 3.2.8 Transformação Glog

A transformação Glog ou Log generalizado foi proposta por Durbin et al. (2002), trata-se de uma modificação da popular transformação logarítmica que se limita a variáveis resposta com valores negativos (DURBIN et al., 2002; MEDINA et al., 2019). Apesar de se estender aos valores negativos, enfatiza-se que a transformação Glog foi projetada, *a priori*, para atender a suposição de distribuição gaussiana. No entanto, em estudos envolvendo sua aplicação,

verifica-se, também, sua boa performance na estabilização de variâncias (ROCKE; DURBIN, 2003; YAN; IRIZARRY, 2006; ROJAS-PERILLA, 2018), o que justifica sua aplicação também para adequação de homocedasticidade.

Apesar de demonstrar bons resultados quanto ao atendimento das suposições gaussianas, o uso da transformação Glog esteve relacionado apenas a análises de dados de expressão gênica (DURBIN et al., 2002; ROCKE; DURBIN, 2003). No entanto, o uso da transformação Glog também tem sido criticado devido ao seu comportamento, que é semelhante à transformação logarítmica para grandes valores de dados. Essa crítica é baseada em um corpo de evidências que indica que é muito severa para dados em grandes valores (SPEED, 2003; KELMANSKY et al., 2013).

Assim, assume-se, na ciência florestal, que amostragens de variáveis dendrométricas geralmente apresentam grande heterocedasticidade, ou seja, medidas de grande intensidade estão associadas a grande variabilidade, além de ausência de normalidade residual. Diante disso, considerando a popular aptidão da transformação Glog para normalização e estabilização de variâncias em grandes conjuntos de dados (KELMANSKY et al., 2013), considera-se que conhecimentos a respeito da performance desse método pode representar uma alternativa ao método tradicional de transformação logarítmica, que, definitivamente, é o mais aplicado na literatura florestal (RIBEIRO et al., 2014; VIBRANS et al., 2015; CYSNEIROS, et al., 2020a).

### 3.2.9 Transformação Dual power

Essa transformação é estabelecida com base em duas funções de potência  $\lambda$ , nas quais uma apresenta potência positiva e outra negativa (YANG, 2006). Diante disso, esse método de transformação caracteriza-se por ser aplicado apenas em variáveis resposta positivas, visando a superação de problemas quanto a não normalidade dos resíduos (BODRO et al., 2019). Em suma, a transformação Dual power foi estabelecida com o intuito de suprir as limitações expostas pela estratégia de transformação Box-Cox (BOX; COX, 1964) com relação ao truncamento no parâmetro de transformação  $\lambda$ .

Entretanto, assim como no método de Box-Cox, a análise Dual power também é fundamentada por uma potência de transformação  $\lambda$ . Todavia, o

diferencial entre esses métodos reside no fato de que a transformação de Box-Cox não assume qualquer transformação aos dados quando  $\lambda = 1$ . A Dual power, que devido a sua singularidade de potência dupla, sempre aplicará transformação nos dados, independentemente do valor assumido por  $\lambda$  (YANG, 2006; BODRO et al., 2019).

Porém, embora apresente conceitos bem fundamentados, a aplicação da técnica Dual power de tratamento de dados ainda não é difundida em estudos alométricos. Apesar disso, esse método pode fornecer novas percepções em estudos dendrométricos de predição, haja vista que as premissas paramétricas são fundamentais nesse tipo de análise.

### 3.2.10 Transformação G power

A estratégia de transformação G power ou de poder generalizado, desenvolvida por Kelmansky et al. (2013), representa uma extensão melhorada da transformação Glog (DURBIN et al., 2002), criticada por promover transformações muito severas em valores altos. Diante disso, o método G power de transformação propõem suavizar os efeitos severos promovidos pelo método Glog em distribuições com assimetrias leves (KELMANSKY; RICCI, 2017).

Sumariamente, a transformação G power é indicada para estabilização da variância (KELMANSKY et al., 2013) e atendimento de normalidade em distribuições assimétricas (ROJAS-PERILLA, 2018, MEDINA et al., 2019), podendo ser aplicada para valores positivos e negativos (KELMANSKY; RICCI, 2017), inclusive, com caudas mais pesadas ou mais leves do que a distribuição normal (TSAI et al., 2017). Em análise comparativa, a transformação G power demonstra melhores resultados para a adequação da normalidade em comparação ao método Glog (KELMANSKY; RICCI, 2017).

No entanto, assim como a maioria das técnicas de transformação descritas na literatura, poucos são os conhecimentos referentes a performance da família de transformação G power em estudos alométricos, especialmente relacionados com a ciência florestal. Dessa forma, consolidar conhecimentos a respeito dessa técnica para adequação das suposições estatísticas paramétricas: normalidade dos resíduos e homocedasticidade das variâncias, é



fundamental para aprimoramento do uso de transformação de variáveis em estudos alométricos com regressão linear.

### 3.2.11 Transformação Log shift

Essa transformação tem base fundamentada no método logarítmico. Entretanto, a diferença principal entre a transformação logarítmica e a Log shift é a escolha da transformação adaptativa conduzida pelos dados. Essa adaptação ocorre em virtude da estimativa de um parâmetro de transformação, denotado por  $\lambda$ . Assim, a transformação logarítmica representa apenas um caso especial dessa família de transformações na qual ocorre quando  $\lambda = 0$ .

De modo geral, a transformação Log shift, que envolve a estimativa ideal de um parâmetro de transformação, é indicada principalmente para corrigir distorções de normalidade dos resíduos (PERE, 2000; FENG et al., 2017; ROJAS-PERILLA, 2018; MEDINA et al., 2019). Apesar disso, na literatura existem aplicações promissoras dessa transformação para a correção simultânea da ausência de normalidade residual e homogeneidade das variâncias (CHINN, 1996). Sabe-se, porém, que nem sempre o uso de transformação vai surtir os efeitos esperados. Logo, confiar que uma transformação responderá as suposições exigidas é uma concepção errônea, haja vista que os resultados dependem muito das características das variáveis.

Contudo, na ciência florestal, apesar da difundida aplicação da transformação logarítmica em variáveis dendrométricas (CHAKRABORTY et al., 2016; FICHER, 2016; CYSNEIROS et al., 2020a), os conhecimentos sobre o uso da técnica Log shift ainda são nulos. Dessa forma, o entendimento do efeito dessa transformação em estudos dendrométricos pode representar uma nova alternativa para tratamentos de variáveis, visando a adequação das suposições estatísticas paramétricas.

### 3.2.12 Transformação raiz quadrada deslocada

A transformação da raiz quadrada deslocada, ou com mudança orientada a dados, foi projetada seguindo a mesma proposta da transformação Log shift, na qual se aplica um parâmetro de transformação orientado aos dados.

Porém, apesar de Bartlett (1937) haver sugerido um deslocamento determinístico para dados pequenos quando se aplica a transformação de raiz quadrada tradicional, a primeira implementação da transformação de raiz quadrada deslocada foi, inicialmente, proposta por Medina et al. (2018). Todavia, esses autores relatam ainda que a aplicação da transformação raiz quadrada deslocada não é promissora para correção da homocedasticidade.

Assim, devido a sua recente introdução na literatura, essa transformação ainda não apresenta consolidação na bibliografia estatística. Entretanto, com relação a transformação convencional de raiz quadrada, que é a base da metodologia em questão, há, na literatura, conceitos bem definidos sobre o seu uso nas mais diversas áreas de conhecimento (BARTLETT, 1937; COHEN et al., 2014; YANG et al., 2016; GHOSH et al., 2021). Diante disso, a aplicação da raiz quadrada deslocada é uma forma de tratamento de dados ainda não utilizada em estudos dendrométricos, o que representa um campo de conhecimento a ser preenchido, especialmente em relação a modelagens lineares.

## 4 MATERIAL E MÉTODOS

### 4.1 BASES DE DADOS

Buscou-se neste estudo realizar modelagens lineares das variáveis dendrométricas altura, volume e biomassa. Nesse sentido, testou-se exclusivamente o ajuste de modelos de regressão linear simples, nos quais priorizou-se a transformação das variáveis dependentes.

Para as modelagens da relação altura-diâmetro e de volume, optou-se, em todos os cenários testados, aplicar unicamente o diâmetro a 1,3 m do solo ( $d$ ) como variável explicativa, visto que esse atributo é o de maior correlação com as variáveis de interesse modeladas. Por outro lado, para as modelagens alométricas de biomassa, foram ajustados modelos lineares simples, dispondo como variáveis independentes a altura das árvores ( $h$ ) e o  $d$ . Dessa forma, em cada um desses cenários, foram testados o efeito das estratégias de transformação para explicação da variável dependente.

Assim, com a intenção de dimensionar o êxito das estratégias de transformação em modelagens lineares com dados de diferentes naturezas, as bases de dados abordadas contemplaram informações dendrométricas de florestas plantadas e naturais sob diferentes condições. Destaca-se que, para as modelagens volumétricas, foram usadas informações dendrométricas de florestas naturais da Amazônia e da Mata Atlântica, além de florestas plantadas de *Pinus oocarpa* Schiede ex Schldl. em idades de 5 e 19 anos. Para as modelagens lineares da relação altura-diâmetro, foram usados também dados de florestas plantadas de *Pinus oocarpa* sob os mesmos estratos de idade. Enquanto isso, nas modelagens de biomassa, os dados utilizados foram provenientes de uma amostragem de árvores de *Mimosa scrabela* Benth., advindas de povoamentos nativos.

Dessa forma, diante de bases de dados com origem e método de obtenção distintos, as modelagens com variáveis transformadas buscaram contemplar variados cenários factíveis em atividades de manejo florestal, na qual a estratégia de regressão linear pode ser aplicada. As características intrínsecas à cada uma dessas bases de dados, no que tange as particularidades de cada

variável dendrométrica abordada, bem como as informações metodológicas de obtenção, estão descritas nos subtópicos seguintes.

#### 4.1.2 Florestas plantadas de *Pinus oocarpa*

A silvicultura com espécies exóticas atualmente assume relevante importância socioeconômica no Brasil, tornando-se imprescindível para a indústria nacional de base florestal, pela geração de matéria prima para diversos setores, tais como celulose, laminação, serraria, compensados e energia. De acordo com o Instituto Brasileiro de Geografia (IBGE), o setor de florestas plantadas nacional alcançou taxas de 10 milhões de hectares cultivados no ano de 2019, dos quais cerca de 2,0 milhões de hectares correspondem exclusivamente aos empreendimentos florestais com o gênero *Pinus*, distribuídos predominantemente nas regiões Centro-Oeste, Sudeste e Sul do Brasil (IBGE, 2019).

Diante desse cenário de constante crescimento de empreendimentos com florestas equiâneas no Brasil, a quantificação eficiente da produção florestal se faz cada vez mais necessária. Nesse sentido, para as modelagens hipsométricas e volumétricas em florestas plantadas, foram usadas bases de dados de dois plantios de *Pinus oocarpa* com estratos de idade contrastantes, especificamente sob as idades de 5 e 19 anos, instalados no município de Agudos, região Sudoeste do estado de São Paulo. Apenas o povoamento com 19 anos recebeu intervenção por desbaste, aos 8, 10, 11, 12 e 15 anos.

Segundo a Classificação Climática de Köppen, o clima da região é do tipo Aw - Clima tropical com inverno seco, a temperatura anual varia de 19 a 21°C e o índice pluviométrico anual é de aproximadamente 1.340 mm (RODRIGUES et al., 2016). A vegetação natural local é classificada como Floresta Ombrófila Densa, com altitude de aproximadamente 900 m (MACHADO et al., 2005; IBGE, 2012).

A amostragem consistiu na medição de 100 árvores por plantio aos 5 e 19 anos, das quais foram mensurados os diâmetros a 1,3 m do solo ( $d$ ) em centímetros e as alturas totais ( $h$ ) em metros para as modelagens das relações hipsométricas. Para as modelagens volumétricas, foi aplicado a cubagem rigorosa das árvores amostradas, que consistiu na medição dos diâmetros ao

longo dos fustes nas posições de 0,1 m; 1,3 m e sucessivamente em intervalos fixos de 2 m até um diâmetro mínimo de 3,0 cm, de modo que os volumes individuais ( $v$ ) foram obtidos pelo método de Smalian. A descrição estatística das variáveis dendrométricas diâmetro, altura e volume é apresentada na Tabela 1.

TABELA 1 - ESTATÍSTICA DESCRITIVA DAS VARIÁVEIS DENDROMÉTRICAS DE POVOAMENTOS EQUIÂNEOS DE *Pinus oocarpa* UTILIZADAS NAS MODELAGENS VOLUMÉTRICAS E DA RELAÇÃO ALTURA-DIÂMETRO.

Variável	Idade (anos)							
	5				19			
	mínimo	média	máximo	desvio padrão	mínimo	média	máximo	desvio padrão
$d$ (cm)	4,4	11,75	20,4	3,11	20,5	30,16	42,2	5,20
$h$ (m)	5,15	9,63	12,68	1,55	23,0	26,38	28,9	1,44
$v$ (m <sup>3</sup> )	0,005	0,064	0,191	0,037	0,398	0,933	1,962	0,358

FONTE: O autor (2022).

A cubagem rigorosa pelo método de Smalian foi executada em função do comprimento de cada seção e das áreas basais de suas extremidades (1), de modo que o volume total do fuste foi obtido pelo somatório dos volumes de cada seção (FINGER, 2006; MACHADO; FIGUEIREDO-FILHO, 2009; AVERY; BURKHART, 2015).

$$v_i = \left( \frac{g_1 + g_2}{2} \right) \cdot l \quad \therefore g_1 = \frac{\pi(d_1)^2}{40.000} \quad \therefore g_2 = \frac{\pi(d_2)^2}{40.000} \quad (1)$$

Em que:  $v$  = volume da seção (m<sup>3</sup>);  $g_1$  = área transversal na base da seção  $i$  da tora (m<sup>2</sup>);  $g_2$  = área transversal no topo da seção  $i$  da tora (m<sup>2</sup>);  $l$  = comprimento da tora (m);  $i$  = número de seções da seção ( $i = 1, 2, 3, \dots, n$ ),  $d_1$  = diâmetro na base da seção (cm);  $d_2$  = diâmetro no topo da seção (cm).

#### 4.1.3 Floresta natural amazônica

A base de dados utilizada para as modelagens volumétricas foi obtida de uma área de concessão florestal situada na Floresta Nacional do Jamari (Flona do Jamari). Essa área está localizada ao Norte do estado de Rondônia, em um território que abrange 223.086,27 hectares (BRASIL, 2000), fazendo

divisas com os municípios de Candeias do Jamari, Itapuã do Oeste e Cujubim, sob as coordenadas geográficas 09° 00' 00" S a 09° 30' 00" S e 62° 44' 05" W a 63° 16' 64" W (MANIESI et al., 2017). O relevo é predominantemente plano a suavemente ondulado, com um clima Tropical Chuvoso do tipo Am (Köppen) (ALVARES et al., 2013; KERBER et al., 2021), típico para a região amazônica (THAINES et al., 2010; ALVARES et al., 2013; SANTOS et al., 2021b). Os períodos de inverno e seca são bem definidos, com média anual de precipitação em torno de 2.400 mm e temperatura média anual de 25°C, predominam na região as tipologias de Floresta Ombrófila, variando de Aberta à Densa (CYSNEIROS et al., 2017).

As informações dendrométricas usadas nas modelagens foram provenientes da cubagem em romaneio das árvores exploradas nas Unidades de Produção Anual dos anos de 2014 e 2015, gerando uma amostragem de 5.204 árvores distribuídas entre 33 espécies. No processo de cubagem, foi utilizado o método de Smalian (1) descrito anteriormente para estimativa volumétrica de florestas plantadas. Destaca-se que esse procedimento foi executado para obtenção do volume comercial da árvore, o qual é representado pelo volume entre a base do fuste até o ponto de inversão morfológica. As estatísticas descritivas das variáveis dendrométricas, volume e diâmetro, usadas nas modelagens estão descritas na Tabela 2.

TABELA 2 - ESTATÍSTICA DESCRITIVA DAS VARIÁVEIS DENDROMÉTRICAS DAP E VOLUME DE UMA FLORESTA NATURAL DA AMAZÔNIA.

<i>n</i>	<i>d</i> (cm)				<i>v</i> (m <sup>3</sup> )			
	mínimo	média	máximo	desvio padrão	mínimo	média	máximo	desvio padrão
5.204	50	81,14	197	21,46	1,6031	7,7907	38,0387	5,3645

FONTE: O autor (2022).

#### 4.1.4 Floresta natural atlântica

A base de dados que representa o domínio da Mata Atlântica foi obtida a partir da cubagem por métodos não-destrutivos de árvores em diferentes fragmentos florestais situados no estado do Rio de Janeiro (CYSNEIROS, 2020). Assim, foram estabelecidos 31 pontos de amostragem sob o domínio Mata Atlântica para a realização da cubagem das árvores. As amostras foram

alocadas em áreas públicas e particulares de diferentes categorias de conservação.

Ao todo, foram cubadas 583 árvores distribuídas entre 84 espécies, pertencentes exclusivamente às três principais tipologias florestais do domínio Mata Atlântica, que correspondem a Floresta Estacional Semidecidual, Floresta Ombrófila Densa e Restinga Arborizada (IBGE, 2012). Quanto à amostragem, 264 árvores pertenceram à Floresta Ombrófila Densa, 209 à Floresta Estacional Semidecidual e 110 à Restinga. Destaca-se, que essas tipologias representam cerca de 98% de cobertura da área do estado e estão presentes em altitudes que variam de 5 a 900 m (SFB, 2018).

Para a região de estudo, de acordo com a classificação de Köppen o clima varia de Tropical (Af, Am e Aw) a Subtropical Úmido (Cfa, Cfb, Cwa e Cwb), com predomínio dos climas Aw - Tropical com Inverno Seco e Cwa - Subtropical com Verão Quente. A precipitação anual varia de 1.000 a 2.200 mm e a temperatura média anual varia de 12 a 24°C (ALVARES et al., 2013).

Das árvores amostradas, foram coletadas as variáveis dendrométricas: diâmetro a 1,3 m do solo ( $d$ ), bem como os diâmetros nas alturas ( $d_i$ ) 0,3 m; 0,5 m; 1 m; 1,3 m; 2 m e a partir desta seção a cada 1 m até a última seção do tronco no ponto de inversão morfológica das árvores. Os volumes das seções foram calculados pelo método de Smalian (1), em que o somatório volumétrico de todas as seções correspondeu ao volume do fuste ( $v$ ). A Tabela 3 detalha as informações das estatísticas descritivas das variáveis mensuradas.

TABELA 3 - ESTATÍSTICA DESCRITIVA DAS VARIÁVEIS DENDROMÉTRICAS DAP E VOLUME DE UMA FLORESTA NATURAL DA MATA ATLÂNTICA.

$n$	$d$ (cm)				$v$ (m <sup>3</sup> )			
	mínimo	média	máximo	desvio padrão	mínimo	média	máximo	desvio padrão
583	10,66	25,04	76,11	9,3	0,028	0,413	3,752	0,409

FONTE: O autor (2022).

#### 4.1.5 Floresta natural de *Mimosa scrabella* Benth

A base de dados com informações de biomassa para os ajustes dos modelos lineares foi obtida de um povoamento nativo de *Mimosa scrabella* Benth. (bracatinga) situado na Região Metropolitana de Curitiba. De acordo com

a classificação de Köppen, o clima dessa região é do tipo Cfb, que corresponde ao clima temperado, mesotérmico úmido, com as quatro estações bem definidas e sem períodos anuais de seca definidos. Os verões são quentes e chuvosos e os invernos frios com eventuais períodos secos, a temperatura média anual se apresenta em torno de 17°C, com mínimas de 11°C e máximas de 24°C. A precipitação média anual se apresenta entre 1.400 e 1600 mm, com leve diminuição no inverno (SOUZA et al., 2014).

Destaca-se que a espécie florestal estudada, *Mimosa scrabella*, caracteriza-se como pioneira, assim, desenvolve-se bem em pleno sol. Ademais, apresenta crescimento rápido e resistência a geadas (URBANO et al., 2008; MACHADO et al., 2011; SILVA et al., 2016). Economicamente, essa espécie é reconhecida principalmente por seu potencial energético, sendo amplamente usada na região como lenha e carvão, além de representar fonte de renda para pequenos agricultores que a cultivam por meio de sistemas agroflorestais (URBANO et al., 2021).

Para a obtenção da biomassa, foi realizada uma amostragem destrutiva de um total de 30 árvores de *Mimosa scrabella*, de modo que as massas das porções dos fustes, galhos, folhas, cascas e raízes foram mensuradas separadamente. Além das massas de cada componente, foram mensurados os diâmetros a 1,3 m do solo ( $d$ ) e as altura totais ( $h$ ) das árvores amostradas. As estatísticas descritivas dessas variáveis dendrométricas estão descritas na Tabela 4.

TABELA 4 - ESTATÍSTICA DESCRITIVA DAS VARIÁVEIS DENDROMÉTRICAS DAP E ALTURA DE UMA AMOSTRAGEM DE ÁRVORES DE *Mimosa scrabella* USADAS PARA MODELAGENS DE BIOMASSA TOTAL.

$n$	$d$ (cm)				$h$ (m)			
	mínimo	média	máximo	desvio padrão	mínimo	média	máximo	desvio padrão
30	4,3	11,22	23,01	5,59	6,9	12,71	19,55	3,67

FONTE: O autor (2022).

Para determinação da biomassa seca dos componentes, foram coletadas amostras logo após a mensuração da massa úmida. Esse procedimento foi realizado com o auxílio de uma balança digital com precisão de 0,1 g. Para o fuste, foi coletada para cada árvore uma amostra transversal do



lenho, retirada na metade da sua altura total. Enquanto isso, para os galhos, folhas, cascas e raízes, foram obtidas subamostras aleatórias de aproximadamente 1 kg para cada fragmento da árvore. As amostras foram secadas em estufa com circulação e renovação de ar a  $65 \pm 5^\circ\text{C}$  até atingir massa constante. A biomassa dos componentes galhos, folhas, cascas e raízes foram obtidas conforme equação descrita em (2).

$$w_c = \frac{mu_c \cdot ms_a}{mu_a} \quad (2)$$

Em que  $w_c$  é biomassa seca do componente (galhos, folhas, cascas ou raízes) para a  $i$ -ésima observação (kg);  $mu_c$  é a massa úmida total do componente (fustes, galhos, folhas, cascas ou raízes) para a  $i$ -ésima observação (kg);  $ms_a$  é a massa seca da amostra do componente (galhos, folhas, cascas ou raízes) para a  $i$ -ésima observação (kg); e  $mu_a$  é a massa úmida da amostra do componente (galhos, folhas, cascas ou raízes) para a  $i$ -ésima observação (kg).

Por outro lado, para a determinação da biomassa do fuste das árvores amostradas, considerou-se a densidade básica da madeira ( $\rho$ ), obtida, por sua vez, pela razão entre a massa seca e o volume das amostras de seções transversais coletadas dos fustes. Assim, com base na densidade básica média ( $\bar{\rho}$ ) e nos volumes dos fustes ( $v$ ), obtidos via cubagem em campo, obteve-se a biomassa do fuste pela seguinte expressão (3).

$$w_{fuste} = \bar{\rho} \cdot v \quad (3)$$

Dessa forma, a biomassa total ( $w_t$ ) em nível de árvore foi obtida a partir do somatório dos estoques de biomassa de todos os componentes. A descrição estatística das biomassas dos fragmentos amostrados, assim como da biomassa total das árvores, é apresentada na Tabela 5.

TABELA 5 - ESTATÍSTICA DESCRITIVA DA BIOMASSA DOS COMPONENTES FUSTE, GALHOS, FOLHAS, CASCA E RAÍZES DE ÁRVORES DE *Mimosa scrabella*.

Componente	Biomassa (kg)			
	mínimo	média	máximo	desvio padrão
Fuste	1,91	47,86	171,64	53,54
Galhos	0,55	14,57	67,37	17,56
Folhas	0,21	2,84	8,20	2,36
Cascas	0,27	5,94	21,76	6,84
Raiz	0,68	12,54	54,58	14,97
Total	3,85	83,75	323,53	93,00

FONTE: O autor (2022)

## 4.2 MÉTODOS DE TRANSFORMAÇÃO DE VARIÁVEIS

Neste estudo, foram abordadas três classes de modelos alométricos considerados primordiais à ciência florestal, especificamente para predição da altura, volume e biomassa de povoamentos florestais. Todavia, vale ressaltar que foram usados exclusivamente modelos lineares simples em todas as etapas de modelagens, em que se usou unicamente o diâmetro com variável explicativa para os ajustes da relação altura-diâmetro e de volume. Por outro lado, para os ajustes da biomassa total de árvores, bem como de seus componentes – fuste, galhos, folhas, casca e raízes – considerou-se a altura e o diâmetro como as variáveis explicativas das modelagens.

Dessa forma, para a adequação das premissas da análise de regressão, aplicou-se diferentes estratégias de transformação nas variáveis dendrométricas de cada modelo, as quais buscaram, sobretudo, o atendimento simultâneo dos pressupostos estabelecidos. Vale destacar que, nas estratégias de transformações potenciais que utilizam um parâmetro lambda ( $\lambda$ ), as transformações foram aplicadas exclusivamente nas variáveis respostas. O parâmetro de transformação  $\lambda$  para essas estratégias foi estimado pelo *software* R, versão 4.0.5 (R Core Team, 2021), por meio do pacote *trafo*, versão 1.0.1 (MEDINA et al., 2018), mediante o método de máxima verossimilhança.

No entanto, para as estratégias na qual não foi necessária a estimativa de  $\lambda$ , o procedimento de transformação foi executado diretamente no momento do ajuste das duas seguintes formas: (i) transformando-se apenas a variável resposta do referido modelo e (ii) transformando-se simultaneamente a variável resposta e a explicativa. Assim, ao todo foram testadas 15 estratégias de

modelagens, incluindo a estruturada com as variáveis em sua escala original. Essas estratégias estão elencadas na Tabela 6, sendo ainda abordadas em mais detalhes nos subtópicos seguintes.

TABELA 6 - ESTRATÉGIAS DE MODELAGEM LINEAR SIMPLES ENVOLVENDO TRANSFORMAÇÕES DE VARIÁVEIS.

Estratégia	Denominação	Modelo transformado
1	Variáveis originais	$Y = \beta_0 + \beta_1 X$
2	Log	$\log(Y) = \beta_0 + \beta_1 X$
3	Recíproca ou Inversa	$\frac{1}{y} = \beta_0 + \beta_1 X$
4	Box-Cox	$\begin{cases} \frac{y^{\lambda}-1}{\lambda} = \beta_0 + \beta_1 X & \text{se } \lambda \neq 0 \\ \log(Y) = \beta_0 + \beta_1 X & \text{se } \lambda = 0 \end{cases}$
5	Manly	$\begin{cases} \frac{\exp^{\lambda y} - 1}{\lambda} = \beta_0 + \beta_1 X & \text{se } \lambda \neq 0 \\ y = \beta_0 + \beta_1 X & \text{se } \lambda = 0 \end{cases}$
6	Bickel-Doksum	$\frac{ y ^{\lambda} \text{sinal}(y) - 1}{\lambda} = \beta_0 + \beta_1 X \quad \text{se } \lambda > 0$

Continua...

Tabela 6 - Continuação...

Estratégia	Denominação	Modelo transformado
7	Yeo-Johnson	$\begin{cases} \frac{(y+1)^\lambda - 1}{\lambda} = \beta_0 + \beta_1 X & \text{se } \lambda \neq 0, y \geq 0 \\ \log(y+1) = \beta_0 + \beta_1 X & \text{se } \lambda = 0, y \geq 0 \\ \frac{(1-y)^{2-\lambda} - 1}{\lambda - 2} = \beta_0 + \beta_1 X & \text{se } \lambda \neq 2, y < 0 \\ -\log(1-y) = \beta_0 + \beta_1 X & \text{se } \lambda = 2, y < 0 \end{cases}$
8	Glog	$\log(y + \sqrt{y^2 + 1}) = \beta_0 + \beta_1 X$
9	Dual power	$\begin{cases} \frac{(y^\lambda - y^{-\lambda})}{\lambda} = \beta_0 + \beta_1 X & \text{se } \lambda \neq 0 \\ \log(y) = \beta_0 + \beta_1 X & \text{se } \lambda = 0 \end{cases}$
10	G power	$\begin{cases} \frac{(y + \sqrt{y^2 + 1})^\lambda - 1}{\lambda} = \beta_0 + \beta_1 X & \text{se } \lambda \neq 0 \\ \log(y + \sqrt{y^2 + 1}) = \beta_0 + \beta_1 X & \text{se } \lambda = 0 \end{cases}$
11	Log shift	$\log(y + \lambda) = \beta_0 + \beta_1 X$
12	Raiz quadrada deslocada	$\sqrt{y + \lambda} = \beta_0 + \beta_1 X$
13	Log-Log	$\log(y) = \beta_0 + \beta_1 \log(X)$
14	Recíproca-Recíproca ou Inversa – Inversa	$\frac{1}{y} = \beta_0 + \beta_1 \frac{1}{X}$

Continua...

Tabela 6 - Continuação...

Estratégia	Denominação	Modelo transformado
15	Raiz-Raiz	$\sqrt{Y} = \beta_0 + \beta_1\sqrt{X}$

em que  $Y$  é a variável resposta do modelo;  $X$  é a variável explicativa do modelo;  $\beta_0$  e  $\beta_1$  são os coeficientes de regressão do modelo linear a serem estimados; e  $\lambda$  é o parâmetro de transformação estimado pelo método de máxima verossimilhança.

FONTE: O autor (2022).

#### 4.2.1 Transformação Logarítmica (Log e Log–Log)

Essa metodologia de transformação é talvez a mais empregada pelos pesquisadores que utilizam a técnica de análise de regressão sobre as mais diversas áreas de conhecimento (OSBORNE, 2010; FENG et al., 2014; FISCHER, 2016; MEDINA et al., 2019). Assim, a transformação Log é a mais popular entre os diferentes tipos de transformações usadas para transformar variáveis em modelos com distribuição de erros não normais e variabilidades desproporcionais (FENG et al., 2014). Neste estudo, ela foi abordada de duas formas: em uma primeira ocasião, usada apenas a variável resposta (4), e, em um segundo momento, aplicada nas variáveis resposta e explicativa, simultaneamente (5).

$$\log(Y) = \beta_0 + \beta_1 X \quad (4)$$

$$\log(Y) = \beta_0 + \beta_1 \log(X) \quad (5)$$

Dessa forma, o procedimento de transformação inversa da variável à escala original consistiu na potenciação dos dados transformados, obtendo-se, portanto, estimativas na escala natural, conforme descrito na expressão 6.

$$y' = \log(Y) \quad \therefore \quad Y = \exp^{y'} \quad (6)$$

#### 4.2.2 Transformação recíproca

A transformação recíproca ou inversa é muito usada na estatística com a finalidade de estabilizar a variância em grandes dispersões de dados, devido a sua capacidade de minimizar o efeito de possíveis valores muito altos da variável de interesse (TUKEY, 1977; OSBORNE, 2010). Essa transformação pode ser útil em situações de assimetria negativa da variável, isso quando é direcionada para estabilização de variância nessas ocasiões (HOYLE, 1973, MEDINA et al., 2019).

Assim, a transformação recíproca (7) foi abordada em duas formas no presente estudo. Em uma primeira situação, efetuou-se apenas as transformações das variáveis dependentes em cada modelagem; e, em um segundo momento, obteve-se a transformação simultânea das variáveis dependente e independente dos modelos ajustados. O destaque é que o processo de retransformação dessa estratégia consiste simplesmente em se obter o inverso da variável transformada, conforme abordado na expressão (7).

$$y' = \frac{1}{y} \quad \therefore \quad y = \frac{1}{y'} \quad (7)$$

Portanto, quando a transformação recíproca foi aplicada apenas na variável resposta dos modelos de regressão, as modelagens foram conduzidas de acordo com a expressão (8). Por outro lado, usou-se o modelo (9) quando a transformação foi aplicada nos dois lados da função.

$$\frac{1}{y} = \beta_0 + \beta_1 X \quad (8)$$

$$\frac{1}{y} = \beta_0 + \beta_1 \frac{1}{X} \quad (9)$$

#### 4.2.3 Transformação de Box-Cox

A transformação de Box-Cox (BOX; COX, 1964) pertence a uma classe de transformações potenciais, cujo método eleva os valores das variáveis à um expoente específico estimado a partir das características apresentadas pelos dados (SAKIA, 1992). Nessa ocasião, o expoente de transformação é um valor de lambda ( $\lambda$ ) que tem a finalidade de transformar a variável explicativa, buscando a distribuição normal dos resíduos e homogeneidade das variâncias. Todavia, um fato particular e limitante dessa estratégia é a admissão apenas de variáveis em escala positiva (ROJAS-PERILLA, 2018, MEDINA, et al., 2019), podendo ser aplicada aos dados conforme a expressão (10).

$$y' = \begin{cases} \frac{y^\lambda - 1}{\lambda} & \text{se } \lambda \neq 0 \\ \log(y) & \text{se } \lambda = 0 \end{cases} \quad (10)$$

A transformação Box-Cox é igual a transformação logarítmica quando  $\lambda = 0$ . Nesse sentido, o valor do parâmetro de transformação  $\lambda$  pode ser estimado iterativamente usando o método de máxima verossimilhança (11), na qual o seu termo logarítmico é definido como.

$$L(\lambda) = k - \frac{n}{2} \text{SSE}(z^{(\lambda)}) \quad (11)$$

Sendo  $k$  uma constante de integração  $z^{(\lambda)} = y^{(\lambda)} / y_m^{(\lambda-1)}$  em que  $y_m$  é uma média geométrica da resposta. SSE é a soma de resíduos quadrado da regressão  $z^{(\lambda)}$ , e  $L$  é a função de verossimilhança. Assim, o valor de  $\lambda$  é ótimo quando  $L$  for máximo (BOX; COX, 1964; SAKIA, 1992; FISCHER, 2016).

Portanto, a família de transformações Box-Cox apresenta uma relativa simplicidade operacional. Além disso, na literatura estatística, essa metodologia apresenta geralmente bons resultados quando se busca homogeneizar as variâncias e normalizar a distribuição residual (SERINALDI et al., 2012; FISCHER, 2016), muito embora nem sempre ela atinja simultaneamente essas premissas (MEDINA, et al., 2019; PÉLLICO NETTO; BEHLING, 2021).

Por conseguinte, quando a estimativa de  $\lambda$  pelo método de Box-Cox foi diferente de zero, as modelagens foram estruturadas conforme o modelo (12).

Caso contrário, quando  $\lambda = 0$ , o procedimento adotado foi o mesmo descrito na expressão (4) para transformação Log. Assim, a retransformação para essa estratégia se resume aos mesmos procedimentos da expressão (5). A expressão (13), por sua vez, detalha o procedimento de retransformação quando  $\lambda \neq 0$ .

$$\frac{y^\lambda - 1}{\lambda} = \beta_0 + \beta_1 X \quad (12)$$

$$y = (\lambda \cdot y' + 1)^{\frac{1}{\lambda}} \quad (13)$$

Nesse contexto, a transformação Box-Cox direcional, a partir dos dados, a transformação mais apropriada (BOX; COX, 1964; OSBORNE, 2010), pois, dependendo do valor de  $\lambda$ , qualquer transformação pertencente à família Box-Cox pode ser implementada (SAKIA, 1992; MEDINA et al., 2019). Assim, a maneira simples com que é feita a interpretação desse parâmetro torna o método Box-Cox um dos mais utilizados no tratamento de dados para procedimentos estatísticos. Dessa forma, essa estratégia pode ser usada em qualquer conjunto de dados em que a variável dependente é sempre positiva, sendo, portanto, uma promissora opção em ramos da ciência florestal.

#### 4.2.4 Transformação de Manly

A transformação de Manly (MANLY, 1976) pertence à família exponencial de transformação. Foi proposta, *a priori*, para satisfazer distribuições que abrangessem valores negativos em uma busca alternativa de preencher as lacunas do método de Box-Cox (1964). A estratégia de Manly (1976) é conhecida pela eficácia em transformar distribuições unimodais em quase simétricas (SAKIA, 1992; CHORTIRAT et al., 2011). Assim, a abordagem da transformação de Manly (1976) no presente estudo se resumiu na aplicação dos passos matemáticos descritos na expressão (14).



$$y' = \begin{cases} \frac{\exp^{\lambda y} - 1}{\lambda} & \text{se } \lambda \neq 0 \\ y & \text{se } \lambda = 0 \end{cases} \quad (14)$$

Dessa forma, nas situações em que o parâmetro  $\lambda$ , estimado pelo método de máxima verossimilhança (11), apresentou valores diferentes de zero, as variáveis dendrométricas transformadas foram ajustadas em regressão linear de acordo com o modelo a seguir (15). Em contrapartida, quando  $\lambda = 0$ , nenhuma transformação foi aplicada aos dados (MANLY, 1976; SAKIA, 1992).

$$\frac{\exp^{\lambda y} - 1}{\lambda} = \beta_0 + \beta_1 X \quad (15)$$

Logo, quando  $\lambda \neq 0$ , o procedimento de retransformação para obtenção da escala original das variáveis consistiu na aplicação da expressão (16), por meio de uma abordagem logarítmica da variável transformada. Em regra, adverte-se que, diante de  $\lambda = 0$ , a retransformação não se aplica.

$$y = \frac{\log(\lambda \cdot y' + 1)}{\lambda} \quad (16)$$

#### 4.2.5 Transformação de Bickel-Doksum

Diante da necessidade de tratamento de dados para análises paramétricas, a transformação de Bickel-Doksum (BICKEL; DOKSUM, 1981) surgiu como uma alternativa de substituição à transformação de Box-Cox (1964). Nesse sentido, a vantagem dessa estratégia de transformação, semelhante à outras metodologias concorrentes de Box-Cox, como Manly (1976), também foi estabelecido sobre propósito de assumir valores negativos, preenchendo a limitação do método de Box-Cox. Assim, os procedimentos matemáticos que precedem o uso da transformação de Bickel-Doksum no presente estudo correspondem aos passos descritos na expressão (17).

$$\frac{|y|^\lambda \text{sinal}(y) - 1}{\lambda} \quad \text{se } \lambda > 0 \quad (17)$$

Dessa forma, a transformação de potência Bickel-Doksum foi concebida especialmente para lidar com curtose em vez de assimetria, em particular para distribuições leptocúrticas e platicúrticas (MEDINA, 2019). Os modelos dendrométricos ajustados com base nessa estratégia foram estruturados com a transformação direcionada na variável resposta, conforme o modelo (18). Todavia, para obtenção da variável dependente retransformada, o procedimento de retransformação foi executado pela abordagem matemática da expressão (19).

$$\frac{|y|^\lambda \text{sinal}(y) - 1}{\lambda} = \beta_0 + \beta_1 X \quad (18)$$

$$\pm [\text{sinal}(y)(\lambda y + 1)]^{\frac{1}{\lambda}} \quad (19)$$

#### 4.2.6 Transformação de Yeo-Johnson

A estratégia de transformação Yeo-Johnson foi proposta para adequar à normalidade variáveis cujas distribuições são assimétricas, seja em conjuntos de dados positivos ou negativos (YEO; JOHNSON, 2000). Dessa forma, essa metodologia representa uma nova família de transformação que pode ser usada sem restrições e que possui muitas das boas propriedades da família de transformações potências de Box-Cox. Essa transformação é definida a partir dos valores das variáveis respostas ( $y$ ) e dos parâmetros de transformações ( $\lambda$ ), conforme os procedimentos matemáticos em (20).

$$y' = \begin{cases} \frac{(y+1)^\lambda - 1}{\lambda} & \text{se } \lambda \neq 0, y \geq 0 \\ \log(y+1) & \text{se } \lambda = 0, y \geq 0 \\ \frac{(1-y)^{2-\lambda} - 1}{\lambda - 2} & \text{se } \lambda \neq 2, y < 0 \\ -\log(1-y) & \text{se } \lambda = 2, y < 0 \end{cases} \quad (20)$$

Vale destacar que, devido as características das variáveis dendrométricas avaliadas, o uso da estratégia de transformação Yeo-Johnson foi fundamentado com base apenas para valores de  $y$  positivos ( $y \geq 0$ ). Dessa forma, com relação as estimativas por máxima verossimilhança dos parâmetros  $\lambda$ , os modelos ajustados foram estruturados conforme a expressão (21) quando  $\lambda \neq 0$ . Todavia, quando  $\lambda \neq 0$ , a estrutura do modelo ajustado foi a mesma da expressão (3) para uso de transformação logarítmica em  $y$ .

$$\frac{(y + 1)^\lambda - 1}{\lambda} = \beta_0 + \beta_1 X \quad (21)$$

Assim, quando  $\lambda = 0$ , a prática de retransformação da variável regressada passou pelo mesmo procedimento descrito na abordagem logarítmica (5). Entretanto, quando se ajustou o modelo (21) para  $\lambda \neq 0$ , a escala original da variável dependente foi obtida sobre o custo da retransformação descrita pela função (22) a seguir.

$$y = (\lambda \cdot y' + 1)^{\frac{1}{\lambda}} - 1 \quad (22)$$

#### 4.2.7 Transformação Glog

A transformação de dados Glog, também conhecida como transformação do logaritmo generalizado, foi proposta por Durbin et al. (2002), sendo recomendada também por Huber et al. (2003) para normalização e estabilização de variâncias de valores próximos a zero. Embora seja uma estratégia de transformação adequada para corrigir a não normalidade e não linearidade, é mais amplamente usada como uma transformação de estabilização de variância (ROJAS-PERILLA, 2018). Dessa forma, a transformação Glog é fundamentada em uma modificação da transformação de base logarítmica, sendo abordada no estudo conforme descrito na expressão (23) a seguir.

$$y' = \log(y + \sqrt{y^2 + 1}) \quad (23)$$

Assim, a transformação Glog foi aplicada exclusivamente às variáveis respostas de cada modelagem avaliada. Nesta conjectura, os modelos ajustados sobre a ótica Glog de transformação apresentaram a estrutura de ajuste descrita no modelo (24). Entretanto, como procedimento de retransformação para a estimativa da variável de interesse em sua escala original, foi usada a expressão (25).

$$\text{Log}\left(y + \sqrt{y^2 + 1}\right) = \beta_0 + \beta_1 X \quad (24)$$

$$y = -\frac{[1 - (\exp^{y'})^2]}{2\exp^{y'}} \quad (25)$$

#### 4.2.8 Transformação Dual power

A transformação Dual power foi proposta por Yang (2006) motivada pela limitação da estratégia de Box-Cox (1964) relacionada ao truncamento no parâmetro de transformação determinado por  $\lambda$ . Portanto, para lidar com esse problema, Yang (2006) propôs recentemente a transformação Dual power, em que, no presente estudo, essa estratégia foi aplicada às variáveis dependentes conforme a expressão (26).

$$y' = \begin{cases} \frac{(y^\lambda - y^{-\lambda})}{2\lambda} & \text{se } \lambda \neq 0 \\ \log(y) & \text{se } \lambda = 0 \end{cases} \quad (26)$$

Ao contrário da transformação de potência Box-Cox, que mantém os dados sem transformação quando  $\lambda = 1$ , a potência Dual power sempre transforma os dados, independentemente do valor que  $\lambda$  assume. Assim, diante de  $\lambda \neq 0$ , as modelagens fundamentadas sob essa técnica de transformação foram abordadas com base no modelo a seguir (27). Todavia, para  $\lambda = 0$ , a transformação log foi usada, conforme expressão (3), anteriormente citada.

$$\frac{(y^\lambda - y^{-\lambda})}{2\lambda} = \beta_0 + \beta_1 X \quad (27)$$

Nessa perspectiva, quando  $\lambda \neq 0$ , a transformação inversa da estratégia Dual power foi obtida conforme as operações descritas na expressão (28), fornecendo estimativas da variável resposta na sua escala original. No entanto, para  $\lambda = 0$ , foram adotados os mesmos procedimentos descritos na expressão (6) diante do uso da transformação Log.

$$y = \left[ \sqrt{1 + \lambda^2 y'^2} + \lambda y' \right]^{\frac{1}{\lambda}} \quad (28)$$

#### 4.2.9 Transformação G power

Como alternativa para a transformação Box-Cox (1964), Kelmansky et al. (2013) e Kelmansky e Ricci (2017) propuseram uma extensão da transformação Glog, denominada de transformação G power. Em sua ação, ela permite valores negativos e com caudas mais pesadas ou mais leves do que a distribuição normal (TSAI et al., 2017). G power é uma transformação propícia para correção de suposições de normalidade e homocedasticidade (ROJAS-PERILLA, 2018, MEDINA et al., 2019), foco principal do presente estudo. Logo, a sua aplicação foi conduzida conforme a expressão (29).

$$y' = \begin{cases} \frac{(y + \sqrt{y^2 + 1})^\lambda - 1}{\lambda} & \text{se } \lambda \neq 0 \\ \log(y + \sqrt{y^2 + 1}) & \text{se } \lambda = 0 \end{cases} \quad (29)$$

A estratégia G power foi conduzida aos dados a partir de um parâmetro de transformação, como em Box-Cox (1964). Todavia, o destaque desse método é que, para  $\lambda = 0$ , a estratégia de transformação recai na transformação Glog (DURBIN et al., 2002), usando-se:  $\log(y + \sqrt{y^2 + 1})$  (expressão 23). Dessa forma, para essa ocasião, o procedimento de ajuste dos dados deve ser o

mesmo descrito pelo modelo citado na expressão (24). No entanto, quando  $\lambda \neq 0$ , o procedimento de ajuste foi baseado na seguinte estruturação (30).

$$\frac{(y + \sqrt{y^2 + 1})^\lambda - 1}{\lambda} = \beta_0 + \beta_1 x \quad (30)$$

Assim, para se obter a restransformação da variável dependente a função aplicada aos dados transformados foi a descrita em (31).

$$y = - \left[ \frac{1 - (\lambda y' + 1)^{\frac{2}{\lambda}}}{2(\lambda y' + 1)^{\frac{1}{\lambda}}} \right] \quad (31)$$

#### 4.2.10 Transformação Log shift

A transformação Log shift (YANG, 1995) é uma extensão da transformação logarítmica tradicional, incluindo um parâmetro de transformação  $\lambda$  baseado em dados como segue na expressão (32). Essa estratégia é promissora para correção de normalidade em distribuições não gaussianas (FENG et al., 2016; ROJAS-PERILLA, 2018; ROJAS-PERILLA, 2020), na qual transforma distribuições assimétricas à direita ou esquerda e do tipo log-normal em simétricas. Merece destaque também a sua utilidade no tratamento de dados com assimetria excessiva, heterocedasticidade e observações influentes (CHINN, 1996; FENG et al., 2016).

$$y' = \log(y + \lambda) \quad (32)$$

Entretanto, como as funções de logaritmo requerem entradas positivas, as operações por essa estratégia são modificadas por intermédio de um parâmetro  $\lambda$ , tornando, portando, o método de transformação de dados Log shift válido para qualquer valor de  $y$ . Assim, os modelos dendrométricos sob os efeitos da transformação Log shift foram ajustados conforme a expressão (33).

$$\log(y + \lambda) = \beta_0 + \beta_1 x \quad (33)$$

Continuamente, as variáveis originais foram estimadas pela função inversa da estratégia logarítmica, destacando a subtração do acréscimo  $\lambda$  em cada modelagem efetuada. Assim, a retransformação da estratégia Log shift foi processada conforme a expressão (34).

$$y = \exp^{y'} - \lambda \quad (34)$$

#### 4.2.11 Transformação Raiz quadrada deslocada

O recente método de transformação por raiz quadrada deslocada foi proposto por Medina et al. (2018). Essa estratégia surgiu como uma extensão da transformação de raiz quadrada tradicional, sendo aplicada no presente estudo de acordo com a expressão (35). Todavia, para sua implementação, um parâmetro de transformação adaptativo, denotado por  $\lambda$ , é estimado de acordo com as características de distribuição do conjunto de dados com a teoria de máxima verossimilhança (expressão 11) (ROJAS-PERILLA et al., 2018; MEDINA, et al., 2019). Assim, a transformação de raiz quadrada deslocada é definida como segue.

$$y = \sqrt{y' + \lambda} \quad (35)$$

Essa estratégia foi direcionada nas modelagens exclusivamente às variáveis respostas. Dessa forma, a estrutura dos modelos ajustados foi descrita da seguinte forma (36).

$$\sqrt{y' + \lambda} = \beta_0 + \beta_1 x \quad (36)$$

Assim, simplificada, a retransformação da variável  $y$  do modelo foi obtida com a aplicação da seguinte operação matemática (37).

$$y = (y'^2 - \lambda) \quad (37)$$

#### 4.2.12 Transformação Raiz - Raiz

A transformação fundamentada no uso da raiz quadrada (BARTLETT, 1937) é uma estratégia tradicional de tratamento de dados, amplamente usada em diversas áreas do conhecimento, com destaque às ciências médicas, sociais e naturais, especialmente para alcançar estabilização de variâncias em pares de dados (MEDINA et al., 2019). Essa transformação é comumente aplicada para lidar com distribuições assimétricas à direita, sendo recomendada também para tratar dados que contenham valores extremamente pequenos, especificamente entre 0 e 1 (BARTLETT, 1937; ROJAS-PERILLA, 2018). Dessa forma, a raiz quadrada foi direcionada aos dados originais conforme a expressão (38).

$$y' = \sqrt{y} \quad \therefore \quad x' = \sqrt{x} \quad (38)$$

Todavia, no presente estudo, essa estratégia de transformação foi implementada simultaneamente às variáveis dependentes e independentes dos modelos dendrométricos ajustados. Logo, os modelos ajustados seguiram a seguinte estrutura (39).

$$\sqrt{y} = \beta_0 + \beta_1 \sqrt{x} \quad (39)$$

Assim, a função inversa dessa estratégia para obtenção das variáveis em escala original foi facilmente processada de acordo com a expressão (40), elevando-se ao quadrado as variáveis transformadas.

$$y = y'^2 \quad \therefore \quad x = x'^2 \quad (40)$$

#### 4.3 AVALIAÇÃO ESTATÍSTICA DOS MODELOS DE REGRESSÃO

Antes dos ajustes dos modelos dendrométricos, foram efetuadas análises exploratórias para cada uma das bases de dados estudadas. Nesse procedimento foram realizadas a construção de diagramas de dispersão, bem como histogramas de frequências para as variáveis modeladas. As variáveis



respostas e explicativas dos ajustes foram submetidas à correlação de Spearman para a verificação do nível de correlação entre elas.

Ademais, as análises gráficas foram também avaliadas com as variáveis em escala logarítmica, devido ao uso consagrado dessa transformação em estudos dendrométricos para linearização de variáveis. Assim, os pares de dados de cada classe de modelo estudada (relação altura-diâmetro, volumétrica e biomassa) foram analisados em escala original e logarítmica. Essas análises gráficas exploratórias estão apresentadas nos Apêndices desse estudo.

Nas modelagens, destaca-se que os ajustes com variáveis transformadas foram abordados somente nos casos de violação de pelo menos um dos pressupostos estabelecidos para a análise de regressão linear. Nessas ocasiões, as modelagens com variáveis transformadas que corrigiram simultaneamente as premissas de normalidade dos resíduos e homocedasticidade das variâncias foram confrontadas estatisticamente para a seleção da estratégia de transformação mais adequada à referida classe de modelo dendrométrico (altura-diâmetro, volume ou biomassa).

Os coeficientes dos modelos de regressão linear foram estimados pelo método dos mínimos quadrados ordinários (*OLS – ordinary least square*) a partir da função *lm* do *software* R, versão 4.0.5 (R CORE TEAM, 2021), em que a significância de cada coeficiente do modelo foi avaliada por meio do teste *t* a 5% de probabilidade de erro. Os critérios estatísticos de ajuste e precisão, bem como a análise residual dos ajustes utilizados para a seleção das melhores equações com a inserção de transformações, estão descritos nos subtópicos seguintes.

#### 4.3.1 Métricas de Qualidade de Ajuste

As métricas de qualidade de ajuste são ferramentas estatísticas direcionadas para avaliar o quão bem um modelo se ajusta às observações da amostra coletada. Diante disso, elas foram abordadas em cada modelagem estatística para avaliar a qualidade das predições do modelo. A seguir, estão apresentadas as métricas empregadas para avaliação dos ajustes.

##### 4.3.1.1 Coeficiente de determinação ajustado ( $R^2_{aj}$ )

O *Coefficiente de determinação ajustado* ou R-quadrado ajustado –  $R^2_{aj}$  (Expressão 41), determina a extensão da variância da variável dependente que pode ser explicada pela variável independente (GUJARATI; PORTE, 2011; BRUCE; BRUCE, 2019). Diante disso, quanto maior o  $R^2_{aj}$ , melhor é a equação de regressão gerada, pois, em termos de proporção, esse coeficiente sugere o quanto de variância é explicada pelo modelo. Nessa perspectiva, esse parâmetro estatístico é usado para determinar a eficiência de ajuste de modelos lineares em análise de regressão.

$$R^2_{aj} = 1 - (1 - R^2) \left( \frac{n - 1}{n - p} \right) \quad (41)$$

Em que,  $R^2_{aj}$  é o coeficiente de determinação ajustado;  $R^2$  é o coeficiente de determinação do ajuste;  $n$  é o número de observações amostradas; e  $p$  é o número de coeficientes de regressão do modelo.

#### 4.3.1.2 Erro padrão da estimativa (SEE)

O *Erro padrão da estimativa* (SEE) foi abordado na sua forma absoluta (Expressão 42) e percentual (Expressão 43). Em resumo, essa métrica corresponde a estatística de precisão da modelagem, informando a qualidade do ajuste e o quanto o modelo erra em média ao estimar a variável dependente. Em análise de regressão, é o parâmetro estatístico que representa a medida de dispersão média entre os valores observados e estimados (CAMPOS; LEITE, 2017; MONTGOMERY et al., 2021). Logo, espera-se que seja o menor possível quando se visa a precisão das predições.

$$SEE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - p}} \quad (42)$$

$$SEE\% = \frac{SEE}{\bar{y}} \cdot 100 \quad (43)$$

Em que: SEE é o erro padrão da estimativa absoluto; SEE% é o erro padrão da estimativa percentual;  $y_i$  representa a amostra i-ésima da variável dependente/resposta do modelo;  $\widehat{y}_i$  corresponde a i-ésima estimativa da variável dependente/resposta do modelo;  $\bar{y}$  é a média da variável dependente observada para o conjunto de amostra;  $n$  é o número de observações; e  $p$  o número de coeficientes de regressão do modelo.

#### 4.3.1.3 Critério de informação Bayesiano (BIC)

O *Critério de Informação Bayesiano* (BIC) (Expressão 44) está fundamentado no princípio da parcimônia para seleção de modelos. Essa métrica foi estabelecida por Schwarz (1978), estando, em partes, intimamente relacionado ao Critério de Informação de Akaike (AKAIKE, 1974), no entanto, com penalidades maiores para a inclusão de variáveis adicionais ao modelo (BRUCE; BRUCE, 2019). Essa métrica foi obtida da seguinte forma.

$$\text{BIC} = n \cdot \log \left( \sum_{i=1}^n (y_i - \widehat{y}_i)^2 / n \right) + p \cdot \log(n) \quad (44)$$

Em que  $n$  é o número de observações da amostra;  $p$  é o número de coeficientes de regressão do modelo;  $y_i$  representa a amostra i-ésima da variável dependente/resposta do modelo; e  $\widehat{y}_i$  corresponde a i-ésima estimativa da variável dependente/resposta do modelo.

#### 4.3.2 Análise residual

Os resíduos dos modelos devem estar de acordo com as premissas da análise paramétrica para a validade das estatísticas abordadas nos tópicos anteriores. A confiabilidade da significância dos coeficientes de regressão fornecidos pelo teste  $t$  e a da Análise de Variância pelo teste de  $F$  são atribuídas também ao atendimento dessas premissas (DRAPER; SMITH, 1998; MONTGOMERY, 2012; MONTGOMERY et al., 2021). Diante disso, a análise residual foi considerada como critério determinante para a seleção das equações

(BENNETT et al., 2013) e para a verificação do atendimento aos pressupostos da regressão linear.

A normalidade dos resíduos foi verificada pelo teste de Lilliefors (LILLIEFORS, 1967; RAZALI; WAH, 2011), com o auxílio do pacote *nortest* (GROSS; LIGGES, 2015). Para avaliação da homocedasticidade, usou-se o teste de Breusch-Pagan (BREUSCH; PAGAN, 1979) por meio do pacote *car* (FOX; WEISBERG, 2019). Todos os testes de hipótese usados consideraram  $\alpha = 0,05$ .

Para as melhores equações de cada classe de modelo, foram analisados os gráficos de relação entre a variável estimada e a observada, gráficos de dispersão dos resíduos estudentizados e os gráficos quantil-quantil (Q-Q) a 95% de probabilidade. Dessa forma, a dispersão dos resíduos estudentizados de cada ajuste em associação ao teste de homocedasticidade foi direcionada para verificação da adequação da estabilização das variâncias. Por outro lado, os gráficos Q-Q gerados, juntamente com o teste de normalidade de Lilliefors, foram usados para verificar a adequação paramétrica da distribuição dos resíduos das melhores estratégias de transformação.

Enquanto isso, as qualidades de predição das melhores estratégias de transformação foram comparadas por meio dos gráficos hnp (*half-normal plot*), os quais consideraram 5% de probabilidade na simulação dos envelopes de confiança. Esse procedimento foi realizado por meio do pacote *hnp* (MORAL et al., 2017).

Os processamentos das bases de dados foram efetuados com o *software* R, versão 4.0.5 (R CORE TEAM, 2021). Utilizou-se para o preparo e adequação dos dados o pacote *dplyr* versão 1.1.3 (WICKHAM, 2020), enquanto isso, para a elaboração dos gráficos, empregou-se o pacote *ggplot2* versão 3.3.5 (WICKHAM, 2016).

## 5 RESULTADOS

### 5.1 ANÁLISE EXPLORATÓRIA DOS DADOS

#### 5.1.1 Base de dados de florestas plantadas

Diante dessa base de dados, as relações de dependência entre as variáveis dendrométricas demonstram evidentes associações não lineares crescentes entre os pares de dados altura-diâmetro, volume-diâmetro e volume-altura. Os diagramas de dispersão atribuem para essas variáveis dispersões exponenciais. Esse comportamento é mais evidente para os povoamentos jovens aos 5 anos (Figura 19 do Apêndice 1).

Especificamente, para as relações altura-diâmetro e volume-altura, essa tendência reduz com o avanço da idade. Portanto, verifica-se que a correlação entre a altura das árvores com o volume e o diâmetro é reduzida com o aumento da idade da floresta (Figuras 19 e 20 do Apêndice 1).

Comparando-se as dispersões de variáveis entre o plantio jovem (5 anos) e o maduro (19 anos), observa-se uma redução abrupta nos valores do coeficiente de correlação de Spearman. Apesar disso, nota-se, para o povoamento tardio aos 19 anos, que as variáveis volume e diâmetro continuam expressando fortes correlações ( $\rho = 0,9713$ ). Com base nisso, infere-se que a relação biológica entre essas variáveis não se reduz com o avanço da idade.

A partir dos histogramas de frequência (Figuras 19 e 20 do Apêndice1), a distribuição das variáveis demonstrou um comportamento tendendo à normalidade apenas para o diâmetro aos 5 anos. Nas demais ocasiões, as variáveis apresentaram distribuições assimétricas, principalmente para o povoamento tardio aos 19 anos (Figura 20 do Apêndice1).

Observa-se, ainda, que mesmo sob o efeito da transformação logarítmica, as variáveis não apresentaram uma distribuição de frequência livre de assimetria. Entretanto, nota-se que, com as variáveis logaritmizadas, os diagramas de dispersão tornam-se mais lineares e com variâncias reduzidas. Cabe ressaltar que esse comportamento foi mais evidente para o povoamento jovem aos 5 anos (Figuras 19 e 20 do Apêndice1).

### 5.1.2 Base de dados de florestas naturais da Amazônia

A dispersão entre volume e diâmetro demonstrou uma acentuada variabilidade para essa base de dados (Figura 21 do Apêndice 2). Salienta-se que o diagrama de dispersão entre essas variáveis demonstrou propensão em aumentar a variabilidade das dispersões de acordo com as dimensões das árvores amostradas.

Destaca-se que esse comportamento sugere uma eventual presença de heterocedasticidade entre as variáveis. Apesar disso, evidencia-se que, entre as variáveis volume e diâmetro, a relação apresentada foi linear e crescente. A correlação de Spearman atribuída foi de 0,784, indicando altas propriedades lineares.

Em contrapartida, acerca das distribuições de frequência, constatou-se que, tanto o diâmetro quanto o volume, apresentaram distribuição assimétrica à direita (Figura 21 do Apêndice 2). Destaca-se que esse comportamento é o esperado em variáveis dendrométricas, especialmente daquelas advindas de florestas tropicais.

No entanto, quando elas foram analisadas em escala logarítmica, verificou-se que, além da redução na variabilidade e aumento da linearização, as distribuições de frequência perderam o caráter assimétrico e se aproximaram de uma distribuição gaussiana (Figura 21 do Apêndice 2).

### 5.1.3 Base de dados de florestas naturais da Mata Atlântica

Devido ao método de coleta baseado em amostragem, a dispersão entre volume e diâmetro apresentou baixa variação para essa base de dados. Apesar disso, destaca-se a eminente tendência exponencial associada à presença perceptível de heterocedasticidade (Figura 22 do Apêndice 3).

Esse comportamento observado demonstra que a relação entre o volume médio das árvores e a sua variância não é constante para diferentes valores de diâmetros. Assim, a tendência é de se aumentar a variabilidade para árvores de maiores dimensões.

As distribuições de frequência das variáveis volume e diâmetro, demonstram distribuições com evidente assimetria à direita. Destaca-se que

essa condição reflete o esperado para a tipologia florestal avaliada, visto que, em amostragens de florestas tropicais, a frequência de árvores tende a diminuir para as maiores classes de diâmetro. Além disso, associa-se também às maiores classes de diâmetro aumentos na variabilidade das dispersões (Figura 22 do Apêndice 3).

No entanto, com as variáveis analisadas em escala logarítmica, verifica-se que a relação não linear entre volume e diâmetro é convertida para uma relação fortemente linear e com variâncias reduzidas. Nota-se, ainda, que as distribuições das variáveis dendrométricas logaritmizadas, tornam-se próxima a normal (Figura 22 do Apêndice 3).

#### 5.1.4 Base de dados de biomassa de povoamentos naturais de *Mimosa scrabella*

##### 5.1.4.1 Biomassa total

Na análise exploratória dessa base de dados, a correlação de Spearman entre diâmetro e altura ( $\rho = 0,8689$ ) demonstrou elevadas propriedades lineares (Figura 23 do Apêndice 4). Contudo, quando estas variáveis foram relacionadas à biomassa total ( $w_{total}$ ) as correlações foram ainda maiores, demonstrando valores superiores a 0,9 (Figura 24 do Apêndice 5). Apesar disso, os diagramas de dispersão evidenciaram comportamentos exponenciais entre os pares de dados. Destaca-se, ainda, que essa observação foi mais evidente entre as variáveis  $w_{total}$  e altura.

Não obstante, enfatiza-se a alta variabilidade na relação entre  $w_{total}$  e a altura das árvores, o que demonstra evidentes propensões à heterocedasticidade, devido à inconstante dispersão entre esses pares de dados. Todavia, com as variáveis transformadas para a escala logarítmica, as dispersões foram fortemente linearizadas, especialmente entre as variáveis  $w_{total}$  e diâmetro (Figura 24 do Apêndice 5).

Destaca-se, ainda, que as distribuições de frequência das variáveis diâmetro e altura apresentaram comportamento assimétrico positivo. Cabe enfatizar que esse é o comportamento habitual para esse tipo de amostragem, uma vez que árvores com menores dimensões apresentam maior densidade em povoamentos desta natureza.

Enfatiza-se que a variável altura, em escala logarítmica, demonstrou distribuição de frequência próxima a distribuição normal (Figura 23 do Apêndice 4). Enquanto isso, a distribuição de frequência do diâmetro logaritmizado, de assimétrica positiva tornou-se bimodal, demonstrando que a transformação não foi capaz de adequar a assimetria (Figura 23 do Apêndice 4).

Acerca da variável  $w_{total}$  na escala original, verificou-se também uma distribuição de frequência com comportamento assimétrico positivo. Destaca-se que, assim como para o diâmetro, os histogramas não foram simetrizados com a transformação logarítmica, mas sim convertidos para uma distribuição bimodal (Figura 24 do Apêndice 5).

#### 5.1.4.2 Biomassa do fuste

A correlação de Spearman demonstrou forte padrão linear entre a biomassa do fuste ( $w_{fuste}$ ) com o diâmetro, assim como também com a altura das árvores. Destaca-se que essas correlações foram superiores a 0,93 (Figura 25 do Apêndice 6).

Sobretudo, enfatiza-se que a dispersão entre essas variáveis marca um comportamento semelhante ao observado para a base de dados anterior, envolvendo a  $w_{total}$ . Nesse sentido, verifica-se uma dispersão exponencial e heterogênea na ocasião em que a  $w_{fuste}$  está associada à altura. Por outro lado, quando a  $w_{fuste}$  é correlacionada com o diâmetro, o comportamento exponencial é menos evidente e a variação da dispersão na nuvem de pontos é substancialmente mais contida (Figura 25 do Apêndice 6).

A análise do histograma de frequência da  $w_{fuste}$  sugere que a maioria das observações estão relacionadas aos menores valores de biomassa. Diante disso, ratifica-se que a variável  $w_{fuste}$  apresenta uma distribuição assimétrica à direita (Figura 25 do Apêndice 6). Entretanto, quando as variáveis estão submetidas à transformação logarítmica, destaca-se que as distribuições de frequência permanecem assimétricas. Apesar disso, a especificação não linear entre as variáveis é corrigida com a logaritimização (Figura 25 do Apêndice 6).



#### 5.1.4.3 Biomassa dos galhos

A partir da análise exploratória, verificou-se que a variável biomassa dos galhos ( $w_{galhos}$ ) demonstrou fortes correlações lineares com o diâmetro e com a altura. Destaca-se, portanto, que a correlação de Spearman demonstrou maiores valores entre a  $w_{galhos}$  e diâmetro ( $\rho = 0,9411$ ) (Figura 26 do Apêndice 7).

Os diagramas de dispersão demonstraram evidente comportamento exponencial entre os pares de dados (Figura 26 do Apêndice 7). Salienta-se que esse comportamento foi mais evidente entre as variáveis  $w_{galhos}$  e altura. Além disso, cabe enfatizar que essas variáveis demonstraram dispersões com evidente presença de heterocedasticidade.

Em contrapartida, o histograma de frequências da variável  $w_{galhos}$  evidenciou um comportamento assimétrico à direita. Esse padrão observado demonstra uma maior densidade de observações para os menores valores  $w_{galhos}$ . Destaca-se, porém, que esse é o comportamento esperado em amostragens de variáveis dendrométricas.

Acerca das análises das variáveis em escala logarítmica, verificou-se que o comportamento exponencial das dispersões passou a apresentar-se nitidamente linear. Enfatiza-se que essa conduta foi observada tanto entre  $w_{galhos}$  e diâmetro quanto entre  $w_{galhos}$  e altura.

Contudo, a logaritimização das variáveis apenas linearizou as dispersões, pois as distribuições de frequências para a  $w_{galhos}$  permaneceram assimétricas. Assim, com a transformação logarítmica na  $w_{galhos}$ , os histogramas apenas transitaram de uma distribuição assimétrica positiva para uma distribuição bimodal (Figura 26 do Apêndice 7).

#### 5.1.4.4 Biomassa das folhas

Acerca da variável biomassa das folhas ( $w_{folhas}$ ), observou-se que o diagrama de dispersão com o diâmetro sugeriu tendências lineares, porém, com variâncias crescentes. A correlação de Spearman entre essas variáveis foi de 0,8518, corroborando o padrão linear existente entre elas. Enquanto isso, quando a  $w_{folhas}$  foi relacionada com a variável altura, as dispersões

demonstraram um evidente padrão não linear, o qual sugeriu um comportamento exponencial. A correlação de Spearman calculada foi de 0,6298.

As frequências de observações da  $w_{folhas}$  mostrou-se semelhante às observadas para os demais componentes de biomassa. Nesse sentido, o histograma de frequências gerado demonstrou um comportamento assimétrico à direita. Destaca-se que essa percepção, além de ser natural para o tipo de amostragem, demonstra que as menores densidades de dados estão associadas aos maiores valores de biomassa (Figura 27 do Apêndice 8).

Em segunda análise, quando as variáveis foram submetidas à transformação logarítmica, houve um aumento de variação na dispersão entre os pares de dados. Observou-se esse comportamento tanto entre a  $w_{folhas}$  e o diâmetro quanto entre a  $w_{folhas}$  e a altura. Apesar disso, as variáveis apresentaram uma disposição sutilmente mais linear.

Ademais, vale enfatizar que a  $w_{folhas}$  logaritmizada demonstrou uma distribuição de frequência mais próxima a normal, contudo, ainda assimétrica. Nesse sentido, verificou-se que a transformação logarítmica na  $w_{folhas}$  promoveu a transição de assimetria à direita para assimetria à esquerda (Figura 27 do Apêndice 8).

#### 5.1.4.5 Biomassa das cascas

Não obstante, quando a biomassa das cascas ( $w_{casca}$ ) foi relacionada ao diâmetro das árvores, os diagramas de dispersão apresentaram uma sutil tendência não linear. Apesar disso, a correlação de Spearman entre essas variáveis foi de 0,9101.

No entanto, esse comportamento não linear foi ainda mais evidente quando a  $w_{casca}$  foi relacionada com a variável altura. Destaca-se que, além de uma dispersão com padrão exponencial, essas variáveis demonstraram uma evidente propensão à heterocedasticidade (Figura 28 do Apêndice 9). Todavia, a correlação entre  $w_{casca}$  e altura foi de 0,9387.

Em continuidade, a análise do histograma de frequência da variável  $w_{casca}$  demonstrou uma distribuição assimétrica à direita, assim como verificado para os demais componente de biomassa. Em contrapartida, quando submetida

a transformação logarítmica, o padrão assimétrico positivo das frequências da  $w_{casca}$  tornou-se bimodal (Figura 28 do Apêndice 9).

Porém, apesar de não demonstrar poder em simetrizar as frequências, a transformação logarítmica linearizou as dispersões da  $w_{casca}$  com as variáveis diâmetro e altura. Além disso, a logaritmização das variáveis homogeneizou a variabilidade entre os pares de dados (Figura 28 do Apêndice 9).

#### 5.1.4.6 Biomassa das raízes

Para essa base de dados, os diagramas de dispersão entre a biomassa das raízes ( $w_{raiz}$ ) e o diâmetro evidenciou tendência crescente e não linear. Ressalta-se que o comportamento esse comportamento observado demonstrou um padrão claramente exponencial.

Contudo, esse desempenho exponencial foi ainda mais explícito no diagrama de dispersão entre  $w_{raiz}$  e altura. Vale enfatizar que a relação entre essas variáveis demonstrou acentuadas variações de dispersão, sugerindo predisposição desses pares de dados à heterocedasticidade. Contudo, apesar das dispersões não lineares, tanto a variável diâmetro quanto a altura demonstraram elevadas correlações de Spearman com a  $w_{raiz}$  (Figura 29 do Apêndice 10).

Enquanto isso, o histograma de frequências da  $w_{raiz}$  demonstrou uma proeminente distribuição assimétrica positiva, de modo que a maioria das observações de  $w_{raiz}$  estiveram relacionadas às menores massas. Entretanto, na escala logarítmica, essa distribuição foi apenas convertida para uma distribuição bimodal, permanecendo assimétrica. Por fim, destaca-se que a transformação logarítmica em  $w_{raiz}$  não normalizou a distribuição de frequências da amostragem. Apesar disso, com as variáveis logaritimizadas obteve-se dispersões linearizadas e com variâncias equalizadas (Figura 29 do Apêndice 10).

## 5.2 ESTRATÉGIAS DE MODELAGEM DA RELAÇÃO ALTURA-DIÂMETRO PARA FLORESTAS PLANTADAS

As modelagens dendrométricas da relação altura-diâmetro em florestas plantadas de *Pinus oocarpa* aos 5 anos, demonstrou que não foi necessário o tratamento das variáveis para atendimento dos pressupostos da análise de regressão. Assim, a homocedasticidade das variâncias, pelo teste de Breusch-Pagan ( $p > 0,05$ ), e a normalidade dos resíduos, pelo teste de Lilliefors ( $p > 0,05$ ), foram contempladas nas modelagens com as variáveis em escala original (Tabela 7).

TABELA 7 - ESTATÍSTICAS DE REGRESSÃO DAS MELHORES ESTRATÉGIAS DE MODELAGEM DA RELAÇÃO ALTURA-DIÂMETRO PARA *Pinus oocarpa* AOS 5 ANOS.

Estratégia	Transformação	Retransformação	R <sup>2</sup> aj	SEE%	BIC	p-BP	p-Li
1	Sem transformação	$y = y$	0,6010	10,17	291,50	0,1904	0,4842
5	Manly	$y = \frac{\log(\lambda \cdot y' + 1)}{\lambda}$	0,6266	9,84	284,89	0,551	0,9916
7	Yeo-Johnson	$y = (\lambda \cdot y' + 1)^{\frac{1}{\lambda}} - 1$	0,6252	9,86	285,25	0,5921	0,9823
4	Box-Cox	$y = (\lambda \cdot y' + 1)^{\frac{1}{\lambda}}$	0,6250	9,86	285,32	0,5991	0,9800

em que R<sup>2</sup>aj é o coeficiente de determinação ajustado do modelo; SEE é o erro padrão da estimativa; BIC é o critério de informação Bayesiano; p-BP é o p-valor do teste de homocedasticidade de Breusch-Pagan; e p-Li é o p-valor do teste de normalidade de Lilliefors.

FONTE: O autor (2022).

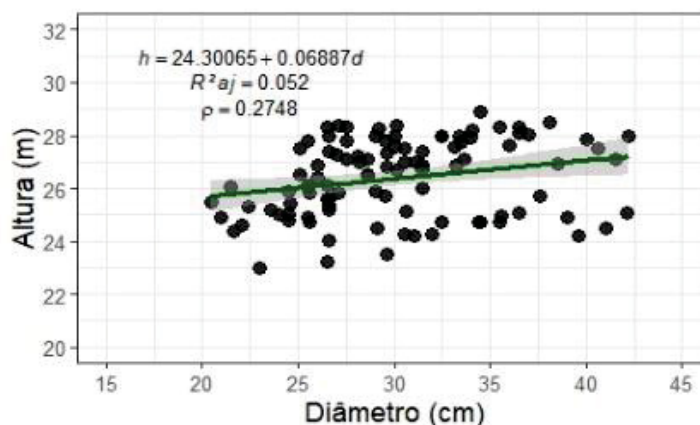
Na comparação entre as modelagens, verificou-se que a incorporação de transformações aos modelos melhorou as estatísticas dos ajustes para predição alométrica. Assim, destaca-se que as transformações potenciais de Manly, seguida de Yeo-Johnson e Box-Cox, melhoraram a performance das modelagens hipsométricas em plantios jovens de *P. oocarpa* (5 anos) (Tabela 7). Todavia, evidencia-se que as aplicações dessas transformações não são estatisticamente exigidas, visto que os pressupostos foram atendidos com as variáveis em escala original.

Em contrapartida, o ajuste da relação altura-diâmetro para povoamentos tardios de *P. oocarpa* (19 anos) demonstrou ausência de correlação entre as variáveis altura e diâmetro. Destaca-se, ainda, que houve baixa relação biológica nesse cenário específico (Figura 1). Dessa forma, a técnica de regressão linear simples não se faz apropriada para predição de altura nesse contexto, haja vista que a relação linear existente entre as variáveis altura e diâmetro apresenta-se

com baixo grau de inclinação, o que afeta consideravelmente a qualidade do ajuste.

Portanto, infere-se que o modelo linear ajustado sob essas condições estima o mesmo valor médio de altura para qualquer diâmetro. Essa ocasião desqualifica esse estrato de idade para abordagens alométricas, visto que nenhum ajuste linear apresentaria resultados satisfatórios nesse cenário.

FIGURA 1 - MODELAGEM DA RELAÇÃO ALTURA-DIÂMETRO PARA *Pinus oocarpa* AOS 19 ANOS.

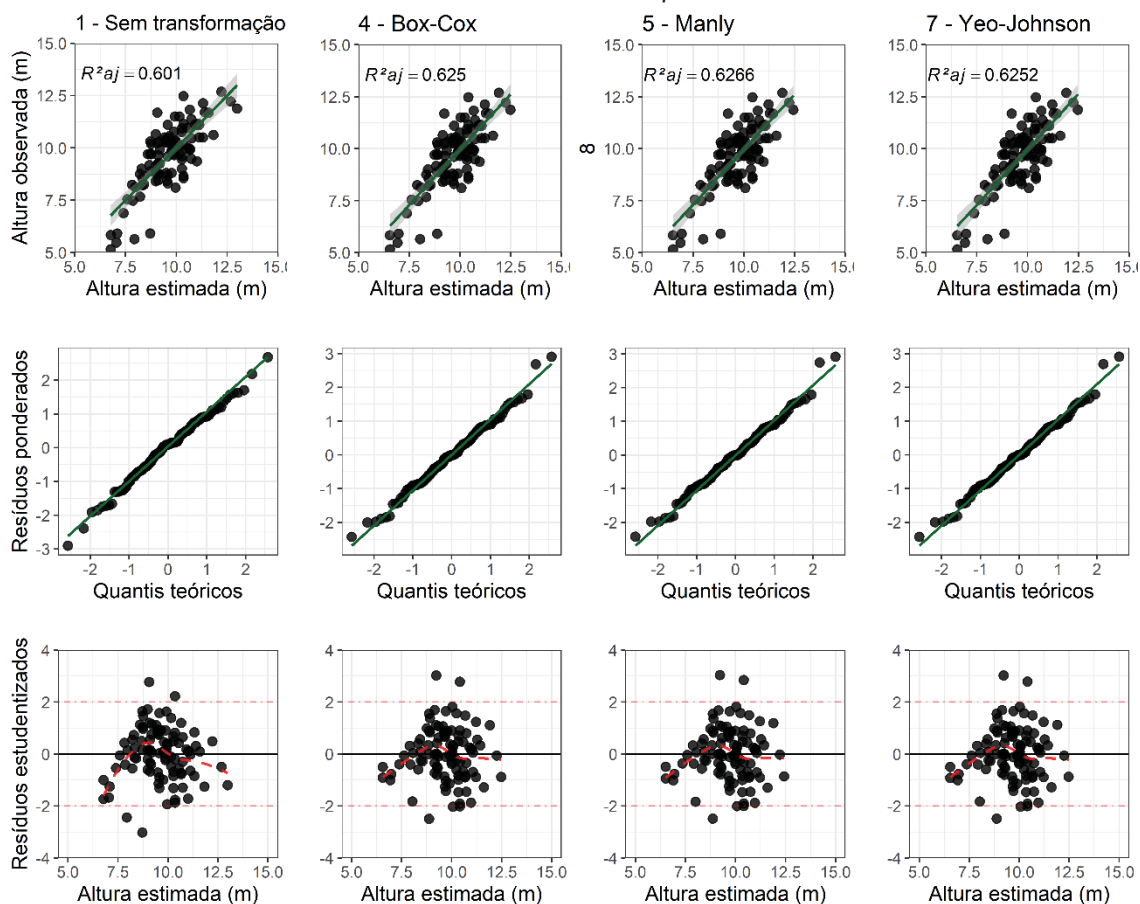


FONTE: O autor (2022).

Por conseguinte, acerca do ajuste da relação altura-diâmetro aos 5 anos, observou-se que a modelagem das estratégias 5, 7 e 4, mesmo promovendo maiores  $R^2_{aj}$  e menores SEE% (Tabela 7), pouco influenciaram na distribuição residual e na disposição dos quantis teóricos (Figura 2). Assim, conforme apresentado nessas análises gráficas, as modelagens com transformações de variáveis apresentaram performances semelhantes a obtida pela modelagem com as variáveis originais.

Vale frisar que o ajuste com as variáveis em escala original, além de atender as premissas da análise de regressão, promoveu coeficientes de regressão significativos. Enquanto isso, os ajustes relacionados às estratégias 7 e 4 (Yeo-Johnson e Box-Cox, respectivamente) foram os únicos a apresentar um de seus coeficientes de regressão não significativos de acordo com o teste  $t$  a 5% de probabilidade. Essas mesmas estratégias apresentaram os maiores erros-padrão (SE) relacionados aos coeficientes de regressão gerados (Tabela 8).

FIGURA 2 - ALTURAS OBSERVADAS VS ESTIMADAS, GRÁFICOS Q-Q E DISTRIBUIÇÃO RESIDUAL DAS MELHORES ESTRATÉGIAS DE MODELAGEM DA RELAÇÃO ALTURA-DIÂMETRO PARA *Pinus oocarpa* AOS 5 ANOS.



FONTE: O autor (2022).

TABELA 8 - COEFICIENTES DE REGRESSÃO E PARÂMETROS DE TRANSFORMAÇÃO DAS MELHORES ESTRATÉGIAS DE MODELAGEM DA RELAÇÃO ALTURA-DIÂMETRO PARA *Pinus oocarpa* AOS 5 ANOS.

Estratégia	Transformação	$\beta_0$	SE $\beta_0$ %	$\beta_1$	SE $\beta_1$ %	$\lambda$
1	Sem transformação	5,0728**	7,5855	0,38823**	81,6011	-
5	Manly	4,65348**	25,9528	1,21424**	8,18948	0,12298
7	Yeo-Johnson	10,8359 <sup>ns</sup>	57,0345	6,2091**	8,19603	2,19692
4	Box-Cox	4,9998 <sup>ns</sup>	82,0693	4,1223**	8,19445	2,06883

em que  $\beta_0$  e  $\beta_1$  são os coeficientes de regressão do modelo; \*\* significativo ao nível 1% de probabilidade; \* significativo ao nível de 5% de probabilidade; <sup>ns</sup> não significativo; SE $\beta_n$ % é o erro padrão percentual dos coeficientes estimados; e  $\lambda$  é o coeficiente de transformação.

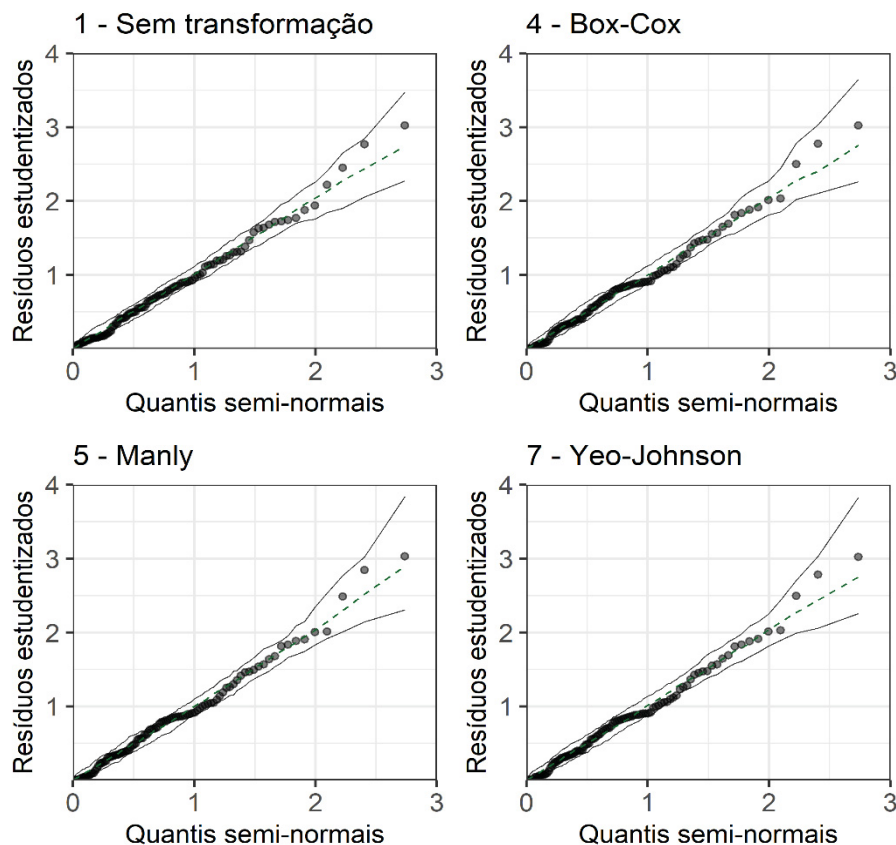
FONTE: O autor (2022).

Nas modelagens da relação altura-diâmetro para *P. oocarpa* aos 5 anos, apenas a transformação inversa (estratégia 3), a inversa-inversa (estratégia 14) e a raiz quadrada deslocada (estratégia 12) não atenderam as premissas de normalidade dos resíduos e homocedasticidade das variâncias ( $p$ -Li e  $p$ -BP <

0,05). Por outro lado, a transformação Log shift (estratégia 11) destacou-se por atender apenas o pressuposto de normalidade residual ( $p-Li = 0,0758$  e  $p-BP = 0,0081$ ).

Acerca das percepções obtidas pelos gráficos *half-normal plot* (hnp) para as modelagens, verificou-se que, na ausência de transformação, os resíduos se mostraram dispostos dentro do envelope de confiança a 5% de probabilidade. Esse resultado corrobora com as interpretações da Tabela 7 sobre a não necessidade de transformar as variáveis para a modelagem da relação altura-diâmetro em povoamento jovem. Acrescenta-se, ainda, que as modelagens com transformação não promoveram melhoras evidentes quanto a qualidade para estimativas (Figura 3).

FIGURA 3 - GRÁFICOS HNP (HALF-NORMAL PLOT) DAS MODELAGENS DA RELAÇÃO ALTURA-DIÂMETRO PARA *Pinus oocarpa* AOS 5 ANOS.



FONTE: O autor (2022).

### 5.3 ESTRATÉGIAS DE MODELAGEM VOLUMÉTRICA PARA FLORESTAS PLANTADAS

Por conseguinte, nas modelagens volumétricas em florestas plantadas, observou-se que a transformação das variáveis foi necessária tanto para os povoamentos jovens (5 anos) quanto para os tardios (19 anos), devido a incompatibilidade dos ajustes lineares simples com os pressupostos clássicos da análise de regressão linear (Tabelas 9 e 11). Vale frisar que as modelagens com variáveis transformadas, além de corrigirem a violação dos pressupostos de normalidade dos erros e homogeneidade das variâncias, melhoraram, sobremaneira, as estatísticas de ajuste e precisão em ambas as idades.

Nesse sentido, considerando especialmente o plantio na idade de 5 anos, observa-se que o ajuste com as variáveis em escala original infringiu os pressupostos de homogeneidade das variâncias ( $p\text{-BP} < 0,01$ ) ao passo que atendeu o de normalidade residual ( $p\text{-Li} > 0,01$ ). Enquanto isso, nos ajustes com variáveis transformadas, essas suposições foram simultaneamente atendidas de acordo com o  $p\text{-valor}$  a 1% de probabilidade dos testes de hipótese (Tabela 9).

TABELA 9 - ESTATÍSTICAS DE REGRESSÃO DAS MELHORES ESTRATÉGIAS DE MODELAGEM VOLUMÉTRICA PARA *Pinus oocarpa* AOS 5 ANOS.

Estratégia	Transformação	Retransformação	R <sup>2</sup> aj	SEE%	BIC	p-BP	p-Li
1	Sem transformação	$y = Y$	0,8779	20,16	-575,75	0,0000	0,0124
4	Box-Cox	$y = (\lambda \cdot Y' + 1)^{\frac{1}{\lambda}}$	0,9383	14,33	-644,02	0,0826	0,9936
6	Bickel-Doksum	$\pm[(Y)(\lambda Y + 1)]^{\frac{1}{\lambda}}$	0,9383	14,33	-644,02	0,0826	0,9936
13	Log-Log	$y = \exp^{Y'}$	0,9379	14,38	-643,37	0,2864	0,7507

em que R<sup>2</sup>aj é o coeficiente de determinação ajustado do modelo; SEE é o erro padrão da estimativa; BIC é o critério de informação Bayesiano; p-BP é o p-valor do teste de homocedasticidade de Breusch-Pagan; e p-Li é o p-valor do teste de normalidade de Lilliefors.

FONTE: O autor (2022).

Cabe destacar que o ajuste sem transformação demonstra uma relevante perda de precisão quando se compara as estatísticas de SEE% com as demais estratégias de ajuste. Salienta-se que a diferença expressa em termos de erro diminui em aproximadamente 6% quando se usa alguma das estratégias de transformação descritas na Tabela 9.

Entretanto, observa-se que as três estratégias de transformações promoveram resultados de ajuste e precisão semelhantes. Todavia, a



transformação Log-Log se destaca devido a sua maior facilidade de aplicação e por promover menores valores de BIC, o que denota maior qualidade de ajuste.

O  $R^2_{aj}$ , por sua vez, apresenta um aumento sutil quando se aplica alguma das transformações selecionadas. Destaca-se que, para as transformações Box-Cox e Bickel-Doksum, as estatísticas de ajuste e precisão, assim como os coeficientes de regressão gerados, mostraram resultados idênticos. Isso se deve a estimativa do parâmetro  $\lambda$  de transformação, que apresentou o mesmo valor para ambas as estratégias (Tabela 10). Portanto, essas estratégias recaem sobre as mesmas propriedades transformadoras, resultando inclusive em uma mesma equação alométrica.

TABELA 10 - COEFICIENTES DE REGRESSÃO E PARÂMETROS DE TRANSFORMAÇÃO DAS MELHORES ESTRATÉGIAS DE MODELAGEM VOLUMÉTRICA PARA *Pinus oocarpa* AOS 5 ANOS.

Estratégia	Transformação	$\beta_0$	SE $\beta_0$ %	$\beta_1$	SE $\beta_1$ %	$\lambda$
1	Sem transformação	-0,066426**	7,5278	0,011069**	3,6137	-
4	Box-Cox	-2,325477**	0,7138	0,056616**	2,4728	0,4212
6	Bickel-Doksum	-2,325475**	0,7095	0,056615**	2,2962	0,4212
13	Log-Log	-8,360080**	1,4342	2,235940**	2,1959	-

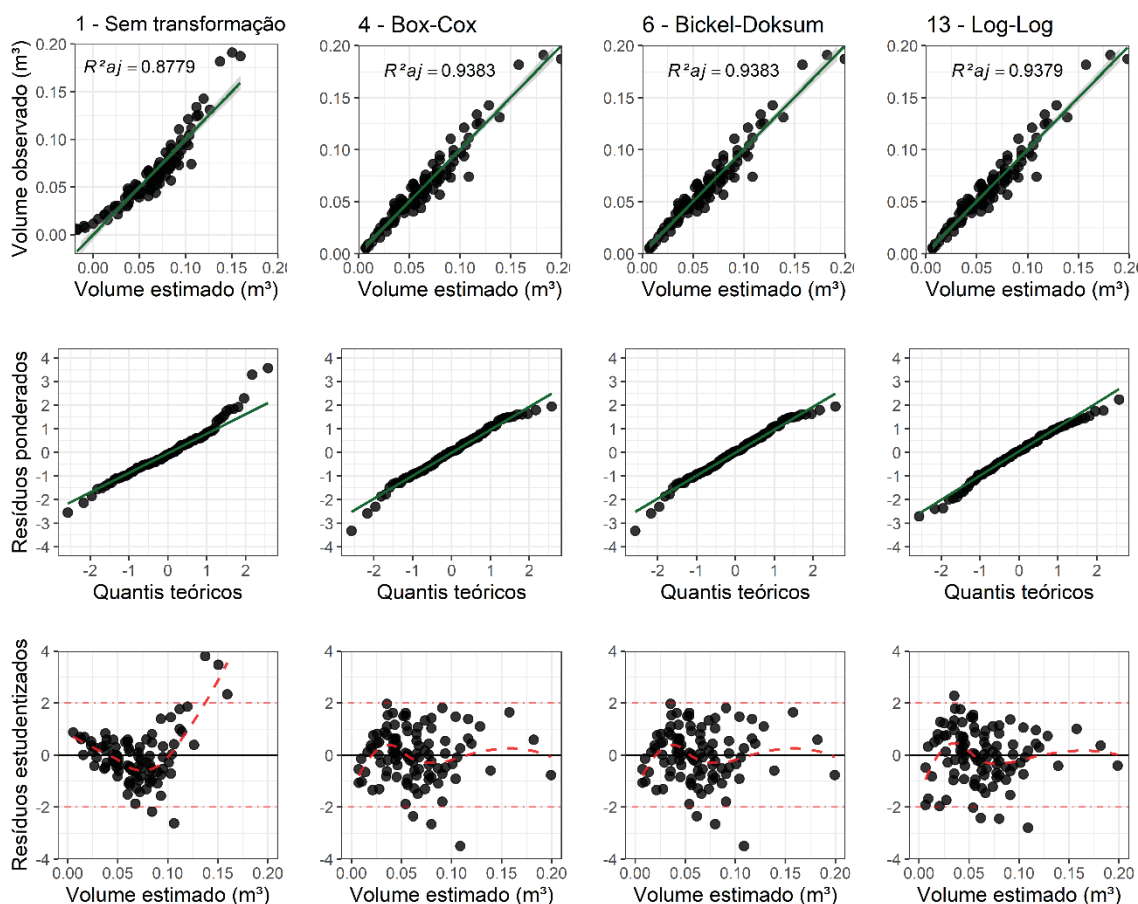
em que  $\beta_0$  e  $\beta_1$  são os coeficientes de regressão do modelo; \*\* significativo ao nível 1% de probabilidade; \* significativo ao nível de 5% de probabilidade; ns não significativo; SE $\beta_n$ % é o erro padrão percentual dos coeficientes estimados; e  $\lambda$  é o coeficiente de transformação.

FONTE: O autor (2022).

Dessa forma, o uso de qualquer uma dessas três técnicas de transformação selecionadas promove a correção simultânea da não normalidade residual e da heterocedasticidade em modelagens volumétricas de povoamentos jovens de *P. oocarpa*. Nesse sentido, como esperado, observa-se uma acentuada tendência não linear para a modelagem com as variáveis originais, quando se confrontam os volumes observados e estimados (Figura 4).

Em virtude disso, a distribuição residual desse ajuste promoveu predisposição à superestimativa para os menores e maiores valores de volume. Esse comportamento promove nítidas tendências à heterocedasticidade. Por outro lado, acerca da normalidade residual, verifica-se uma clara divergência nos quantis superiores do gráfico Q-Q, apontando uma fuga das caudas não esperada para uma distribuição gaussiana dos resíduos.

FIGURA 4 - VOLUMES OBSERVADOS VS ESTIMADOS, GRÁFICOS Q-Q E DISTRIBUIÇÃO RESIDUAL DAS MELHORES ESTRATÉGIAS DE MODELAGEM VOLUMÉTRICA PARA *Pinus oocarpa* AOS 5 ANOS.



FONTE: O autor (2022).

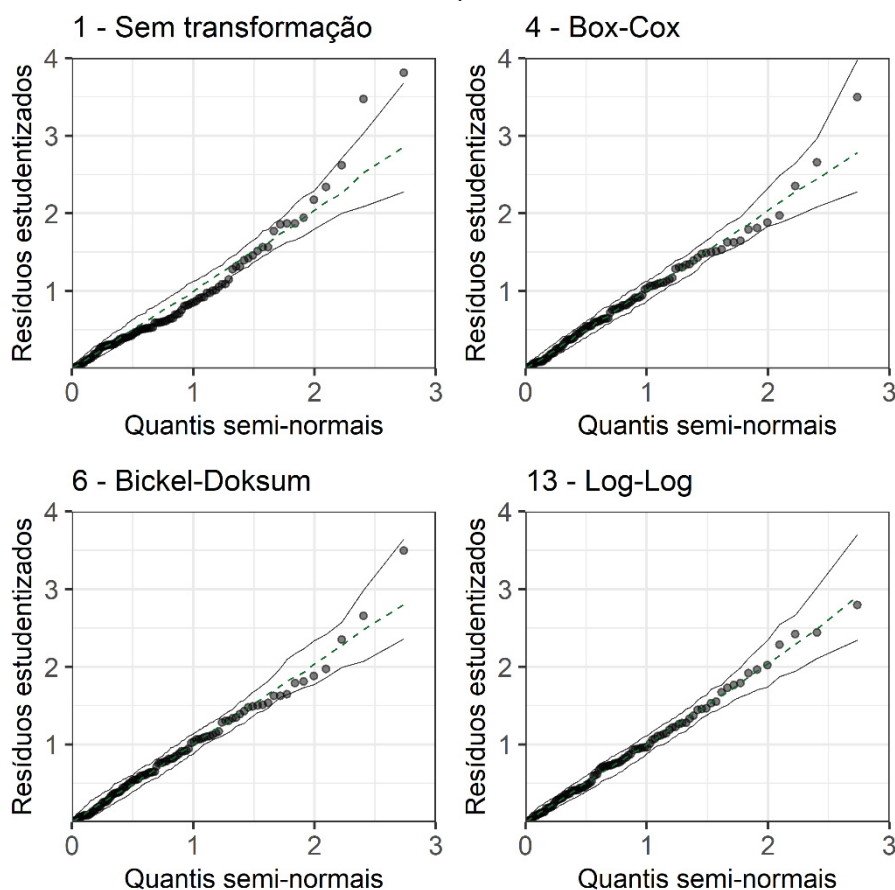
A tendência linear na dispersão entre os volumes observados *versus* estimado foi alcançada para as três estratégias selecionadas (4, 6 e 13). Sobre essa perspectiva, as transformações surtiram efeito direto na estrutura residual das modelagens, de modo que a distribuição dos resíduos estudentizados dispõem-se aleatoriamente com variações constantes e livre de tendenciosidades. Ressalta-se que esse comportamento é exatamente o oposto ao observado na dispersão residual da estratégia 1 (Sem transformação) (Figura 3).

Da mesma maneira, os gráficos Q-Q gerados pelos ajustes com transformações forneceram uma distribuição dos quantis teóricos mais linearizada, demonstrando adequação da distribuição normal dos resíduos para essas estratégias de ajuste. Pelas análises gráficas, infere-se que a estratégia Log-Log (13) foi a que melhor linearizou a distribuição teórica dos quantis,

promovendo menores fugas das caudas na análise dos gráficos Q-Q, principalmente nos quantis inferiores (Figura 4).

Cabe enfatizar que, pela análise dos gráficos hnp, a ausência de transformação resultou em resíduos dispostos além do limite superior do envelope de confiança a 5% de probabilidade (Figura 5), corroborando a inadequação do modelo, conforme Tabela 9. Os métodos de transformação Box-Cox e Bickel-Doksum proporcionam resíduos dentro dos envelopes de confiança, porém próximos aos seus limites. Por fim, a transformação Log-Log mostrou-se a mais adequada.

FIGURA 5 - GRÁFICOS HNP (HALF-NORMAL PLOT) DAS MODELAGENS VOLUMÉTRICAS PARA *Pinus oocarpa* AOS 5 ANOS.



FONTE: O autor (2022).

Seguindo esse mesmo comportamento, as modelagens volumétricas aplicadas ao povoamento tardio de *P. oocarpa* aos 19 anos também promoveram adequação residual somente com uso de transformações. Nesse cenário, as

estratégias que obtiveram maior destaque foram a Log-Log, a Log e a Dual Power, fundamentadas em propriedades logarítmicas.

Todavia, ao contrário do observado nos ajustes volumétricos aos 5 anos, os modelos com transformações não aperfeiçoaram as estatísticas de ajuste e precisão. Contudo, destaca-se que houve diferenças estatística entre as estratégias de modelagem com transformações. Nesse sentido, salienta-se que a estratégia 13 (Log-Log) forneceu melhores estatísticas de ajuste e precisão, as quais foram muito próximas às apresentadas pela modelagem com as variáveis originais (Tabela 11).

TABELA 11 - ESTATÍSTICAS DE REGRESSÃO DAS MELHORES ESTRATÉGIAS DE MODELAGEM VOLUMÉTRICA PARA *Pinus oocarpa* AOS 19 ANOS.

Estratégia	Transformação	Retransformação	R <sup>2</sup> aj	SEE%	BIC	p-BP	p-Li
1	Sem transformação	$y = y$	0,9399	9,41	-190,82	0	0,0092
13	Log-Log	$y = \exp^{y'}$	0,9354	9,75	-183,72	0,6071	0,0584
2	Log	$y = \exp^{y'}$	0,9002	12,12	-140,21	0,0789	0,0631
9	Dual power	$y = [\sqrt{1 + \lambda^2 y'^2} + \lambda y']^{\frac{1}{\lambda}}$	0,9002	12,12	-140,21	0,0789	0,0631

em que R<sup>2</sup>aj é o coeficiente de determinação ajustado do modelo; SEE é o erro padrão da estimativa, BIC é o critério de informação Bayesiano; p-BP é o p-valor do teste de homocedasticidade de Breusch-Pagan; e p-Li é o p-valor do teste de normalidade de Lilliefors.

FONTE: O autor (2022).

Ademais, evidencia-se que as estratégias Log e Dual power promoveram estatísticas de ajuste e precisão idênticas, devido a estimativa do parâmetro de transformação  $\lambda$  fornecido pela transformação Dual power (estratégia 9). Dessa forma, por  $\lambda$  ser igual a zero, atribuiu-se a propriedade logarítmica na transformação da variável dependente, conforme descrito no tópico 4.2.8.

Diante disso, tanto a estratégia Log quanto a Dual power incidem em propriedades logarítmicas para transformação da variável resposta. Assim, como esperado, todas as estatísticas de ajustes fornecidas por essas transformações mostraram-se idênticas. Ressalta-se, ainda, que os coeficientes de regressão das equações geradas por esses modelos foram significativos à 1%, de acordo com o teste *t* (Tabela 12).

TABELA 12 - COEFICIENTES DE REGRESSÃO E PARÂMETROS DE TRANSFORMAÇÃO DAS MELHORES ESTRATÉGIAS DE MODELAGEM VOLUMÉTRICA PARA *Pinus oocarpa* AOS 19 ANOS.

Estratégia	Transformação	$\beta_0$	SE $\beta_0$ %	$\beta_1$	SE $\beta_1$ %	$\lambda$
1	Sem transformação	-1,082671**	4,7937	0,066834**	2,543616	-
13	Log-Log	-7,56433**	2,36505	2,18833**	2,403659	-
2	Log	-2,285393**	2,63412	0,071090**	2,672668	-
9	Dual power	-2,285393**	2,63412	0,071090**	2,672668	0,0000

em que  $\beta_0$  e  $\beta_1$  são os coeficientes de regressão do modelo; \*\* significativo ao nível 1% de probabilidade; \* significativo ao nível de 5% de probabilidade; ns não significativo; SE $\beta_n$ % é o erro padrão percentual dos coeficientes estimados; e  $\lambda$  é o coeficiente de transformação.

FONTE: O autor (2022).

Nessa conjectura, a estratégia 13, que se baseia no uso da transformação logarítmica tanto na variável dependente quanto na independente, configura-se como a mais adequada para modelagem volumétrica do povoamento tardio de *P. oocarpa*. Isso se deve à performance obtida nos ajustes e na adequação dos pressupostos de normalidade residual e homogeneidade das variâncias (Tabela 11). A superioridade dessa estratégia é comprovada quando se comparam as dispersões residuais geradas por cada uma das estratégias de ajuste selecionadas (Figura 6).

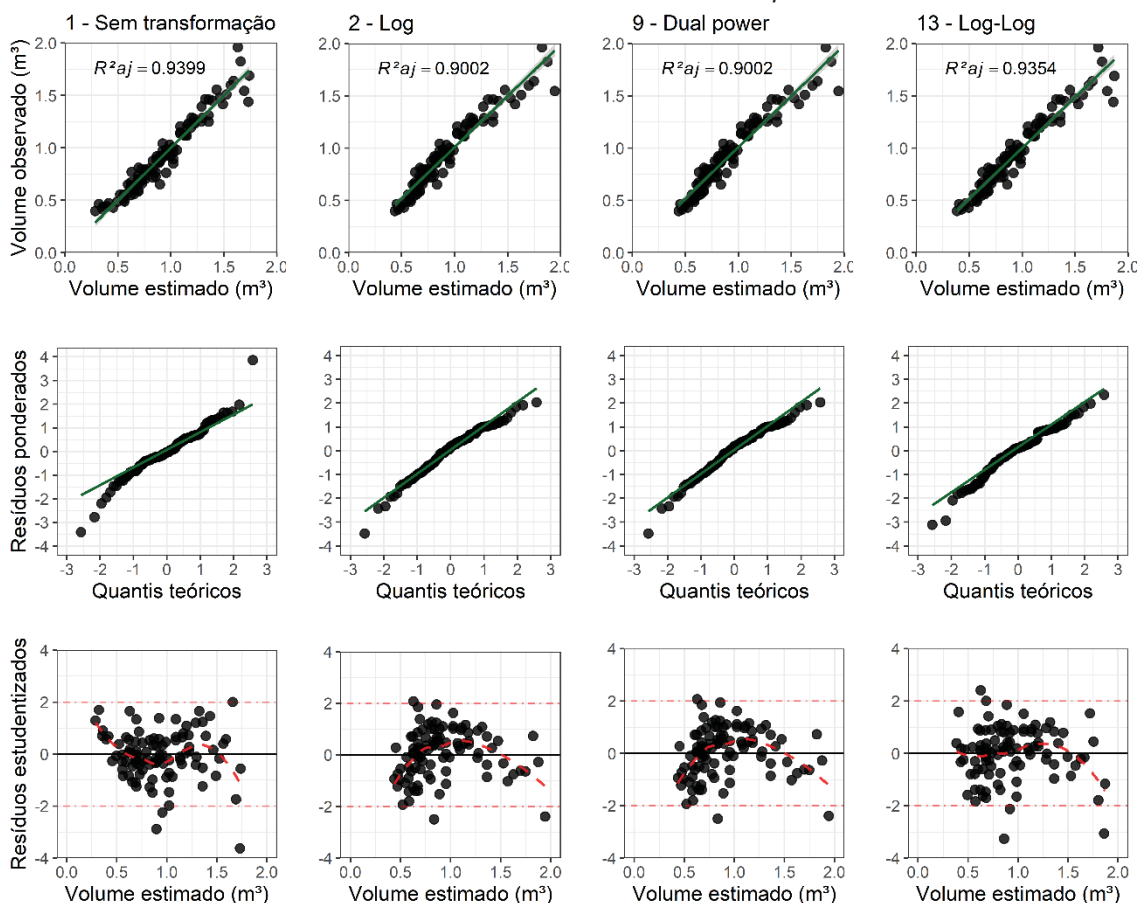
Com base nos gráficos de dispersão residual, sugere-se que as variâncias mais homogêneas dos erros são, indiscutivelmente, associadas à estratégia Log-Log. Enquanto isso, na modelagem com as variáveis originais (estratégia 1), as abordagens gráficas evidenciaram claras violações de homogeneidade nas variâncias e normalidade residual (Figura 6).

Assim, a distribuição não gaussiana dos resíduos gerado por essa estratégia é manifestamente comprovada nas distribuições dos quantis teóricos (gráficos Q-Q). Dessa forma, observa-se divergências proeminentes nos quantis inferiores e superiores gerados no ajuste. Esse comportamento dos quantis normais indicam uma distribuição enviesada dos erros em formato “S” (Figura 6).

A distribuição dos resíduos, por sua vez, também salienta a qualidade inferior desse ajuste, de modo que a dispersão dos erros gerados é inconsistente e tendenciosa ao longo do eixo das abscissas, exibindo variâncias crescentes e heterogêneas em clássicos formatos de “funil” ou “leque”. Essas implicações, com base na distribuição dos resíduos e dos quantis teórico, confluem com as

estatísticas obtidas pelos testes de aderência de Breusch-Pagan, para homogeneidade das variâncias, e de Lilliefors, para normalidade dos resíduos.

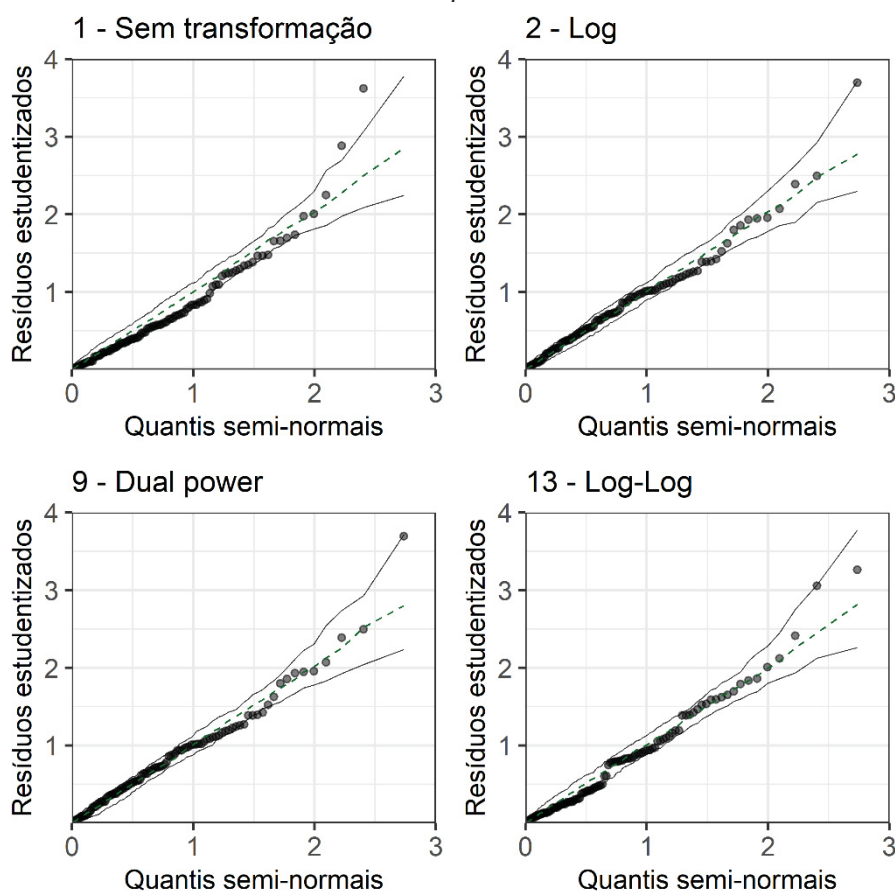
FIGURA 6 - VOLUMES OBSERVADOS VS ESTIMADOS, GRÁFICOS Q-Q E DISTRIBUIÇÃO RESIDUAL DAS MELHORES ESTRATÉGIAS DE MODELAGEM VOLUMÉTRICA PARA *Pinus oocarpa* AOS 19 ANOS.



FONTE: O autor (2022).

Pela análise dos gráficos hnp, na Figura 7, observou-se que a ausência de transformação resultou em resíduos dispostos além do limite inferior do envelope de confiança a 5% de probabilidade. Essa percepção corroborou a inadequação imposta ao modelo, conforme Tabela 11. Os métodos de transformação Log, Dual power e Log-Log proporcionam resíduos dentro dos envelopes de confiança, porém próximos aos seus limites. Contudo, forneceram estimativas com qualidade superior à modelagem sem transformação (estratégia 1).

FIGURA 7 - GRÁFICOS HNP (HALF-NORMAL PLOT) DAS MODELAGENS VOLUMÉTRICAS PARA *Pinus oocarpa* AOS 19 ANOS.



FONTE: O autor (2022).

#### 5.4 ESTRATÉGIAS DE MODELAGEM VOLUMÉTRICA PARA FLORESTAS NATURAIS

Devido à alta heterogeneidade entre os pares de dados das florestas naturais avaliadas, o ajuste volumétrico com as variáveis em escala original (estratégia 1) mostrou-se inconsistente quanto ao atendimento dos pressupostos paramétricos da análise de regressão. Destaca-se, ainda, que somente nas modelagens volumétricas para a Mata Atlântica foi possível alcançar a correção das premissas da regressão linear. Logo, nenhuma das estratégias de transformação estudadas forneceu resultados satisfatórios quando ajustada à base de dados de árvores amazônicas para modelagem do volume comercial.

Assim, para as modelagens do volume do fuste na Mata Atlântica, verificou-se que, apesar de algumas estratégias de transformação mostrarem-se próximas da correção das ausências de normalidade residual e

homogeneidade das variâncias, apenas a transformação Log-Log (estratégia 13) foi capaz de corrigi-las simultaneamente. Ainda, enfatiza-se que essa estratégia, além de promover o atendimento das suposições da análise de regressão linear, acarretou melhoras significativas às estatísticas de ajuste ( $R^2_{aj}$  e BIC) e precisão (SEE). Destaca-se que essas qualidades atribuídas à transformação Log-Log estão relacionadas a sua capacidade de linearização entre volume e diâmetro.

Enquanto isso, a estratégia de transformação Box-Cox (4), a segunda mais promissora, adequou apenas a distribuição residual à normalidade, deixando ainda prevalecer a heterocedasticidade, de acordo com os testes de Lilliefors e Breusch-Pagan, respectivamente. (Tabela 13).

TABELA 13 - ESTATÍSTICAS DE REGRESSÃO DAS MELHORES ESTRATÉGIAS DE MODELAGEM VOLUMÉTRICA DE UMA AMOSTRAGEM DE ÁRVORES NATURAIS DA MATA ATLÂNTICA.

Estratégia	Transformação	Retransformação	$R^2_{aj}$	SEE%	BIC	p-BP	p-Li
1	Sem transformação	$y = y$	0,8195	42,09	-368,86	0,0000	0,0000
13	Log-Log	$y = \exp^{y'}$	0,9011	36,08	-548,57	0,1898	0,9317
4	Box-Cox	$y = (\lambda \cdot y' + 1)^{\frac{1}{\lambda}}$	0,8950	40,28	-420,06	0,1944	0,0000

em que  $R^2_{aj}$  é o coeficiente de determinação ajustado do modelo; SEE é o erro padrão da estimativa; BIC é o critério de informação Bayesiano; p-BP é o p-valor do teste de homocedasticidade de Breusch-Pagan; e p-Li é o p-valor do teste de normalidade de Lilliefors. FONTE: O autor (2022).

Portanto, diante dos resultados da Tabela 13, observa-se uma inquestionável superioridade estatística da modelagem com o uso da estratégia Log-Log (13). Enfatiza-se, ainda, que apenas essa estratégia foi capaz de promover uma distribuição constante das variâncias, adequando conjuntamente os pressupostos de homocedasticidade e normalidade residual.

Em consonância a isto, verificou-se que, mesmo infringindo as suposições da regressão linear, a modelagem volumétrica com as variáveis originais apresentou coeficientes de regressão significativos de acordo com o teste  $t$  ao nível de 5%. Salienta-se, porém, que essa inferência é passível de descrédito estatístico, visto que os resíduos gerados não se adequaram à análise paramétrica (Tabela 13).



TABELA 14 - COEFICIENTES DE REGRESSÃO E PARÂMETROS DE TRANSFORMAÇÃO DAS MELHORES ESTRATÉGIAS DE MODELAGEM VOLUMÉTRICA DE UMA AMOSTRAGEM DE ÁRVORES NATURAIS DA MATA ATLÂNTICA.

Estratégia	Transformação	$\beta_0$	SE $\beta_0$ %	$\beta_1$	SE $\beta_1$ %	$\lambda$
1	Sem transformação	-0,583920**	3,5416	0,039814**	1,9339	-
13	Log-Log	-8,375460**	-1,1779	2,261330**	1,3731	-
4	Box-Cox	-2,455771**	-0,9329	0,060425**	1,4067	0,3205

em que  $\beta_0$  e  $\beta_1$  são os coeficientes de regressão do modelo; \*\* significativo ao nível 1% de probabilidade; \* significativo ao nível de 5% de probabilidade; <sup>ns</sup> não significativo; SE $\beta_n$ % é o erro padrão percentual dos coeficientes estimados; e  $\lambda$  é o coeficiente de transformação.

FONTE: O autor (2022).

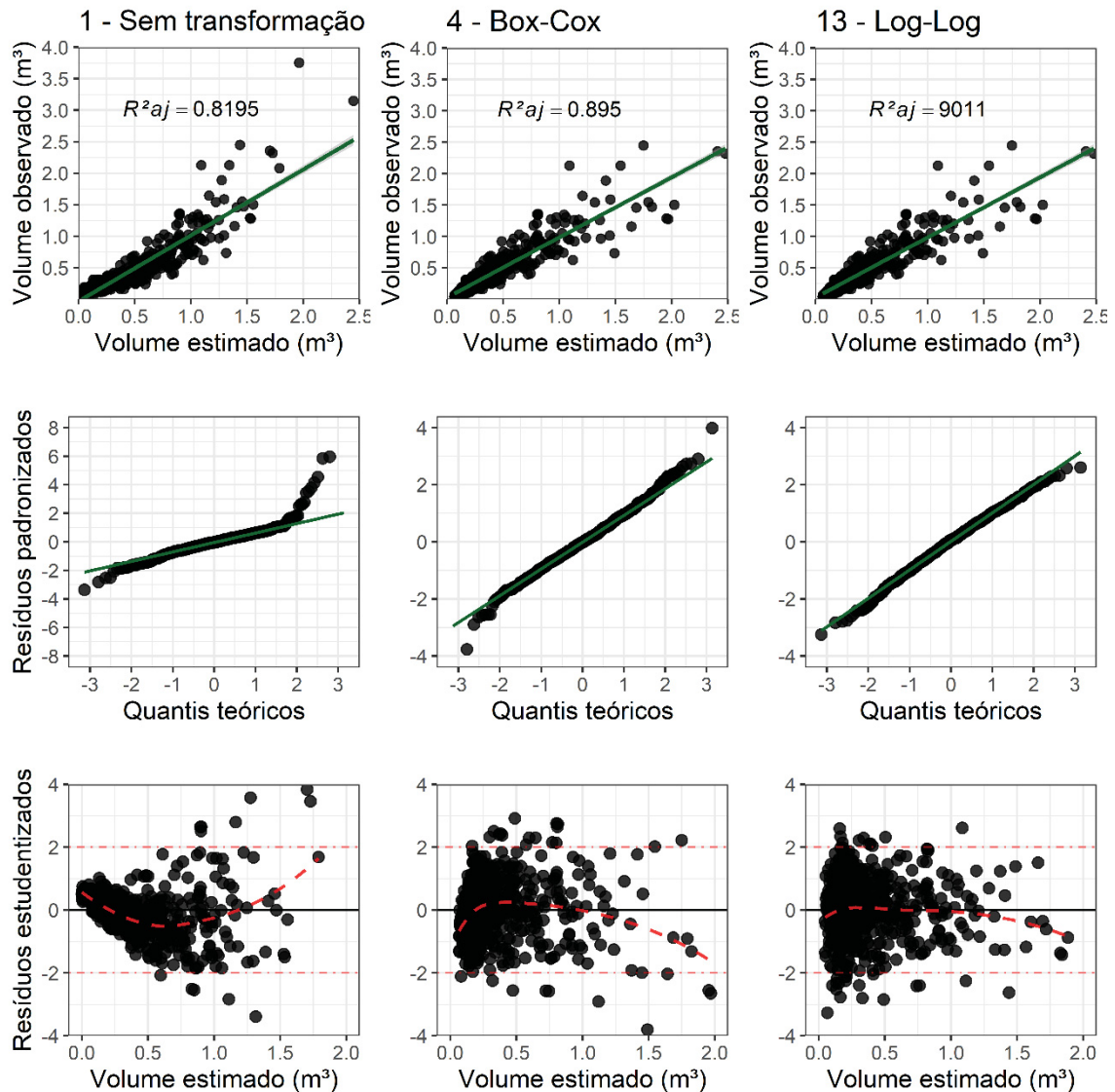
Por outro lado, como esperado, as modelagens com transformações de variáveis também promoveram coeficientes de regressão significativos. Assim, acerca desses resultados os coeficientes  $\beta_0$  e  $\beta_1$  ajustados são estatisticamente habilitados para predições e apropriados para o uso em equações alométricas.

Destaca-se que os testes de hipóteses à normalidade dos resíduos e homocedasticidade indicaram que apenas a transformação Log-Log corrigiu simultaneamente essas suposições. Apesar disso, verifica-se pelas análises dos gráficos Q-Q e distribuição dos resíduos estudentizados melhoras proeminentes nas modelagens quando se aplica transformações nas variáveis durante o processo de ajuste (Figura 8).

Observa-se, ainda, que a distribuição dos erros na modelagem com a transformação Box-Cox melhorou acentuadamente a distribuição dos erros gerados. Destaca-se que o comportamento residual da estratégia Box-Cox se equiparou ao obtido com o emprego do método de transformação Log-Log. Enfatiza-se que essa similaridade é corroborada pelo teste de hipótese de Breusch-Pagan a 1%, o qual demonstra que ambas as estratégias apresentaram variâncias constantes dos resíduos (Tabela 13).

Em vista disso, enfatiza-se que a disposição dos resíduos da regressão linear simples sem transformação demonstrou dispersão altamente enviesada, com superestimativas progressivas do volume. Enquanto isso, o gráfico Q-Q demonstrou, a partir das distorções dos quantis normais, uma evidente fuga da normalidade residual. Essas inferências são corroboradas pelos testes estatísticos de hipóteses ( $p$ -valor < 0,05).

FIGURA 8 - VOLUMES OBSERVADOS VS ESTIMADOS, GRÁFICOS Q-Q E DISTRIBUIÇÃO RESIDUAL DAS MELHORES ESTRATÉGIAS DE MODELAGEM VOLUMÉTRICA DE UMA AMOSTRAGEM DE ÁRVORES NATURAIS DA MATA ATLÂNTICA.

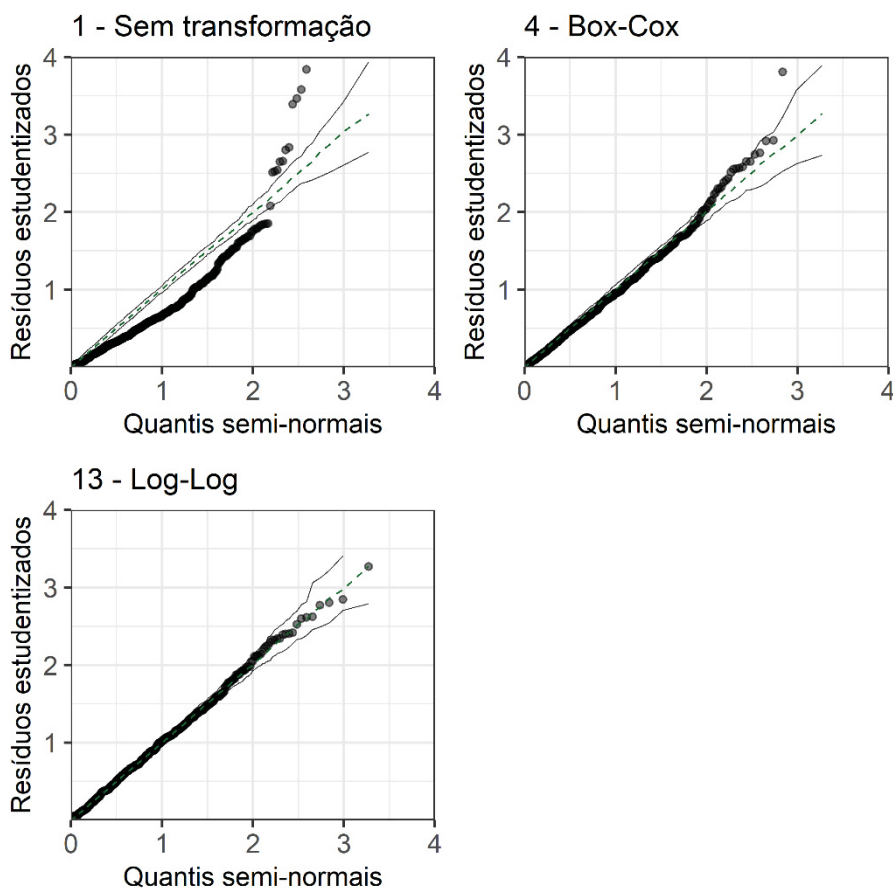


FONTE: O autor (2022).

Portanto, verifica-se que a logaritmização se mostra eficiente à adequação das suposições da regressão linear. Apesar disso, as propriedades logarítmicas de transformação presentes nas estratégias 2, 8 e 11 (Log, Glog e Log shift, respectivamente) não forneceram os resultados esperados. Assim, sugere-se que a propriedade logarítmica de transformação deve ser implementada tanto na variável dependente quanto na independente, visto que quando direcionadas apenas à variável  $y$  do modelo não promoveram a adequação simultânea das suposições paramétricas.

Acerca das análises dos gráficos hnp na Figura 9, destaca-se que a ausência de transformação (1) e o método Box-Cox (4) resultaram em resíduos dispostos além dos limites do envelope de confiança a 5% de probabilidade, corroborando a inadequação do modelo, conforme Tabela 13. O método de transformação Log-Log (13) proporcionou resíduos dentro dos envelopes de confiança. Assim, de acordo com as análises dos gráficos hnp, a transformação Log-Log (13) foi unânime entre as estratégias, exibindo exímia qualidade preditiva para o volume do fuste de árvores da Mata Atlântica.

FIGURA 9 - GRÁFICOS HNP (HALF-NORMAL PLOT) DAS MODELAGENS VOLUMÉTRICAS PARA A UMA AMOSTRAGEM DE ÁRVORES DA MATA ATLÂNTICA.



FONTE: O autor (2022).

Em via de regra, infere-se, pelas análises gráficas, que as transformações Log-Log e Box-Cox promoveram a linearização requerida em análise de regressão. Todavia, especifica-se que o método de Box-Cox, apesar de se configurar uma alternativa convincente para modelagem volumétrica, não

se impõe integralmente como mediadora das pressuposições da análise de regressão linear, devido a habilidade em corrigir apenas a normalidade dos resíduos (gráfico Q-Q da Figura 8). Por outro lado, a alternativa Log-Log de transformação mostra-se mais eficiente ao atendimento simultâneo das premissas exigidas.

Dessa forma, as estratégias de transformação para a modelagem do volume do fuste na Mata Atlântica surtiram os efeitos almejados. Entretanto, quando aplicadas para a modelagem volumétrica em florestas de produção na Amazônia, os resultados foram menos promissores.

Assim, as metodologias de transformação propostas não alcançaram os resultados esperados, mostrando-se técnicas incapazes de corrigir as pronunciadas fugas da normalidade residual e heterogeneidade das variâncias, típicas à tipologia florestal amazônica. Apesar disso, as estratégias Log-Log e Box-Cox, mesmo não atendendo as suposições da regressão linear, promoveram os melhores ajustes dentre as estratégias envolvendo transformação (Tabela 15).

TABELA 15 - ESTATÍSTICAS DE REGRESSÃO DAS MELHORES ESTRATÉGIAS DE MODELAGEM VOLUMÉTRICA DE UMA AMOSTRA DE ÁRVORES NATURAIS DA FLORESTA AMAZÔNICA.

Estratégia	Transformação	Retransformação	R <sup>2</sup> aj	SEE%	BIC	p-BP	p-Li
1	Sem transformação	$y = y$	0,6304	41,86	27.095,40	2,22e-16	2,2e-16
13	Log-Log	$y = \exp^{y'}$	0,6402	42,75	27.315,54	4,67e-14	0,0005
4	Box-Cox	$y = (\lambda \cdot y' + 1)^{\frac{1}{\lambda}}$	0,6174	51,02	29.154,29	2,22e-16	0,0016

em que R<sup>2</sup>aj é o coeficiente de determinação ajustado do modelo; SEE é o erro padrão da estimativa; BIC é o critério de informação Bayesiano; p-BP é o p-valor do teste de homocedasticidade de Breusch-Pagan; e p-Li é o p-valor do teste de normalidade de Lilliefors.

FONTE: O autor (2022).

À vista disso, a modelagem da estratégia 4 (Box-Cox), além de não atender os pressupostos requeridos, não se mostrou estatisticamente superior aos ajustes das estratégias 1 e 13 (Sem transformação e Log-Log, respectivamente). Essas, por sua vez, promoveram R<sup>2</sup>aj, BIC e SEE% semelhantes entre si. Apesar disso, a aplicação da estratégia Log-Log na modelagem ainda se mostrou inapropriada, visto que os testes de hipóteses não indicaram atendimento da normalidade residual ( $p-Li < 0,05$ ) e da homogeneidade das variâncias ( $p-BP < 0,05$ ).

A semelhança entre as modelagens das estratégias selecionadas se estendeu também aos erros padrões relacionados aos coeficientes de regressão. Destaca-se que todos os ajustes forneceram coeficientes de regressão significativos de acordo com o teste  $t$  a 1%.

Porém, cabe enfatizar que essas inferências estatísticas, acerca da significância do teste  $t$ , tornam-se duvidosas quando não há o atendimento da normalidade residual e da homogeneidade das variâncias nas modelagens ( $p$ -Li e  $p$ -BP na Tabela 15). Nessa perspectiva, eventuais estimativas promovidas pelas equações ajustadas poderão ser inapropriadas.

TABELA 16 - COEFICIENTES DE REGRESSÃO E PARÂMETROS DE TRANSFORMAÇÃO DAS MELHORES ESTRATÉGIAS DE MODELAGEM VOLUMÉTRICA DE UMA AMOSTRA DE ÁRVORES NATURAIS DA FLORESTA AMAZÔNICA.

Estratégia	Transformação	$\beta_0$	SE $\beta_0$ %	$\beta_1$	SE $\beta_1$ %	$\lambda$
1	Sem transformação	-8,314319**	2,12669	0,198477**	1,06158	-
13	Log-Log	-6,629850**	1,33442	1,946860**	1,03911	-
4	Box-Cox	0,121692**	16,33222	0,021701**	1,09212	0,0068

em que  $\beta_0$  e  $\beta_1$  são os coeficientes de regressão do modelo; \*\* significativo ao nível 1% de probabilidade; \* significativo ao nível de 5% de probabilidade; ns não significativo; SE $\beta_n$ % é o erro padrão percentual dos coeficientes estimados; e  $\lambda$  é o coeficiente de transformação.

FONTE: O autor (2022).

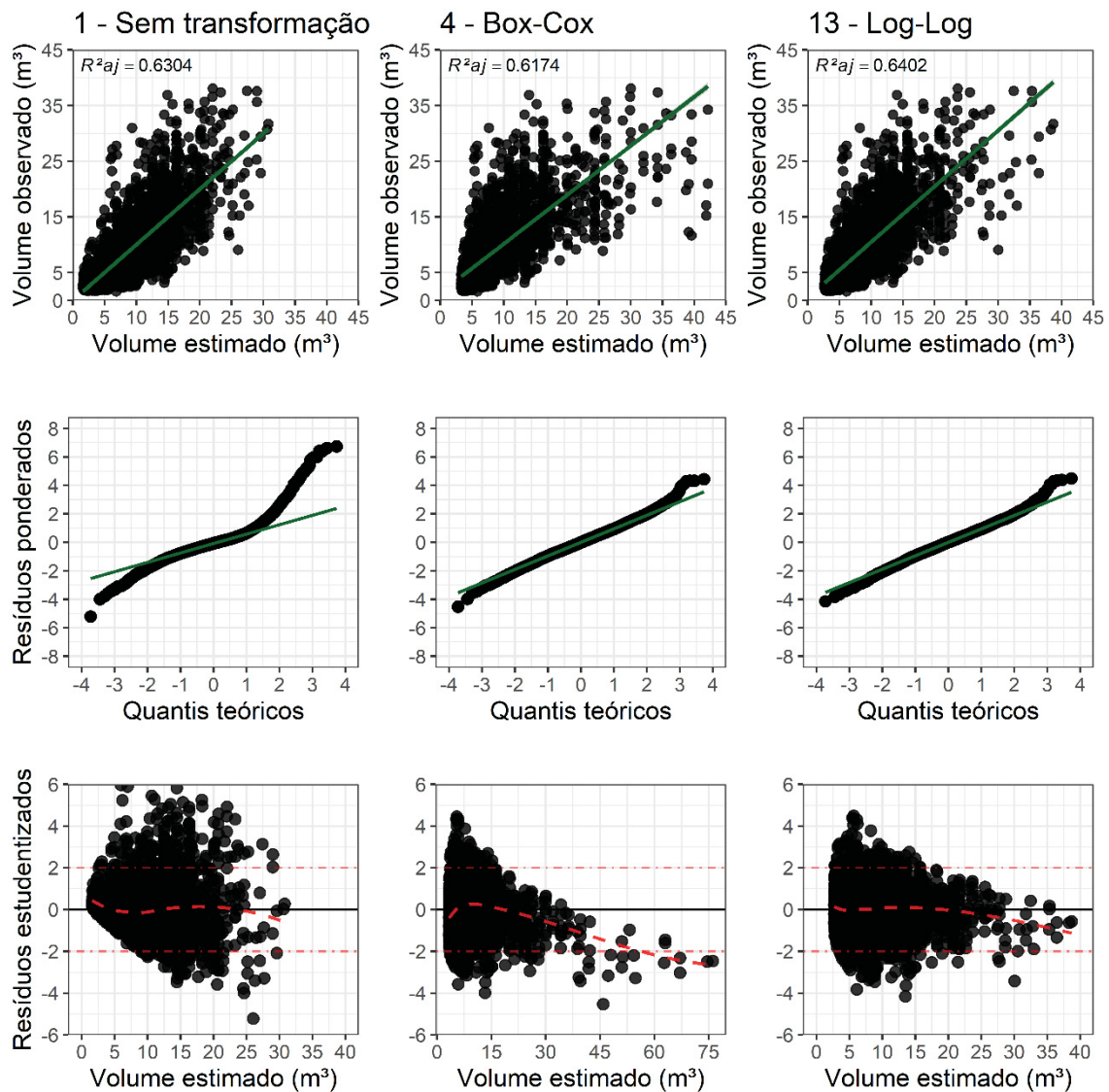
Com base na comparação gráfica entre as estratégias de ajuste 1, 13 e 4, observa-se distribuições não constantes dos resíduos e evidencia de presença de viés nas estimativas (Figura 10). Contudo, verificou-se que as melhoras alométricas obtidas com o emprego de transformação na modelagem foram insignificantes.

Assim, depreende-se que o emprego de transformações na modelagem do volume mostrou-se uma técnica inapropriada. Todavia, a estrutura residual fornecida pelo ajuste da estratégia Log-Log (13) apresentou menores amplitudes e enviesamentos em comparação com as estratégias de Box-Cox (4) e de variáveis originais (1); porém, destaca-se que o enviesamento residual ainda permaneceu evidente. Infere-se, portanto, que a heterocedasticidade residual torna-se mais explícita quando se aplica a estratégia de transformação Box-Cox na modelagem.

Em contrapartida, os gráficos Q-Q demonstram que as estratégias Log-Log e Box-Cox promoveram distribuições residuais com menores distorções,

indicando maior proximidade dessas estratégias à distribuição gaussiana dos resíduos (Figura 10). Nesse sentido, as análises gráficas corroboram os resultados dos testes de Breusch-Pagan para homogeneidade das variâncias e de Lilliefors para normalidade residual.

FIGURA 10 - VOLUMES OBSERVADOS VS ESTIMADOS, GRÁFICOS Q-Q E DISTRIBUIÇÃO RESIDUAL DAS MELHORES ESTRATÉGIAS DE MODELAGEM VOLUMÉTRICA DE UMA AMOSTRA DE ÁRVORES NATURAIS DA FLORESTA AMAZÔNICA.

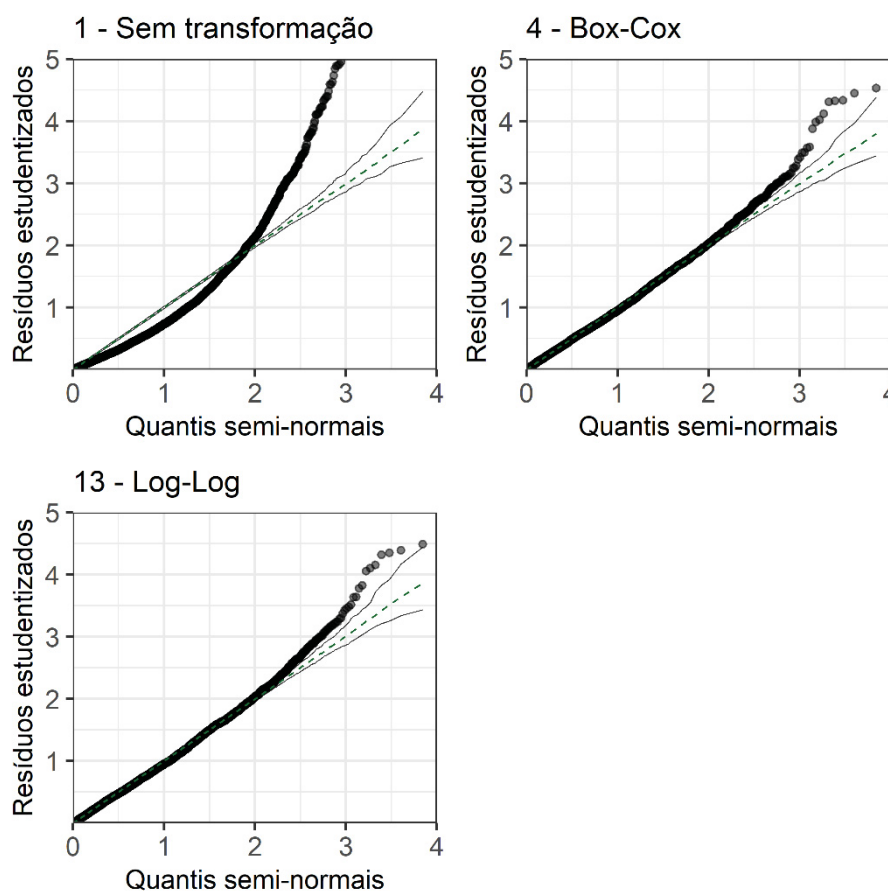


FONTE: O autor (2022).

Assim, nota-se que o ajuste com estratégias de transformação promoveu melhorias às estruturas residuais do modelo. Apesar disso, nenhuma delas foi promissora para o atendimento simultâneo dos pressupostos da regressão linear em modelagem volumétrica.

Diante disso, os gráficos hnp corroboram essas perspectivas (Figura 11), visto que as modelagens com ausência de transformação e com as transformações de Box-Cox e Log-Log resultaram em resíduos além dos limites de confiança a 5% de probabilidade. Porém, a implementação das transformações proporcionou modelagens de volume com melhor qualidade preditiva.

FIGURA 11 - GRÁFICOS HNP (HALF-NORMAL PLOT) DAS MODELAGENS VOLUMÉTRICAS DE UMA AMOSTRA DE ÁRVORES NATURAIS DA FLORESTA AMAZÔNICA.



FONTE: O autor (2022).

## 5.5 ESTRATÉGIAS DE MODELAGEM DA BIOMASSA TOTAL DE ÁRVORES DE *Mimosa scrabella*

Acerca da alometria da biomassa total ( $w_{total}$ ) de árvores de *Mimosa scrabella*, observou-se que a normalidade residual e a homogeneidade das variâncias foram atendidas à 1% de significância quando se usou a altura como

variável explicativa (Tabela 17). Contudo, na ocasião em que o diâmetro foi a variável explicativa do modelo, apenas a normalidade residual foi contemplada ( $p\text{-Li} > 0,05$ ), o que demonstrou, portanto, a possibilidade da aplicação de transformações.

Evidencia-se que, em todos os cenários abordados, as transformações de base logarítmica e raiz quadrada se destacaram como as estratégias de transformações mais promissoras. Por outro lado, as estratégias 3 e 14 (inversa e inversa-inversa), fundamentadas sobre a propriedade recíproca de transformação, apresentaram os piores ajustes.

TABELA 17 - ESTATÍSTICAS DE REGRESSÃO DAS MELHORES ESTRATÉGIAS DE MODELAGEM DA BIOMASSA TOTAL DE *Mimosa scrabella* EM FUNÇÃO DA ALTURA E DO DIÂMETRO.

Modelo	Estratégia	Transformação	Retransformação	R <sup>2</sup> aj	SEE%	BIC	p-Li	p-BP
w <sub>total</sub> (kg) em função da altura	1	Sem transformação	$y = y$	0,7059	60,23	328,52	0,0191	0,0102
	12	Raiz deslocada	$y = (y'^2 - \lambda)$	0,7206	58,70	326,97	0,3572	0,0715
	13	Log-Log	$y = \exp^{y'}$	0,7975	63,28	331,48	0,5711	0,6708
	15	Raiz-Raiz	$y = y'^2$	0,8283	58,28	326,54	0,4460	0,1913
w <sub>tota</sub> (kg) em função do diâmetro	1	Sem transformação	$y = y$	0,9519	24,35	274,18	0,2494	0,0070
	5	Manly	$y = \frac{\log(\lambda \cdot y' + 1)}{\lambda}$	0,9689	19,59	261,13	0,1875	0,1419
	13	Log-Log	$y = \exp^{y'}$	0,9832	18,38	257,30	0,9833	0,0603
	15	Raiz-Raiz	$y = y'^2$	0,9863	16,59	251,17	0,1013	0,6148

em que R<sup>2</sup>aj é o coeficiente de determinação ajustado do modelo; SEE é o erro padrão da estimativa; BIC é o critério de informação Bayesiano; p-BP é o p-valor do teste de homocedasticidade de Breusch-Pagan; e p-Li é o p-valor do teste de normalidade de Lilliefors.

FONTE: O autor (2022).

Dessa forma, na alometria da  $w_{total}$ , várias estratégias obtiveram bons resultados quanto ao atendimento simultâneo das premissas de normalidade residual e homogeneidade das variâncias (Tabela 17). Assim, para a modelagem alométrica, cuja variável independente foi a altura total das árvores, apenas as estratégias 1, 3 e 14 não atenderam às premissas requeridas. Porém, quando se utilizou o diâmetro como variável independente, muitas das estratégias testadas não obtiveram os resultados almejados, especificamente as estratégias 1, 3, 4, 6, 7, 9, 10, 12 e 14.

Vale enfatizar que, dentre as quinze estratégias de ajustes testadas, seis foram exitosas quando se usou o diâmetro como variável regressora. Por outro lado, quando se empregou a altura, doze estratégias foram bem-sucedidas.



Destaca-se que nas modelagens em que se aplicou o diâmetro, as transformações Log, Manly, GLog, Log shift, Log-Log e Raiz-Raiz foram as estratégias que atenderam simultaneamente as premissas requeridas. Enquanto isso, nas modelagens em função da altura, as estratégias promissoras foram as anteriormente citadas, em conjunto com as transformações Box-Cox, Bickel-Doksum, Yeo-Johnson, Dual de Yang, G power e Raiz ponderada.

Dentre essas estratégias, selecionou-se, em cada ocasião, apenas aquelas que obtiveram estatísticas de ajuste e precisão superiores às apresentadas pelos modelos sem transformação. Diante disso, a Tabela 17 detalha as estatísticas de regressão obtidas pelas transformações de melhores performances. Ressalta-se, que as metodologias mais hábeis se repetem para altura e diâmetro como variáveis independentes, com exceção das estratégias 12 e 5.

Enfatiza-se que o uso das transformações nas modelagens alométricas da  $w_{total}$ , além de atender as suposições paramétricas da regressão linear, melhoraram consideravelmente as estatísticas de  $R^2_{aj}$ , BIC e SEE (Tabela 17). Assim, sugere-se que as transformações duplas, Log-Log e Raiz-Raiz, mostraram-se as mais promissoras na modelagem da  $w_{total}$ .

Nesse sentido, com base nas análises gráficas dos ajustes, sugere-se que a incorporação dessas transformações às modelagens acarretou melhoras às estruturas residuais dos ajustes. Salienta-se que as transformações promoveram ganhos na homogeneização dos resíduos e na linearização dos quantis teóricos (Figura 12).

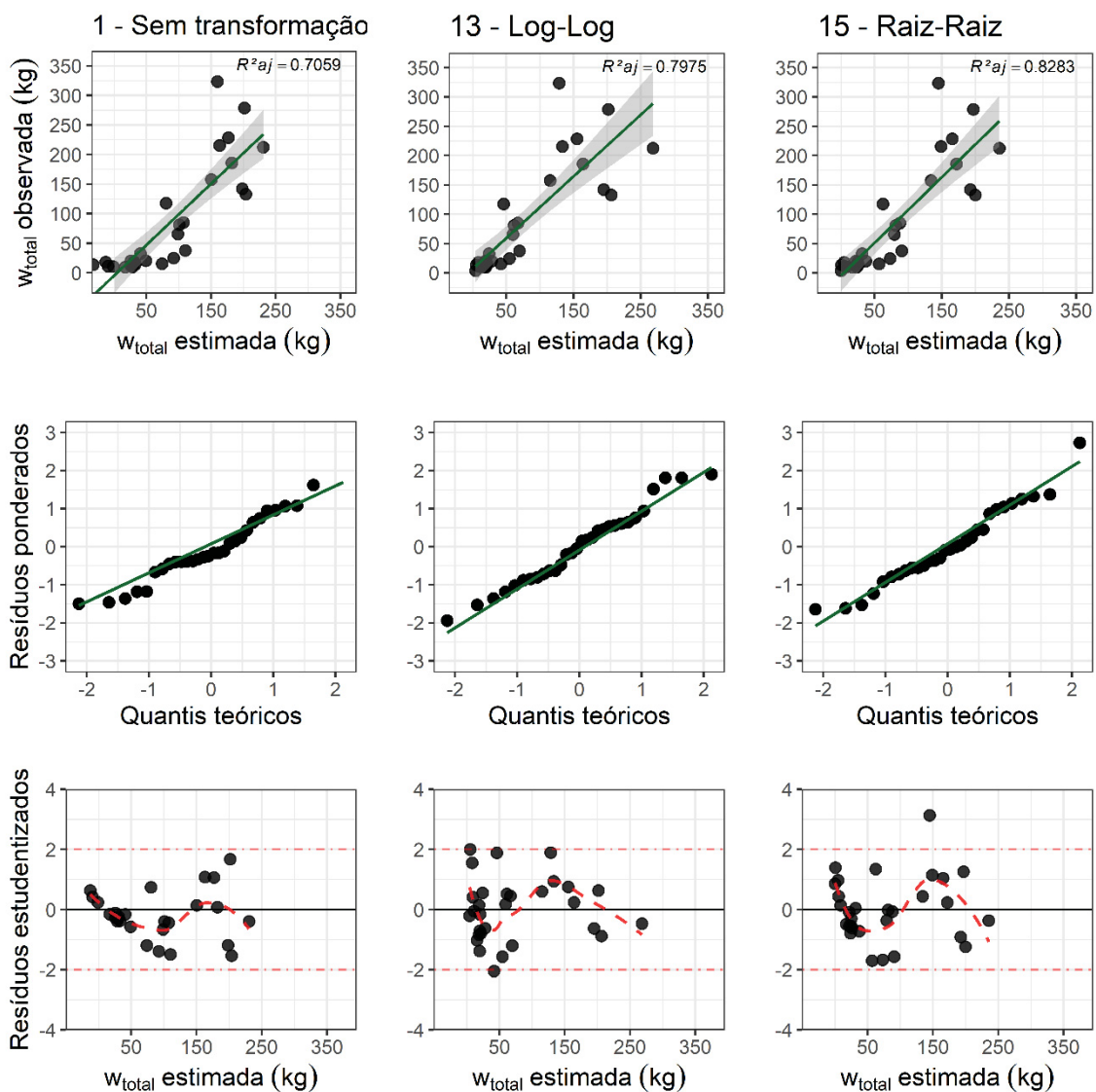
Em vista disso, nas modelagens da  $w_{total}$  em função da altura das árvores, verificou-se a habilidade das transformações Log-Log e Raiz-Raiz em linearizar as dispersões da  $w_{total}$ . Assim, eventuais padrões exponenciais dessa variável foram convertidos para uma tendência mais próxima à linear. Essa capacidade é evidenciada nas dispersões entre a  $w_{total}$  observada *versus* a estimada entre as modelagens com transformação (Log-Log e Raiz-Raiz) e a modelagem com as variáveis originais (Figura 12).

Os gráficos de dispersão residual corroboraram os resultados do teste de Breusch-Pagan a 5% (Tabela 17), que, por sua vez, demonstrou variâncias heterogêneas apenas para estratégia 1 ( $p-BP < 0,05$ ). Portanto, quando as variáveis do modelo foram submetidas a transformação Log-Log (13) ou Raiz-

Raiz (15) a variação heterogênea e enviesada dos resíduos, presente na estratégia 1, foi homogeneizada em torno do eixo das abscissas (Figura 12).

Dessa forma, enfatiza-se, que essas transformações demonstraram resultados contundentes para a homogeneização de variâncias nas análises de regressão linear da  $w_{total}$  em função da altura. Essa mesma eficácia é observada quanto ao atendimento da normalidade residual, analisada por meio dos quantis normais dos gráficos Q-Q (Figura 12).

FIGURA 12 - BIOMASSA TOTAL OBSERVADA VS ESTIMADA, GRÁFICOS Q-Q E DISTRIBUIÇÃO RESIDUAL DAS MELHORES ESTRATÉGIAS DE MODELAGEM DE BIOMASSA TOTAL DE *Mimosa scrabella* EM FUNÇÃO DA ALTURA.

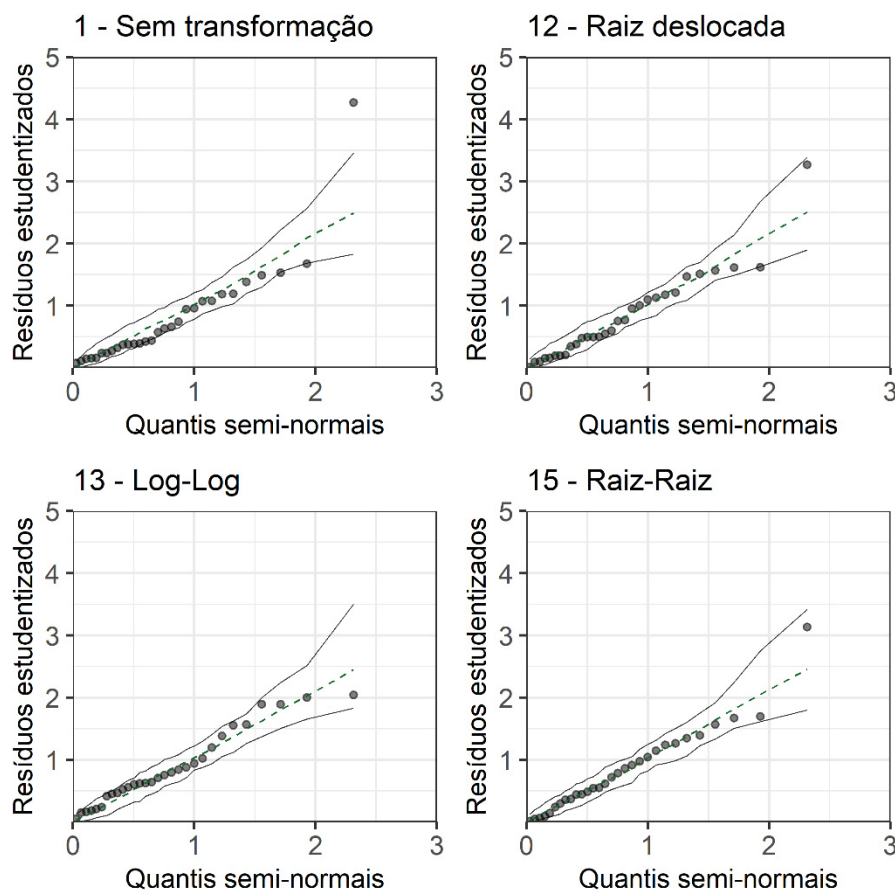


FONTE: O autor (2022).

Nessa perspectiva, destaca-se que as transformações apresentaram performances semelhantes para a estimativa da biomassa total. Assim, com base na análise dos gráficos hnp, as estratégias Raiz deslocada (12), Log-Log (13) e Raiz-Raiz (15) resultaram em resíduos dentro do envelope de confiança. Entretanto, a estratégia Raiz-Raiz sugeriu melhores qualidades de estimativa, visto que os resíduos gerados foram dispostos com menores distorções e respeitando os limites do envelope de confiança do gráfico hnp (Figura 13).

A modelagem com a estratégia 1 (sem transformação) não se demonstrou apropriada, pois alguns resíduos se projetaram além dos intervalos de confiança estabelecidos (Figura 13). Enquanto isso, as transformações raiz deslocada e Log-Log (12 e 13), apesar de promissoras, demonstraram qualidade de estimativa inferior a sugerida pela estratégia Raiz-Raiz (15).

FIGURA 13 - GRÁFICOS HNP (HALF-NORMAL PLOT) DAS MODELAGENS DA BIOMASSA TOTAL DE ÁRVORES DE *Mimosa scrabella* EM FUNÇÃO DA ALTURA.



FONTE: O autor (2022).

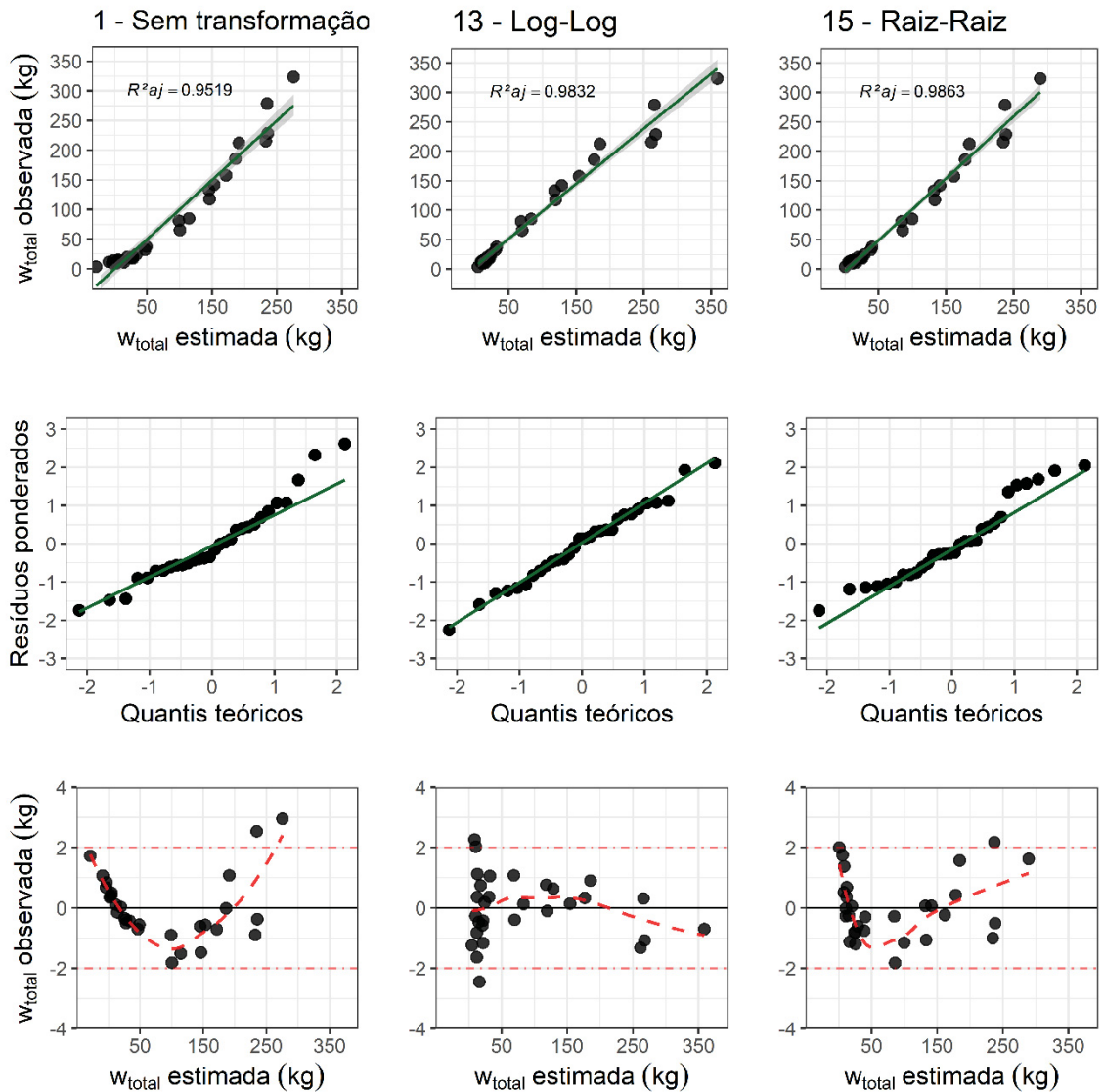
Não obstante, na ocasião em que a  $w_{total}$  foi estimada em função do diâmetro, as transformações selecionadas também proporcionaram ganhos estatísticos às modelagens. Na Figura 14, nota-se que os valores de  $w_{total}$  observados *versus* preditos foram linearizados com a aplicação das transformações logarítmica e de raiz quadrada. Essa capacidade influenciou diretamente na qualidade residual das modelagens, por promoverem a redução de eventuais enviesamentos.

Desse modo, observa-se que, quando a modelagem é executada com as variáveis em escala original (1), os resíduos gerados são dispostos com clara presença de tendenciosidade e variâncias heterocedásticas. No entanto, com a aplicação das estratégias de transformações Log-Log (13) e Raiz-Raiz (15), a consistência dos erros melhora consideravelmente, de modo que o enviesamento é equalizado e as variâncias tornam-se notadamente mais homogêneas (Figura 14). Esses resultados gráficos corroboram os obtidos no teste de aderência de Breusch-Pagan a 5%, nas situações em que a  $w_{total}$  é estimada em função do diâmetro das árvores (Tabela 17).

Os gráficos Q-Q gerados por essa classe de modelo também seguem as mesmas conclusões do teste de hipótese de Lilliefors a 5% para normalidade residual. Assim, na modelagem com variáveis originais, os quantis teóricos superiores divergem da tendência linear. Contudo, destaca-se que esse distúrbio é corrigido com a aplicação das transformações.

Sobretudo, em uma análise mais criteriosa dos gráficos Q-Q, o método de transformação Log-Log projeta os quantis teóricos em uma disposição mais uniforme e linear em relação ao método Raiz-Raiz. Entretanto, as sutis distorções dos quantis apresentadas por esse último método de transformação não comprometem a distribuição normal dos resíduos, de acordo com o teste de Lilliefors ao nível de 1% de probabilidade ( $p-Li > 0,01$ ).

FIGURA 12 - BIOMASSA TOTAL OBSERVADA VS ESTIMADA, GRÁFICOS Q-Q E DISTRIBUIÇÃO RESIDUAL DAS MELHORES ESTRATÉGIAS DE MODELAGEM DE BIOMASSA TOTAL DE *Mimosa scrabella* EM FUNÇÃO DO DIÂMETRO.



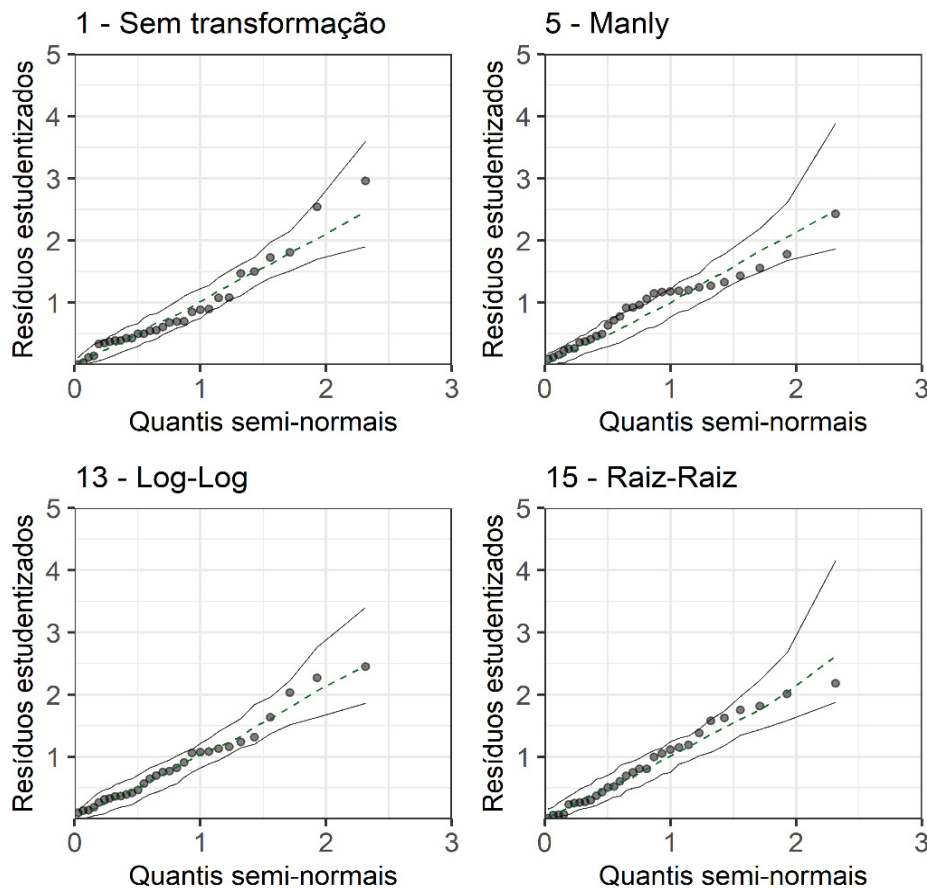
FONTE: O autor (2022).

Acerca dos resultados obtidos pela análise dos gráficos hnp, observou-se que a transformação de Manly projetou resíduos externos aos limites do envelope de confiança. Portanto, o uso dessa estratégia demonstrou qualidade de estimativa inferior à estratégia 1, que não envolveu uso de transformação (Figura 15).

Por outro lado, as estratégias Log-Log e Raiz-Raiz resultaram em resíduos dentro do envelope de confiança. Todavia, a disposição dos resíduos

por meio das transformações Log-Log sugeriu melhores qualidades de estimativa, corroborando com as estatísticas apresentadas na Tabela 17.

FIGURA 13 - GRÁFICOS HNP (HALF-NORMAL PLOT) DAS MODELAGENS DA BIOMASSA TOTAL DE ÁRVORES DE *Mimosa scrabella* EM FUNÇÃO DO DIÂMETRO.



FONTE: O autor (2022).

Na Tabela 18 estão apresentados, por estratégia de modelagem, os coeficientes de regressão e suas respectivas significâncias de acordo com o teste  $t$ . Para todos os cenários descritos, os coeficientes foram significativos a 1%. Contudo, como a estratégia 1 não apresentou adequação dos pressupostos residuais de normalidade e homocedasticidade, a inferência de significância pelo teste  $t$  foi válida apenas para as estratégias 13 e 15. Ademais, as equações geradas por essas estratégias foram mais eficientes (maior  $R^2_{aj}$  e menor BIC) e precisas (menor SEE%), o que as destacou como as mais acuradas para predição da  $w_{total}$  de árvores.

TABELA 18 - COEFICIENTES DE REGRESSÃO DAS MELHORES ESTRATÉGIAS DE MODELAGEM DA BIOMASSA TOTAL DE *Mimosa scrabella* EM FUNÇÃO DA ALTURA E DO DIÂMETRO.

Modelo	Estratégia	Transformação	$\beta_0$	SE $\beta_0$ %	$\beta_1$	SE $\beta_1$ %
$w_{total}$ (kg) em função da altura	1	Sem transformação	-188,921**	17,8561	21,4600**	11,90121
	13	Log-Log	-6,2363**	13,4246	3,9787**	8,359514
	15	Raiz-Raiz	-22,1017**	13,0904	8,4707**	9,582443
$w_{total}$ (kg) em função do diâmetro	1	Sem transformação	-98,6074**	8,58495	16,2546**	4,169281
	13	Log-Log	-2,25959**	6,22148	2,59651**	2,303477
	15	Raiz-Raiz	-11,6437**	4,45048	5,9727**	2,590118

em que  $\beta_0$  e  $\beta_1$  são os coeficientes de regressão do modelo; \*\* significativo ao nível 1% de probabilidade; \* significativo ao nível de 5% de probabilidade; <sup>ns</sup> não significativo; e SE $\beta_n$ % é o erro padrão percentual dos coeficientes estimados.

FONTE: O autor (2022).

## 5.6 ESTRATÉGIAS DE MODELAGEM DA BIOMASSA DO FUSTE DE ÁRVORES DE *Mimosa scrabella*

Para os ajustes da biomassa do fuste ( $w_{fuste}$ ), verificou-se que os modelos lineares cuja variável explicativa foi a altura, as suposições da regressão linear foram atendidas a 1% de probabilidade (Tabela 19;  $p-Li$  e  $p-BP > 0,01$ ). Porém, nas modelagens com o diâmetro, as variâncias foram consideradas heterocedásticas ( $p-BP < 0,05$ ). Diante disso, verificou-se que o uso de transformações se mostrou promissor nesse cenário, visto que foi possível corrigir ambos os pressupostos ao nível de significância estabelecido (Tabela 19).

Todavia, quando a  $w_{fuste}$  foi modelada em função da altura total, as estratégias de propriedade recíproca (3 e 14) promoveram a violação do pressuposto de homogeneidade das variâncias ( $p-BP < 0,05$ ). Por outro lado, quando as modelagens ocorreram em função do diâmetro, a violação da homogeneidade esteve relacionada às estratégias 1, 3, 12 e 14.

Diante disso, verifica-se que para ambas as estruturas de modelo, a maioria das estratégias de transformação foram promissoras para adequação dos pressupostos da regressão. Destaca-se, ainda, que nenhuma estratégia violou a suposição de normalidade residual.

Por conseguinte, as transformações mais eficientes aos modelos estruturados em função da altura das árvores de acordo com as estatísticas de ajuste e precisão das modelagens foram as associadas às estratégias 12, 13 e

15. Em contrapartida, quando a estrutura do modelo esteve relacionada ao uso do diâmetro, as estratégias de melhor performance foram a 4, 10 e 15. As estatísticas das modelagens de cada uma dessas estratégias de acordo com a variável explicativa usada estão descritas na Tabela 19.

TABELA 19 - ESTATÍSTICAS DE REGRESSÃO DAS MELHORES ESTRATÉGIAS DE MODELAGEM DA BIOMASSA DO FUSTE DE *Mimosa scrabella* EM FUNÇÃO DA ALTURA E DO DIÂMETRO.

Modelo	Estratégia	Transformação	Retransformação	R <sup>2</sup> aj	SEE%	BIC	p-Li	p-BP
$w_{fuste}$ (kg) em função da altura	1	Sem transformação	$y = y'$	0,7663	54,08	288,49	0,4286	0,0183
	12	Raiz deslocada	$y = (y'^2 - \lambda)$	0,8666	50,49	284,37	0,7883	0,0622
	13	Log-Log	$y = \exp^{y'}$	0,8964	55,63	290,19	0,4609	0,6624
	15	Raiz-Raiz	$y = y'^2$	0,8469	50,42	284,29	0,7125	0,2555
$w_{fuste}$ (kg) em função do diâmetro	1	Sem transformação	$y = y'$	0,9339	28,75	250,59	0,2914	0,0025
	4	Box-Cox	$y = (\lambda \cdot y' + 1)^{\frac{1}{\lambda}}$	0,9648	26,85	246,48	0,9946	0,0631
	10	G Power	$y = - \left[ \frac{1 - (\lambda y' + 1)^{\frac{2}{\lambda}}}{2(\lambda y' + 1)^{\frac{1}{\lambda}}} \right]$	0,9648	26,91	246,62	0,9956	0,0627
	15	Raiz-Raiz	$y = y'^2$	0,9616	24,12	240,05	0,9580	0,3744

em que R<sup>2</sup>aj é o coeficiente de determinação ajustado do modelo; SEE é o erro padrão da estimativa; BIC é o critério de informação Bayesiano; p-BP é o p-valor do teste de homocedasticidade de Breusch-Pagan; e p-Li é o p-valor do teste de normalidade de Lilliefors.

FONTE: O autor (2022).

Constatou-se, portanto, que o uso de transformações, além de trazer melhorias às estruturas residuais dos modelos, promoveram melhorias consideráveis às estatísticas de regressão. Diante disso, as modelagens com variáveis transformadas forneceram valores de R<sup>2</sup>aj, SEE e BIC mais promissores.

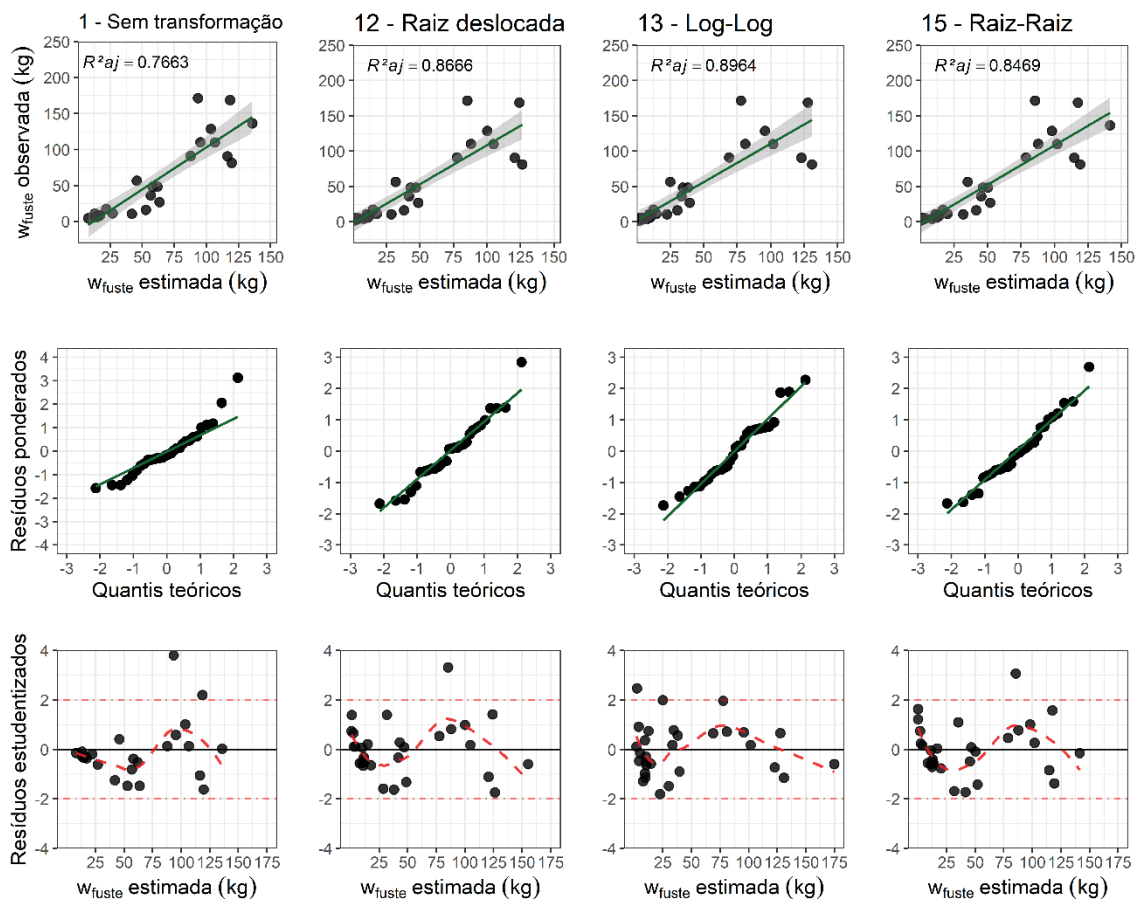
Nessa conjectura, a partir da análise dos gráficos Q-Q e da distribuição dos resíduos estudentizados para as modelagens em função da altura, verifica-se uma predisposição à violação da normalidade residual e da homogeneidade das variâncias para a estratégia 1, cuja modelagem é efetuada com as variáveis em escala original (Figura 16).

Para a estratégia 1, observa-se que a disposição dos quantis teóricos do gráfico Q-Q demonstram eventuais fugas da normalidade nas extremidades superiores. Destaca-se que esse comportamento evidencia uma propensão a não normalidade na distribuição dos erros gerados, embora o teste de hipótese de Lilliefors a 5% tenha aferido normalidade dos resíduos. Por outro lado, os gráficos Q-Q gerados pelas modelagens das demais estratégias não



evidenciaram qualquer fuga de normalidade, corroborando os resultados do teste de hipótese para normalidade residual.

FIGURA 14 - BIOMASSA DO FUSTE OBSERVADA VS ESTIMADA, GRÁFICOS Q-Q E DISTRIBUIÇÃO RESIDUAL E DAS MELHORES ESTRATÉGIAS DE MODELAGEM DE BIOMASSA DO FUSTE DE *Mimosa scrabella* EM FUNÇÃO DA ALTURA.



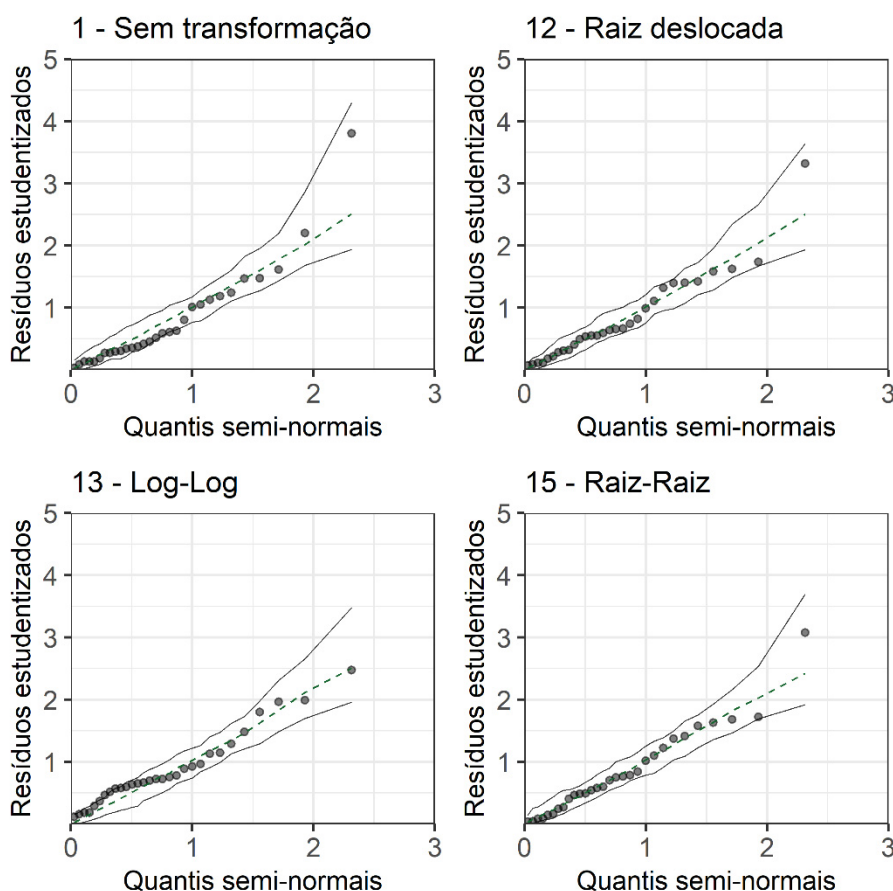
FONTE: O autor (2022).

Com relação aos gráficos de distribuição dos resíduos studentizados, observou-se distribuições heterocedásticas e enviesadas dos resíduos apenas para a estratégia 1, na qual o ajuste foi executado com variáveis em escala original. Os resíduos gerados por este ajuste se dispersaram em um formato de “funil”, de modo que os resíduos referentes aos maiores valores de biomassa ( $W_{fuste}$ ) apresentam maior variabilidade. Salienta-se que, apesar da análise residual sugerir essas interpretações, o teste de hipótese de Breusch-Pagan ao nível de 5% de probabilidade demonstrou aderência deste ajuste à homogeneidade das variâncias.

Entretanto, acerca das análises dos gráficos hnp, todas as modelagens selecionadas dispuseram seus resíduos respeitando os limites do envelope de

confiança, inclusive quando não se aplicou transformação nas variáveis (Figura 17). Esse comportamento correspondeu ao esperado de acordo com as estatísticas de ajuste da Tabela 19. Contudo, cabe enfatizar que a implementação de transformações no modelo acarretou melhoras às estimativas. Diante disso, destacou-se a estratégia de transformação Raiz-Raiz (15).

FIGURA 15 - GRÁFICOS HNP (HALF-NORMAL PLOT) DAS MODELAGENS DA BIOMASSA DO FUSTE DE ÁRVORES DE *Mimosa scrabella* EM FUNÇÃO DA ALTURA.



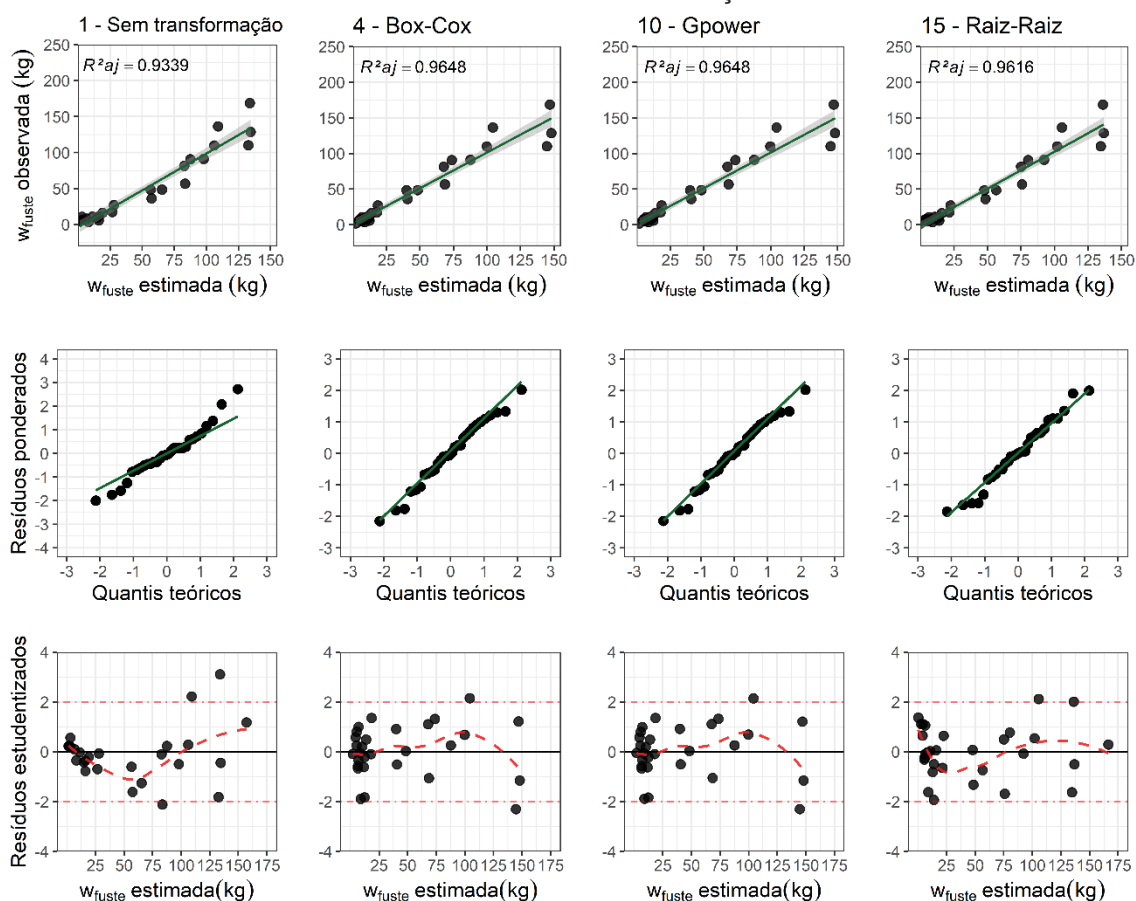
FONTE: O autor (2022).

Não obstante, para as modelagens da  $w_{fuste}$  cuja variável explicativa foi o diâmetro, as percepções quanto a normalidade residual e homogeneidade das variâncias obtidas pelas análises gráficas foram as mesmas fornecidas pelos testes de hipóteses de Lilliefors e Breusch-Pagan a 5%. Assim, de acordo com a dispersão dos resíduos estudentizados, apenas a modelagem com as variáveis na escala original violou o pressuposto de homogeneidade das variâncias.

Com relação à normalidade residual, foi possível observar sutis distorções nos gráficos Q-Q das estratégias 1 e 10 (Variáveis originais e G power, respectivamente) (Figura 18). Apesar disso, não houve indícios de violação dessa suposição para estas modelagens de acordo com o teste de Lilliefors a 5% (Tabela 19).

Enfatiza-se, ainda, que os métodos de transformação de Box-Cox e Raiz-Raiz foram os que promoveram simultaneamente melhores distribuições dos quantis teóricos e dos resíduos estudentizados. Para essas estratégias, foram observadas distribuições mais lineares dos quantis teóricos nos gráficos Q-Q e dispersões mais homogêneas nos gráficos de distribuição dos resíduos estudentizados (Figura 18).

FIGURA 16 - BIOMASSA DO FUSTE OBSERVADA VS ESTIMADA, GRÁFICOS Q-Q E DISTRIBUIÇÃO RESIDUAL DAS MELHORES ESTRATÉGIAS DE MODELAGEM DE BIOMASSA DO FUSTE DE *Mimosa scrabella* EM FUNÇÃO DO DIÂMETRO.

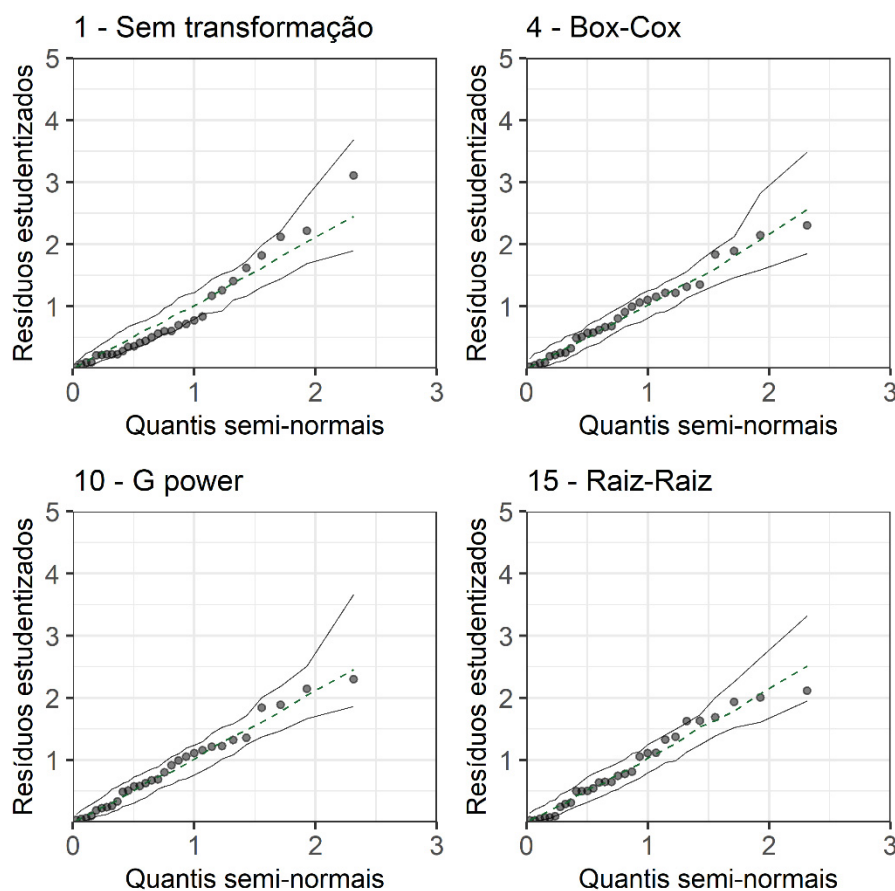


FONTE: O autor (2022).

Por outro lado, enfatiza-se que a linearidade entre as  $w_{fuste}$  observadas e estimadas foi semelhante em todos os cenários de modelagem (Figuras 16 e 18). Contudo, essa similaridade entre os pares de dados observados e estimados foi mais evidente quando se usou o diâmetro para a predição da  $w_{fuste}$ .

Acerca da eficiência nas estimativas por cada estratégia, observou-se que a ausência de transformação se mostrou inadequada para estimativas, visto que muitos resíduos ultrapassaram o limite inferior do envelope de confiança do gráfico hnp (Figura 19). Destaca-se, porém, que as transformações demonstraram qualidade de estimativa promissoras, dentre as quais se sobressaíram os métodos Box-Cox e G power, com performances muito semelhantes.

FIGURA 17 - GRÁFICOS HNP (HALF-NORMAL PLOT) DAS MODELAGENS DA BIOMASSA DO FUSTE DE ÁRVORES DE *Mimosa scrabella* EM FUNÇÃO DO DIÂMETRO.



FONTE: O autor (2022).

Na Tabela 20 estão apresentados os coeficientes de regressão gerados e suas respectivas significâncias de acordo com o teste  $t$  para cada estratégia de modelagem. Em todas as modelagens, os coeficientes estimados foram significativos a 1%.

Entretanto, para a estratégia 1, a inferência estatística acerca do teste  $t$  torna-se descredibilizada devido a inadequação do pressuposto de homogeneidade das variâncias. Ressalta-se, ainda, que as equações obtidas pelas estratégias com variáveis transformadas, além de mais eficientes (maior  $R^2_{aj}$  e menor BIC) e precisas (menor SEE%), se destacam como as mais adequadas para o objetivo ao qual foram propostas, visto que atenderam simultaneamente aos pressupostos requeridos para a análise por regressão linear.

TABELA 20 - COEFICIENTES DE REGRESSÃO E PARÂMETROS DE TRANSFORMAÇÃO DAS MELHORES ESTRATÉGIAS DE MODELAGEM DA BIOMASSA DO FUSTE DE *Mimosa scrabella* EM FUNÇÃO DA ALTURA E DO DIÂMETRO.

Modelo	Estratégia	Transformação	$\beta_0$	SE $\beta_0$ %	$\beta_1$	SE $\beta_1$ %	$\lambda$
$w_{fuste}$ (kg) em função da altura	1	Sem transformação	-115,3736**	15,0039	12,847433**	10,2020	
	12	Raiz deslocada	-6,762161**	13,9319	0,981462**	7,2647	-0,9069
	13	Log-Log	-7,864368**	8,8398	4,382083**	6,3006	
	15	Raiz-Raiz	-18,102585**	10,5151	6,783555**	7,8719	
$w_{fuste}$ (kg) em função do diâmetro	1	Sem transformação	-56,152365**	10,1750	9,271509**	4,9333	
	4	Box-Cox	-3,733010**	12,6225	1,063956**	3,5433	0,4388
	10	G Power	-4,10678200**	15,3137	1,419926**	3,5424	0,4356
	15	Raiz-Raiz	-9,058392**	6,2715	4,577430**	3,7051	

em que  $\beta_0$  e  $\beta_1$  são os coeficientes de regressão do modelo; \*\* significativo ao nível 1% de probabilidade; \* significativo ao nível de 5% de probabilidade; ns não significativo; SE $\beta_n$ % é o erro padrão percentual dos coeficientes estimados; e  $\lambda$  é o coeficiente de transformação.

FONTE: O autor (2022).

## 5.7 ESTRATÉGIAS DE MODELAGEM DA BIOMASSA DOS GALHOS DE ÁRVORES DE *Mimosa scrabella*

As modelagens da biomassa dos galhos ( $w_{galhos}$ ) em função da altura e diâmetro em suas escalas originais violaram o pressuposto de homogeneidade das variâncias, ao passo que atenderam o de normalidade residual ao nível de 1% de probabilidade pelo teste de Lilliefors. Todavia, com a aplicação de estratégias de transformação de variáveis, foi possível corrigir simultaneamente essas premissas (Tabela 21).

Porém, destaca-se que, na ocasião em que a  $w_{galhos}$  foi modelada em função da altura total, somente as estratégias com propriedades recíprocas (3 e 14) não atenderam simultaneamente os pressupostos da regressão linear. Nesse sentido, todas as demais estratégias de transformação foram promissoras para esse cenário de ajuste da  $w_{galhos}$ .

Enquanto isso, acerca das modelagens em função do diâmetro, apenas as estratégias 2, 8 e 13 (Log, Glog e Log-Log, respectivamente) atenderam simultaneamente as suposições da regressão linear. Nesse caso, as transformações com propriedades logarítmicas foram as mais promissoras na adequação dos pressupostos paramétricos.

Destaca-se que as transformações selecionadas promoveram melhorias nas estatísticas de regressão. Entretanto, destaca-se que, apesar do aumento na explicação dos modelos (maior  $R^2_{aj}$ ), a aplicação das transformações Log e Glog promoveram SEE maiores aos obtidos pela modelagem sem transformação. Essas percepções demonstram que as modelagens que envolvem essas transformações apresentam perda de precisão (Tabela 21).

TABELA 21 - ESTATÍSTICAS DE REGRESSÃO DAS MELHORES ESTRATÉGIAS DE MODELAGEM DA BIOMASSA DE GALHOS DE *Mimosa scrabella* EM FUNÇÃO DA ALTURA E DO DIÂMETRO.

Modelo	Estratégia	Transformação	Retransformação	$R^2_{aj}$	SEE%	BIC	p-Li	p-BP
$w_{galhos}$ (kg) em função da altura	1	Sem transformação	$y = Y$	0,5273	82,85	242,74	0,0172	0,0033
	5	Manly	$y = \frac{\log(\lambda \cdot Y' + 1)}{\lambda}$	0,7106	92,29	249,21	0,6511	0,4909
	8	Glog	$y = -\frac{[1 - (\exp^{y'})^2]}{2\exp^{y'}}$	0,6679	95,18	251,06	0,6560	0,9512
	9	Dual power	$y = \left[ \sqrt{1 + \lambda^2 y'^2} + \lambda y' \right]^{\frac{1}{\lambda}}$	0,6650	92,16	249,13	0,5647	0,8181
$w_{galhos}$ (kg) em função do diâmetro	1	Sem transformação	$y = Y$	0,8574	45,50	206,78	0,0124	0,0000
	2	Log	$y = \exp^{y'}$	0,9012	73,06	235,19	0,1706	0,9613
	8	Glog	$y = -\frac{[1 - (\exp^{y'})^2]}{2\exp^{y'}}$	0,9162	66,78	229,80	0,4163	0,5301
	13	Log-Log	$y = \exp^{y'}$	0,9448	38,82	197,25	0,5196	0,4519

em que  $R^2_{aj}$  é o coeficiente de determinação ajustado do modelo; SEE é o erro padrão da estimativa; BIC é o critério de informação Bayesiano; p-BP é o p-valor do teste de homocedasticidade de Breusch-Pagan; e p-Li é o p-valor do teste de normalidade de Lilliefors.

FONTE: O autor (2022).

Assim, com base nas estatísticas de ajuste ( $R^2_{aj}$  e BIC) e precisão (SEE), a estratégia fundamentada na transformação de Manly (estratégia 5) obteve a melhor performance dentre as modelagens em função da altura. Dessa

forma, considerando os  $R^2_{aj}$  obtidos, a transformação de Manly aumentou a explicação do modelo em relação ao modelo sem transformação. Enquanto isso, para as modelagens em função do diâmetro, a transformação Log-Log (estratégia 13) obteve melhor desempenho de ajuste promovendo o aumento do  $R^2_{aj}$  em relação a modelagem com as variáveis originais.

Nesse sentido, as transformações de Manly e Log-Log mostram-se estatisticamente mais apropriadas para a aplicação em regressão linear da  $w_{galhos}$ . No entanto, salienta-se que, para o atendimento da normalidade residual e homogeneidade das variâncias, todas as demais estratégias de transformação selecionadas impõem-se como aptas a substituir as modelagens com variáveis em escala original.

Diante disso, as análises gráficas das modelagens com as estratégias selecionadas demonstram as mesmas percepções evidenciadas pelos testes de hipóteses de Lilliefors e Breusch-Pagan a 5% de probabilidade. Assim, a partir das análises dos gráficos Q-Q e dos gráficos de resíduos, infere-se que apenas as modelagens sem transformações violaram as premissas exigidas.

Dessa forma, observou-se pelas distribuições dos quantis teóricos indícios de violação da premissa de normalidade residual para o ajuste da estratégia sem transformação (1) nas modelagens em função da altura (Figura 20). Apesar disso, os resultados dos testes de hipóteses de Lilliefors a 1% de probabilidade demonstra para o ajuste dessa estratégia aderência à normalidade residual.

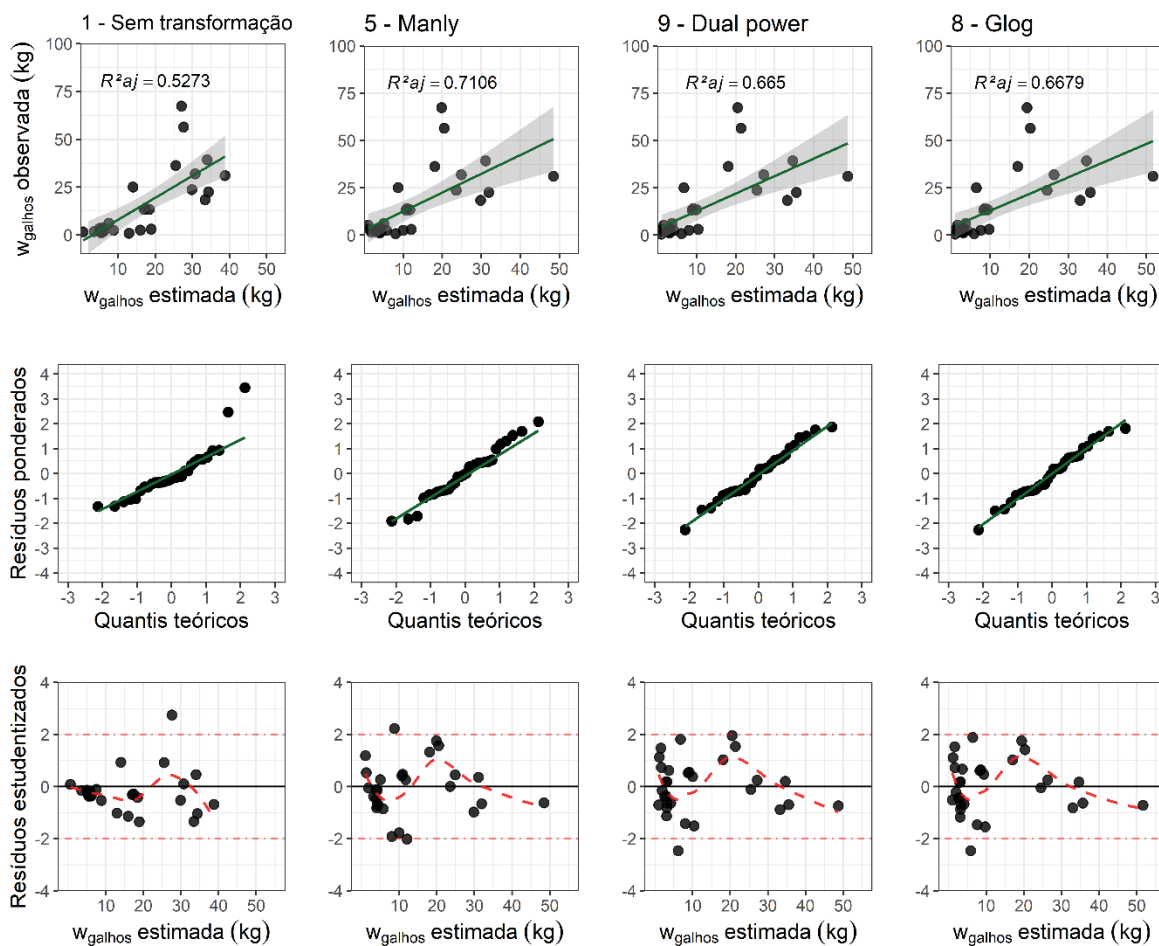
Não obstante, os gráficos Q-Q das modelagens das demais estratégias selecionadas expressam a adequação dessa premissa, visto que a consistente disposição linear dos quantis supõem a distribuição normal dos resíduos (Figura 20). Dessa forma, dentre os modelos em que se adotou a altura como variável explicativa, aqueles que envolveram as transformações de Manly, Glog e Dual power (estratégias 5, 8 e 9, respectivamente) promoveram melhoras às estruturas residuais.

Considerando-se ainda a estrutura de modelo envolvendo a altura, o diagnóstico de homogeneidade das variâncias obtido pelos gráficos dos resíduos estudentizados das estratégias 5, 8 e 9 seguiu as mesmas evidências fornecidas pelo teste de hipótese de Breusch-Pagan a 5% de probabilidade (Tabela 21). Assim, o comportamento heterocedástico das variâncias referentes a

modelagem da estratégia 1 foi corrigido com o uso das transformações de Manly, Dual power e Glog (Figura 20).

Por outro lado, pela dispersão entre os pares de dados de  $w_{galhos}$  observada *versus* estimada, verificou-se que as modelagens com as transformações de Manly Glog e Dual power tenderam a superestimar os maiores valores de  $w_{galhos}$ , bem com subestimar os menores. Apesar disso, essas modelagens equalizaram a evidente heterogeneidade residual presente no ajuste com as variáveis sem transformação (estratégia 1).

FIGURA 18 - BIOMASSA DOS GALHOS OBSERVADA VS ESTIMADA, GRÁFICOS Q-Q E DISTRIBUIÇÃO RESIDUAL DAS MELHORES ESTRATÉGIAS DE MODELAGEM DE BIOMASSA DOS GALHOS DE *Mimosa scrabella* EM FUNÇÃO DA ALTURA.



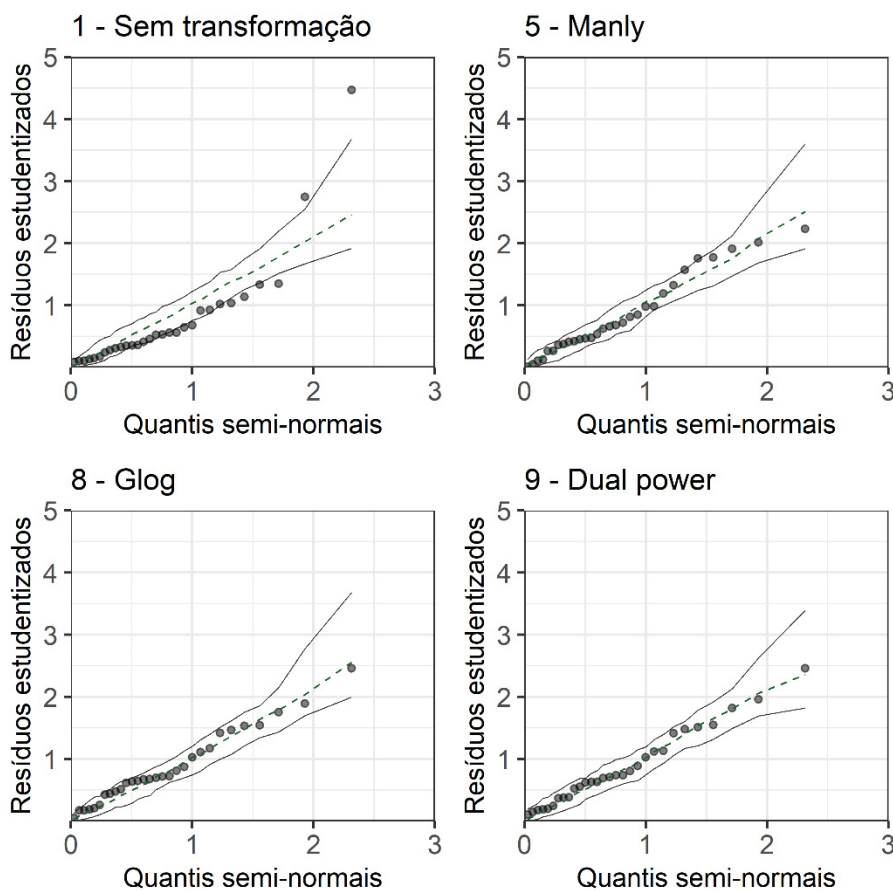
FONTE: O autor (2022).

Com relação a qualidade preditiva dessas estratégias analisadas nos gráficos hnp, observou-se que a estratégia 1, com ausência de transformação, se mostrou inadequada para modelagem da  $w_{galhos}$ , pois os resíduos gerados pela modelagem com as variáveis em escala original ultrapassaram os limites



inferior e superior do envelope de confiança a 5% de probabilidade (Figura 21). Enquanto isso, as transformações Glog (8) e Dual power (9) mostraram melhores performances preditivas, ao passo que a transformação de Manly (5), apesar promissora, dispôs alguns resíduos sobrepostos ao limite do envelope de confiança.

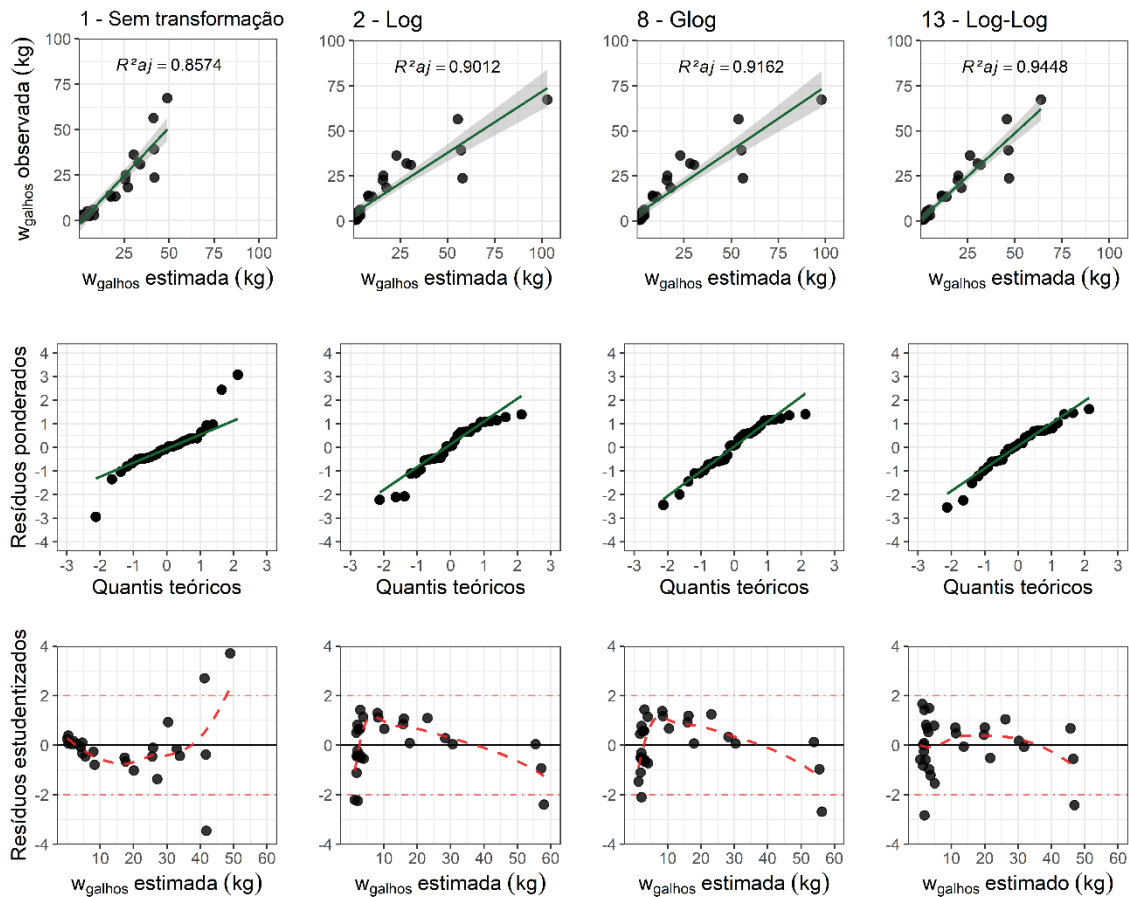
FIGURA 19 - GRÁFICOS HNP (HALF-NORMAL PLOT) DAS MODELAGENS DA BIOMASSA DOS GALHOS DE ÁRVORES DE *Mimosa scrabella* EM FUNÇÃO DA ALTURA.



FONTE: O autor (2022).

Quando a alometria da  $w_{galhos}$  foi conduzida em função do diâmetro, as modelagens com variáveis transformadas apresentaram estruturas residuais mais satisfatórias. Diante disso, no emprego da estratégia 1 para modelagem de  $w_{galhos}$ , constatou-se uma distribuição residual heterocedástica, com variâncias crescentes que demonstram um clássico formato de “funil” ou “leque”. Contudo, essa violação de homocedasticidade residual foi corrigida com a incorporação das transformações Log, Glog e Log-Log às variáveis do modelo (Figura 22).

FIGURA 20 - BIOMASSA DOS GALHOS OBSERVADA VS ESTIMADA, GRÁFICOS Q-Q E DISTRIBUIÇÃO RESIDUAL DAS MELHORES ESTRATÉGIAS DE MODELAGEM DE BIOMASSA DOS GALHOS DE *Mimosa scrabella* EM FUNÇÃO DO DIÂMETRO.



FONTE: O autor (2022).

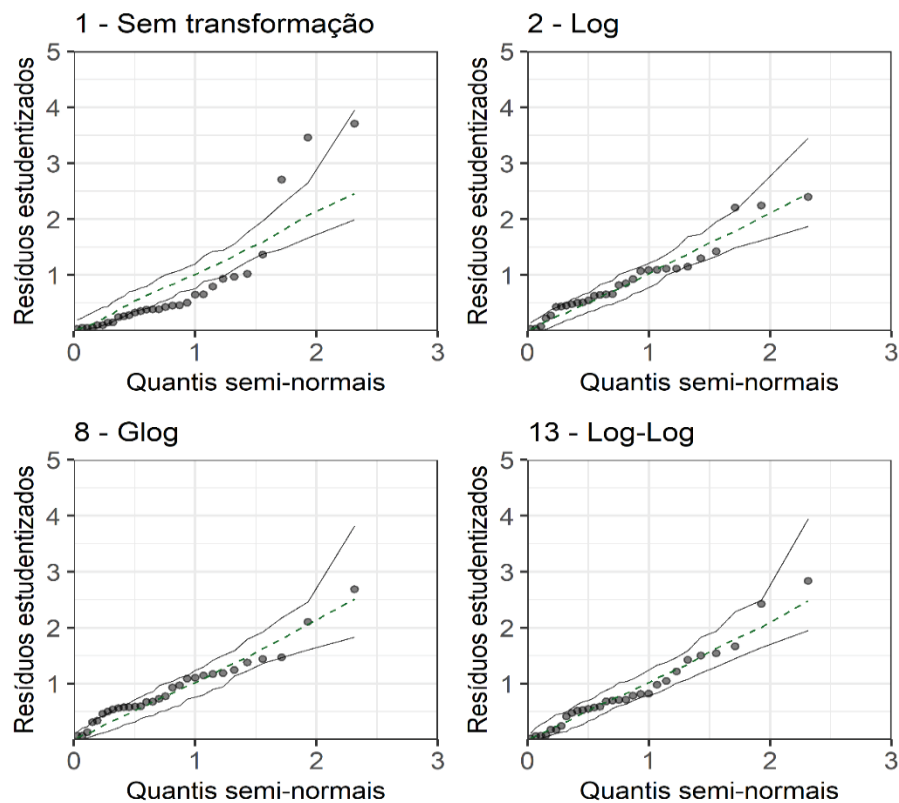
Acerca da normalidade residual, verifica-se, pelo gráfico Q-Q da estratégia 1, fugas da linearidade tanto nos quantis superiores quanto nos inferiores. Apesar de evidente, esse comportamento não atribui distribuições residuais não normais para o ajuste da estratégia sem transformação, visto que o teste de Lilliefors indicou distribuição normal dos resíduos ao nível de significância de 1%.

Todavia, salienta-se que, no esforço das transformações em adequar a modelagem às suposições paramétricas, houve uma propensão em superestimar algumas das maiores observações de  $w_{galhos}$ . Assim, diante da análise entre as  $w_{galhos}$  observadas *versus* estimada, nota-se que as transformações Log (2) e Glog (8) projetaram suas estimativas distantes dos valores reais da variável de interesse. Entretanto, com o uso da estratégia 13 (Log-Log), na qual se aplicou a transformação logarítmica tanto na variável

dependente quanto na independente, as estimativas tornaram-se mais próximas aos valores originais.

Os gráficos hnp ratificaram essas evidências, visto que a estratégia Log-Log (13) apresentou melhor qualidade preditiva. Todavia, destaca-se que as transformações Log e Glog também forma promissoras, distribuindo resíduos dentro do limite de confiabilidade estabelecidos de 5%. Ressalta-se que a modelagem com ausência de transformação mostrou-se inadequada (Figura 23), corroborando com as estatísticas apresentadas na Tabela 21.

FIGURA 21 - GRÁFICOS HNP (HALF-NORMAL PLOT) DAS MODELAGENS DA BIOMASSA DOS GALHOS DE ÁRVORES DE *Mimosa scrabella* EM FUNÇÃO DO DIÂMETRO.



FONTE: O autor (2022).

Destaca-se que todos os modelos apresentaram coeficientes significativos de acordo com o teste  $t$  ao nível de 1% de probabilidade. A Tabela 22 descreve, para cada estratégia de modelagem selecionada, os coeficientes de regressão obtidos e seus erros padrão relacionados.

TABELA 22 - COEFICIENTES DE REGRESSÃO E PARÂMETROS DE TRANSFORMAÇÃO DAS MELHORES ESTRATÉGIAS DE MODELAGEM DA BIOMASSA DE GALHOS DE *Mimosa scrabella* EM FUNÇÃO DA ALTURA E DO DIÂMETRO.

Modelo	Estratégia	Transformação	$\beta_0$	SE $\beta_0$ %	$\beta_1$	SE $\beta_1$ %	$\lambda$
$w_{galhos}$ (kg) em função da altura	1	Sem transformação	-30,28976**	26,6615	3,530986**	17,3181	
	5	Manly	-9,955560**	20,8939	1,338388**	11,7679	-0,0580
	8	Glog	-1,151353**	44,1046	0,296128**	12,9674	
	9	Dual power	-2,416363**	24,9756	0,349634**	13,0708	0,2290
$w_{galhos}$ (kg) em função do diâmetro	1	Sem transformação	-18,17273**	15,1485	2,918930**	7,5507	
	2	Log	-0,756697**	23,7347	0,234294**	6,1461	
	8	Glog	0,075020**	211,0104	0,226064**	5,6179	
	13	Log-Log	-4,417925**	6,5189	2,734038**	4,4806	

em que  $\beta_0$  e  $\beta_1$  são os coeficientes de regressão do modelo; \*\* significativo ao nível 1% de probabilidade; \* significativo ao nível de 5% de probabilidade; ns não significativo; SE $\beta_n$ % é o erro padrão percentual dos coeficientes estimados; e  $\lambda$  é o coeficiente de transformação.

FONTE: O autor (2022).

## 5.8 ESTRATÉGIAS DE MODELAGEM DA BIOMASSA DAS FOLHAS DE ÁRVORES DE *Mimosa scrabella*

Nesta etapa do estudo, as modelagens da biomassa das folhas ( $w_{folhas}$ ) demonstraram a necessidade do uso de transformação, pois nas modelagens abordando a altura e o diâmetro como variáveis preditoras infringiu-se o pressuposto de homogeneidade das variâncias (Tabela 23). As violações dessa suposição da análise de regressão linear foram determinadas a 1% de probabilidade, por meio do teste de hipótese de Breusch-Pagan.

Assim, de acordo os testes de hipóteses, verificou-se para os modelos em que o diâmetro foi empregado como variável explicativa que as estratégias 3, 9, 13 e 14 violaram a normalidade dos resíduos, ao passo que apenas a estratégia 1 infringiu a homocedasticidade das variâncias. Por outro lado, nos modelos envolvendo a altura para predição da  $w_{folha}$ , apenas as estratégias 3 e 14 violaram a premissa de normalidade dos erros, enquanto a estratégia 1 violou exclusivamente a homogeneidade das variâncias.

Em suma, enfatiza-se que, para as duas estruturações de modelo destinadas a alometria da  $w_{folha}$ , destacaram-se as técnicas de transformação de Manly, G power e Raiz quadrada deslocada (estratégias 5, 10 e 12, respectivamente) como as mais promissoras. Ressalta-se que essas estratégias se sobressaíram estatisticamente às demais para os dois cenários de ajuste (Tabela 23).

TABELA 23 - ESTATÍSTICAS DE REGRESSÃO DAS MELHORES ESTRATÉGIAS DE MODELAGEM DA BIOMASSA DE FOLHAS DE *Mimosa scrabella* EM FUNÇÃO DA ALTURA E DO DIÂMETRO.

Modelo	Estratégia	Transformação	Retransformação	R <sup>2</sup> aj	SEE%	BIC	p-Li	p-BP
<i>w<sub>folhas</sub></i> (kg) em função da altura	1	Sem transformação	$y = y$	0,4535	61,43	126,55	0,8714	0,0048
	5	Manly	$y = \frac{\log(\lambda \cdot y' + 1)}{\lambda}$	0,4636	63,68	128,70	0,7052	0,2268
	10	G power	$y = - \left[ \frac{1 - (\lambda y' + 1)^{\frac{2}{\lambda}}}{2(\lambda y' + 1)^{\frac{1}{\lambda}}} \right]$	0,4423	64,07	129,07	0,5361	0,3446
	12	Raiz deslocada	$y = (y'^2 - \lambda)$	0,4602	61,78	126,89	0,9900	0,0542
<i>w<sub>folhas</sub></i> (kg) em função do diâmetro	1	Sem transformação	$y = y$	0,8085	36,37	95,09	0,0438	0,0057
	5	Manly	$y = \frac{\log(\lambda \cdot y' + 1)}{\lambda}$	0,8127	34,71	92,30	0,2254	0,4432
	10	G power	$y = - \left[ \frac{1 - (\lambda y' + 1)^{\frac{2}{\lambda}}}{2(\lambda y' + 1)^{\frac{1}{\lambda}}} \right]$	0,8058	34,93	92,68	0,1224	0,6481
	12	Raiz deslocada	$y = (y'^2 - \lambda)$	0,8116	35,13	93,01	0,1211	0,2603

em que R<sup>2</sup>aj é o coeficiente de determinação ajustado do modelo; SEE é o erro padrão da estimativa; BIC é o critério de informação Bayesiano; p-BP é o p-valor do teste de homocedasticidade de Breusch-Pagan; e p-Li é o p-valor do teste de normalidade de Lilliefors.

FONTE: O autor (2022).

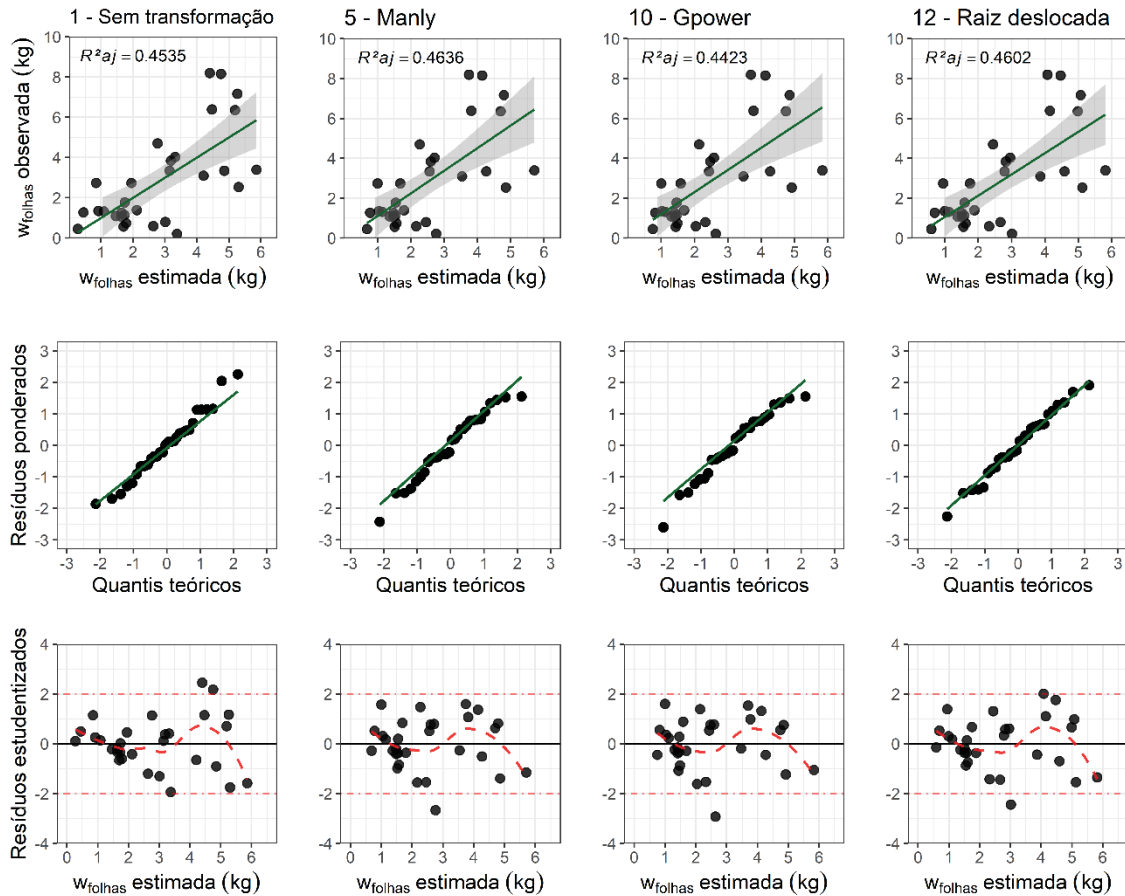
Apesar da aplicação dessas transformações agregarem melhorias nas estatísticas das regressões, foram ínfimas as mudanças nos valores dos R<sup>2</sup>aj, SEE% e BIC em comparação às modelagens com as variáveis em escala original (estratégia 1). Todavia, destaca-se que, quando se empregou o diâmetro como variável preditora, os modelos apresentaram maior capacidade explicativa da *w<sub>folha</sub>*. Assim, para os ajustes com o diâmetro, os R<sup>2</sup>aj foram de aproximadamente 0,80, enquanto que, para os com a altura, eles foram de aproximadamente 0,45 (Tabela 23).

De acordo com as percepções estatísticas obtidas pelas regressões lineares, a estratégia fundamentada na transformação de Manly promoveu as melhores performances para as duas conjecturas de modelos alométricos testadas (Tabela 23). Diante disso, esta estratégia forneceu, em comparação as outras, maiores R<sup>2</sup>aj e menores SEE% e BIC, proporcionando os melhores ajustes independentemente da variável explicativa empregada.

A Figura 24 detalha o comportamento residual para os modelos estruturados com a altura em  $\mathcal{X}$ . Assim, com base nos gráficos Q-Q das modelagens, verificou-se que, para nenhuma estratégia, foi observada fuga da normalidade residual. Contudo, a análise gráfica das distribuições residuais dos ajustes indicou heterocedasticidade à estratégia 1, visto que os resíduos

estudentizados demonstraram variâncias não constantes e heterogêneas em torno do eixo das abscissas.

FIGURA 22 - BIOMASSA DE FOLHAS OBSERVADA VS ESTIMADA, GRÁFICOS Q-Q E DISTRIBUIÇÃO RESIDUAL DAS MELHORES ESTRATÉGIAS DE MODELAGEM DE BIOMASSA DAS FOLHAS DE *Mimosa scrabella* EM FUNÇÃO DA ALTURA.



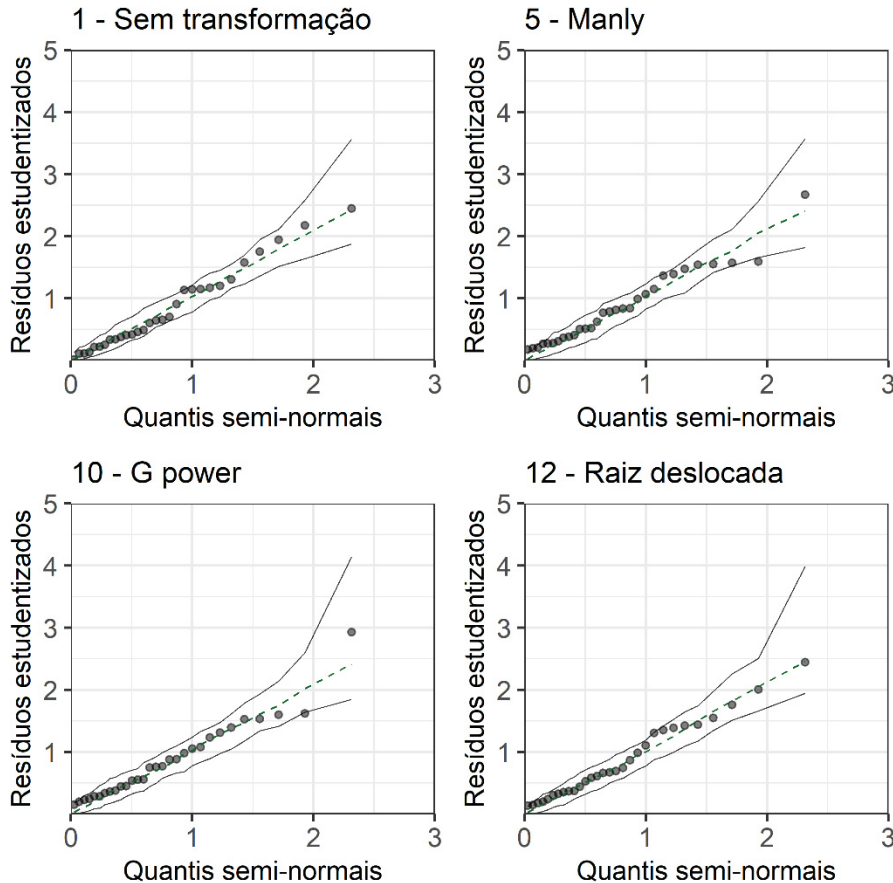
FONTE: O autor (2022).

Na análise gráfica da  $w_{folhas}$  observada versus estimada, nota-se que as mudanças nas dispersões foram ínfimas entre as modelagens com transformações e as com variáveis originais (Figura 24). Portanto, enfatiza-se que as transformações selecionadas se mostraram eficientes à finalidade proposta, pois à medida que adequaram o pressuposto violado (homogeneidade das variâncias), não infringiram aquele que outrora estava em adequação (normalidade residual). Essa evidência demonstra objetividade das transformações.

Acerca da eficiência preditiva das estratégias selecionadas, observou-se que a transformação de Raiz deslocada (12) foi mais adequada (Figura 25). Destaca-se que, apesar da violação da suposição de homogeneidade das

variâncias, a modelagem sem transformação também promoveu boa eficiência preditiva, visto que os resíduos gerados respeitaram o envelope de confiança estabelecidos no gráfico hnp.

FIGURA 23 - GRÁFICOS HNP (HALF-NORMAL PLOT) DAS MODELAGENS DA BIOMASSA DAS FOLHAS DE ÁRVORES DE *Mimosa scrabella* EM FUNÇÃO DA ALTURA.



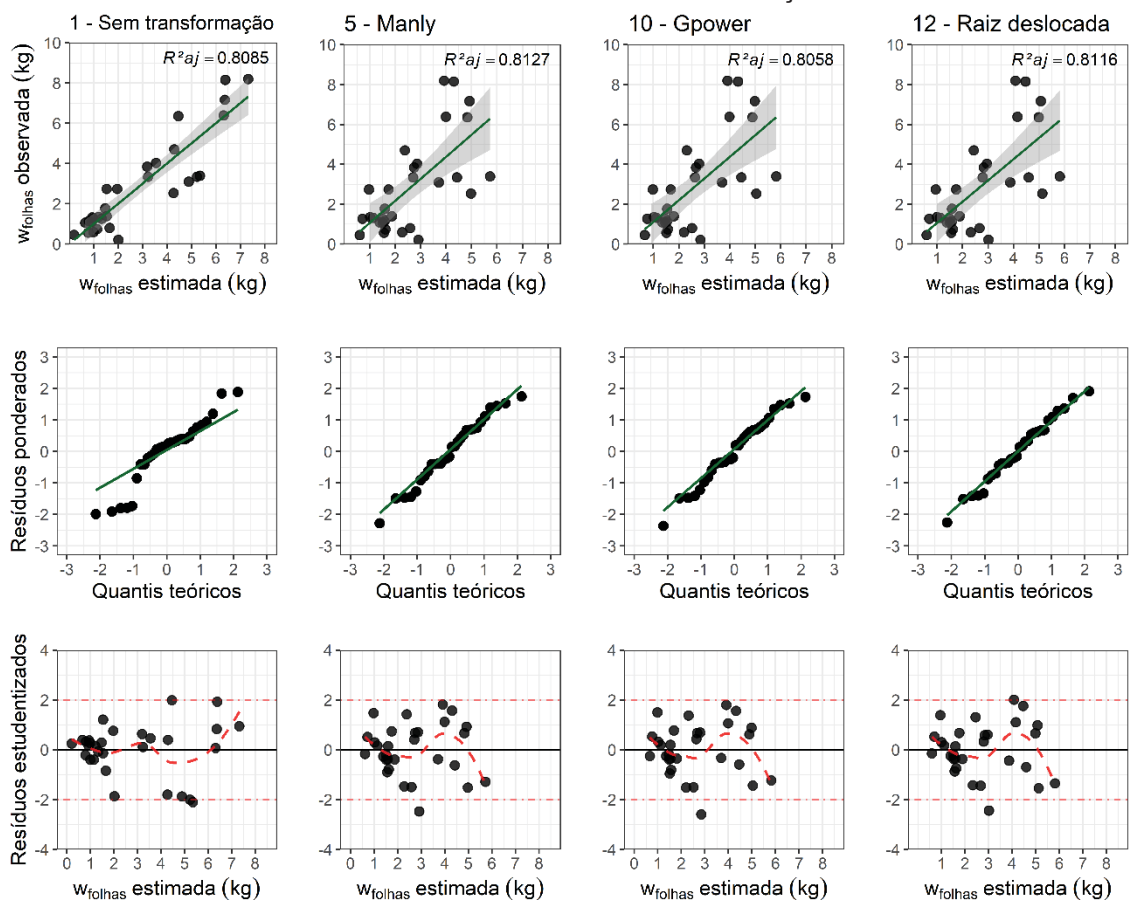
FONTE: O autor (2022).

Para os modelos estruturados com o diâmetro como preditor da  $w_{folhas}$ , a regressão linear com as variáveis na escala original não atendeu ao pressuposto de homogeneidade das variâncias de acordo com testes de hipótese de Breusch-Pagan, ao nível de 1% de probabilidade (Tabela 23). A distribuição dos resíduos, por sua vez, também reforça a violação dessa premissa, visto que a dispersão dos erros gerados é inconsistente e exhibe variâncias não constantes ao longo do eixo das abscissas (Figura 26).

Enquanto isso, a distribuição dos quantis teóricos (gráfico Q-Q) do ajuste da estratégia 1 demonstrou um comportamento que tendeu para a não

normalidade dos resíduos. Verificou-se, portanto, distribuições não lineares dos quantis inferiores e superiores para essa estratégia, sugerindo uma distribuição enviesada dos resíduos. Apesar disso, de acordo com o teste de hipótese de Lilliefors a 1%, não houve violação da suposição de distribuição normal para os resíduos fornecidos pelo ajuste dessa estratégia (Tabela 23).

FIGURA 24 - BIOMASSA DE FOLHAS OBSERVADA VS ESTIMADA, GRÁFICOS Q-Q E DISTRIBUIÇÃO RESIDUAL DAS MELHORES ESTRATÉGIAS DE MODELAGEM DE BIOMASSA DE FOLHAS DE *Mimosa scrabella* EM FUNÇÃO DO DIÂMETRO.

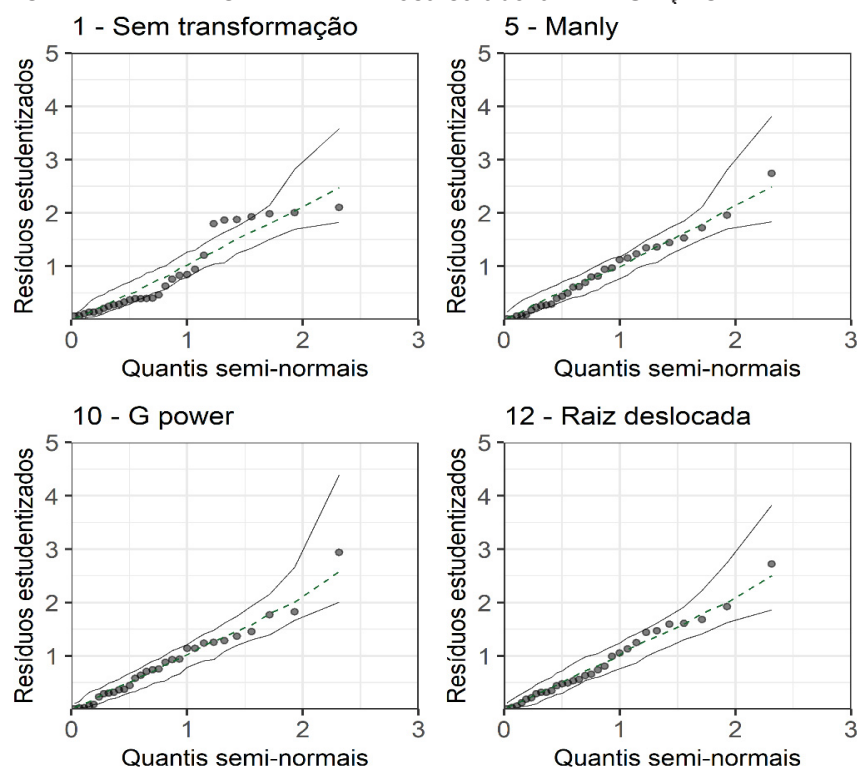


FONTE: O autor (2022).

De acordo com os gráficos hnp gerados, apenas os modelos com transformação foram adequados para predição. Nesse sentido, as modelagens com as estratégias 5, 10 e 12 promoveram resíduos dentro dos intervalos de confiança estabelecidos. Observações diferentes foram atribuídas aos resíduos gerados pela estratégia 1, os quais sugeriram inadequação preditiva (Figura 27).



FIGURA 25 - GRÁFICOS HNP (HALF-NORMAL PLOT) DAS MODELAGENS DA BIOMASSA DAS FOLHAS DE ÁRVORES DE *Mimosa scrabella* EM FUNÇÃO DA DIÂMETRO.



FONTE: O autor (2022).

Dessa forma, os coeficientes de regressão gerados pelos modelos selecionados apresentaram-se significativos a 1%, conforme o teste  $t$  (Tabela 24). Destaca-se que a equação gerada pelo ajuste da estratégia 1 está sujeita a inconfiabilidade, haja vista que os resíduos gerados na modelagem não se adequaram às suposições da análise paramétrica ( $p$ -Li e  $p$ -BP < 0,05).

TABELA 24 - COEFICIENTES DE REGRESSÃO E PARÂMETROS DE TRANSFORMAÇÃO DAS MELHORES ESTRATÉGIAS DE MODELAGEM DA BIOMASSA DAS FOLHAS DE *Mimosa scrabella* EM FUNÇÃO DA ALTURA E DO DIÂMETRO.

Modelo	Estratégia	Transformação	$\beta_0$	SE $\beta_0$ %	$\beta_1$	SE $\beta_1$ %	$\lambda$
$w_{folhas}$ (kg) em função da altura	1	Sem transformação	-2,7745438**	41,9743	0,44147**	19,9787	
	5	Manly	-0,7429238**	69,4686	0,19949**	19,6000	-0,2329
	10	G power	-0,3478095**	118,0436	0,15229**	20,4216	0,0520
	12	Raiz deslocada	0,4127596**	69,6287	0,11036**	19,7535	0,7900
$w_{folhas}$ (kg) em função do diâmetro	1	Sem transformação	-1,4353399**	29,8159	0,380616**	9,0117	
	5	Manly	-0,3369358**	68,7259	0,208761**	8,8618	-0,1605
	10	G power	-0,2079230**	110,2697	0,202158**	9,1018	0,3070
	12	Raiz deslocada	0,7527564**	14,0005	0,094680**	8,8720	0,7900

em que  $\beta_0$  e  $\beta_1$  são os coeficientes de regressão do modelo; \*\* significativo ao nível 1% de probabilidade; \* significativo ao nível de 5% de probabilidade; <sup>ns</sup> não significativo; SE $\beta_n$ % é o erro padrão percentual dos coeficientes estimados; e  $\lambda$  é o coeficiente de transformação.

FONTE: O autor (2022).

## 5.9 ESTRATÉGIAS DE MODELAGEM DA BIOMASSA DA CASCA DE ÁRVORES DE *Mimosa scrabella*

Nas modelagens da biomassa da casca ( $w_{casca}$ ), a maioria das estratégias de ajuste estudada forneceram bons resultados. Sobretudo, destaca-se que, para as regressões cujas variáveis não estiveram sob efeito de transformações, violou-se o atendimento de pelo menos um dos pressupostos requeridos (Tabela 25). Dessa forma, o uso de transformações no processo de modelagem da  $w_{casca}$  mostrou-se necessário.

Nessa perspectiva, nos modelos lineares em que a altura esteve como variável independente, seis estratégias demonstraram a inadequação de pelo menos uma das premissas da análise de regressão, de modo que em cinco ocasiões a premissa violada foi a homocedasticidade das variâncias. Em contrapartida, nos modelos lineares cujo diâmetro foi a variável independente, sete estratégias infringiram o atendimento dos pressupostos requeridos, dentre as quais quatro violaram exclusivamente o pressuposto de homocedasticidade das variâncias.

Acerca dessas informações, destaca-se que independentemente da variável explicativa utilizada nas modelagens, nenhuma estratégia de ajuste violou as premissas de normalidade residual e homogeneidade das variâncias simultaneamente.

As estratégias de transformação que forneceram o atendimento simultâneo dos pressupostos exigidos para a análise de regressão linear foram a 2, 5, 7, 8, 10, 11, 13 e 15, para os modelos em que a variável independente foi o diâmetro; e 2, 4, 5, 6, 7, 8, 9, 10 e 11, para os modelos em que a variável independente foi a altura. Todavia, dentre essas estratégias bem-sucedidas de ajuste, destacaram-se estatisticamente as estratégias 7, 10 e 13 para os modelos com o diâmetro e as estratégias 4, 8 e 9 para aqueles com a altura.

Assim, os ajustes com essas estratégias, além da adequação conjunta dos pressupostos ao nível de 1% de probabilidade ( $p-Li$  e  $p-BP > 0,01$ ), promoveram melhoras consideráveis nas estatísticas de ajuste e precisão. Entretanto, nas modelagens em que a  $w_{casca}$  esteve relacionada com a altura, houve perda de precisão dos modelos envolvendo transformação. Nessas

ocasiões, os valores percentuais de SEE demonstraram aumento em relação a estratégia sem transformação (Tabela 25).

TABELA 25 - ESTATÍSTICAS DE REGRESSÃO DAS MELHORES ESTRATÉGIAS DE MODELAGEM DA BIOMASSA DE CASCA DE *Mimosa scrabella* EM FUNÇÃO DA ALTURA E DO DIÂMETRO.

Modelo	Estratégia	Transformação	Retransformação	R <sup>2</sup> aj	SEE%	BIC	p-Li	p-BP
<i>W<sub>casca</sub></i> (kg) em função da altura	1	Sem transformação	$y = Y$	0,6888	64,21	173,61	0,0936	0,0013
	4	Box-Cox	$y = (\lambda \cdot y' + 1)^{\frac{1}{\lambda}}$	0,8754	79,23	186,22	0,3116	0,1829
	8	Glog	$y = -\frac{[1 - (\exp^{y'})^2]}{2\exp^{y'}}$	0,8685	70,84	179,50	0,1172	0,0695
	9	Dual power	$y = \left[ \sqrt{1 + \lambda^2 y'^2} + \lambda y' \right]^{\frac{1}{\lambda}}$	0,8743	72,50	180,90	0,2315	0,0945
<i>W<sub>casca</sub></i> (kg) em função do diâmetro	1	Sem transformação	$y = Y$	0,8822	39,51	144,48	0,3081	0,0003
	7	Yeo-Johnson	$y = (\lambda \cdot y' + 1)^{\frac{1}{\lambda}} - 1$	0,9374	35,34	137,78	0,1237	0,2587
	10	G power	$y = -\left[ \frac{1 - (\lambda y' + 1)^{\frac{2}{\lambda}}}{2(\lambda y' + 1)^{\frac{1}{\lambda}}} \right]$	0,9370	35,43	137,94	0,1288	0,2804
	13	Log-Log	$y' = \log(Y)$	0,9155	33,06	133,79	0,3237	0,1589

em que R<sup>2</sup>aj é o coeficiente de determinação ajustado do modelo; SEE é o erro padrão da estimativa; BIC é o critério de informação Bayesiano; p-BP é o p-valor do teste de homocedasticidade de Breusch-Pagan; e p-Li é o p-valor do teste de normalidade de Lilliefors.

FONTE: O autor (2022).

No entanto, pelas estatísticas de R<sup>2</sup>aj de cada estratégia, observou-se que, com a incorporação de transformações nas variáveis dos modelos, ocorreram melhoras no potencial de explicação de *w<sub>casca</sub>*. Essas percepções foram ainda mais evidentes quando a *w<sub>casca</sub>* foi ajustada em função do diâmetro, pois o R<sup>2</sup>aj do ajuste sem transformação oscilou de 0,88 para aproximadamente 0,94 com a aplicação de alguma estratégia de transformação. Destaca-se, ainda, que melhorias semelhantes estiveram relacionadas às estatísticas de SEE% e BIC dos modelos (Tabela 25).

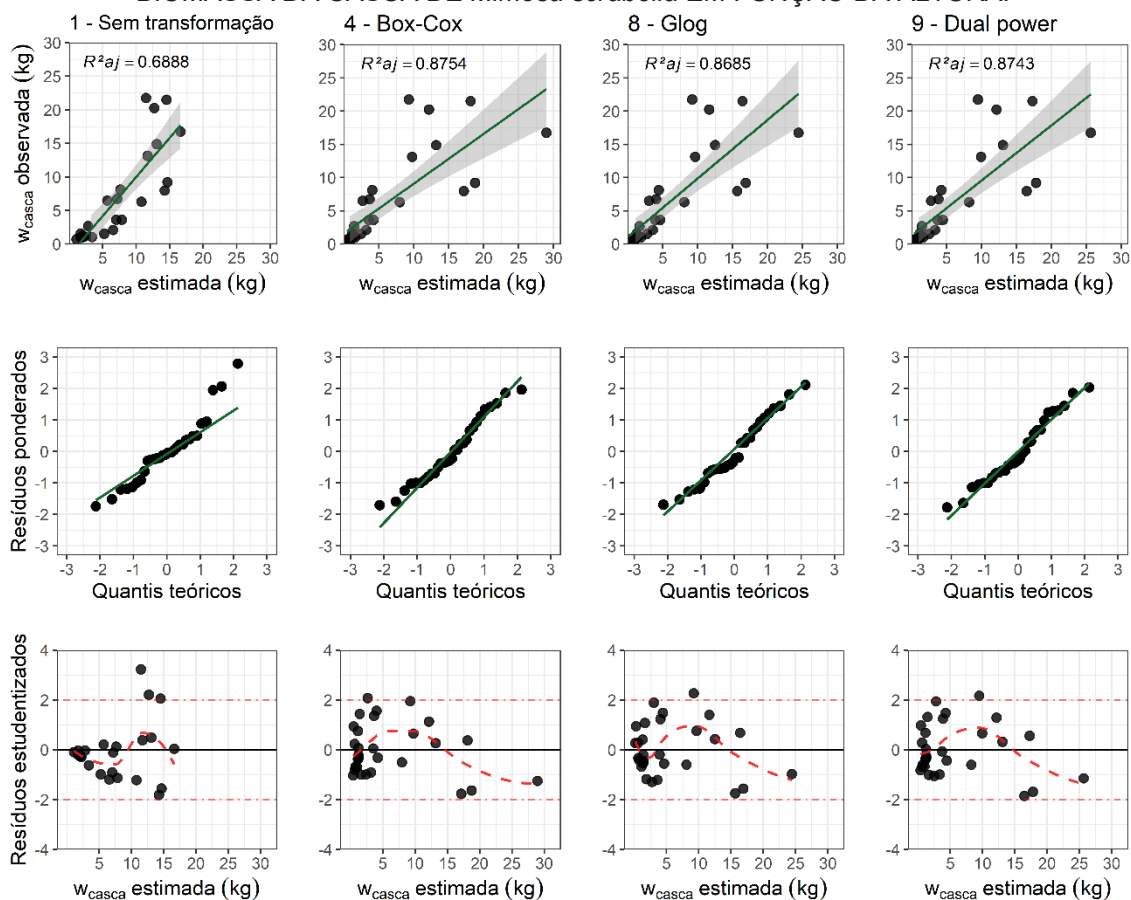
As análises gráficas dos ajustes, por sua vez, retrataram as mesmas evidências obtidas pelas estatísticas de regressão (Figura 28). Entretanto, a distribuição dos quantis teórico no gráfico Q-Q demonstrou uma disposição distorcida para os quantis positivos do ajuste do modelo envolvendo a estratégia 1. Porém, essas moderadas distorções nos quantis superiores não caracterizaram violação da suposição gaussiana para o termo de erro dessa modelagem (*p-Li* = 0,0936) (Tabela 25).

Os resíduos gerados pelos modelos selecionados envolvendo estratégias de transformações demonstraram distribuições uniformes e ausência

de heterocedasticidade. Entretanto, com relação a modelagem da estratégia 1, na qual o emprego das variáveis ocorreu na escala original, a distribuição residual demonstrou uma propensão a heterocedasticidade. Destaca-se que essas evidências gráficas acerca da homocedasticidade foi comprovada pelo teste de hipótese de Breusch-Pagan ao nível 1% de probabilidade.

Dessa forma, as técnicas de transformações de Box-Cox, Glog e Dual power (estratégias 4, 8 e 9, respectivamente) mostraram-se eficazes quando direcionadas à adequação da homocedasticidade das variâncias. No entanto, enfatiza-se que, diante das projeções alométricas, essas metodologias de transformação demonstraram tendências em superestimar os maiores valores de  $w_{casca}$  (Figure 28). Isso ocorreu devido ao esforço das transformações em equalizarem as dispersões mais proeminentes dos pares de dados, com o propósito de corrigir a vigente violação da homocedasticidade das variâncias.

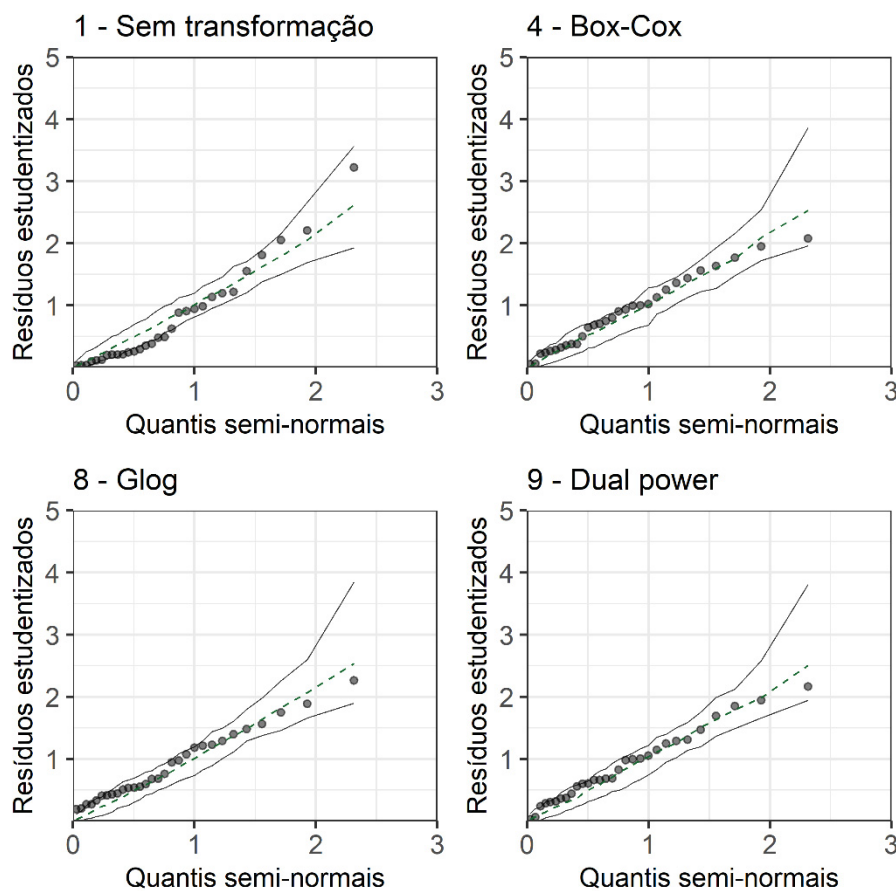
FIGURA 26 - BIOMASSA DA CASCA OBSERVADA VS ESTIMADA, GRÁFICOS Q-Q E DISTRIBUIÇÃO RESIDUAL E DAS MELHORES ESTRATÉGIAS DE MODELAGEM DE BIOMASSA DA CASCA DE *Mimosa scrabella* EM FUNÇÃO DA ALTURA.



FONTE: O autor (2022).

Contudo, os gráficos hnp demonstraram que o uso das transformações promoveu melhorias nas estimativas da  $w_{casca}$ . Verificou-se, portanto, para a estratégia 1, que alguns resíduos relacionados aos menores semi-quantis ultrapassaram o limite inferior do envelope de confiança, desqualificando o uso de variáveis em escala original para fins preditivos. Enquanto isso, os comportamentos dos resíduos gerados nas modelagens envolvendo as transformações Box-Cox, Glog e Dual power foram semelhantes (Figura 29).

FIGURA 27 - GRÁFICOS HNP (HALF-NORMAL PLOT) DAS MODELAGENS DA BIOMASSA DA CASCA DE ÁRVORES DE *Mimosa scrabella* EM FUNÇÃO DA ALTURA.

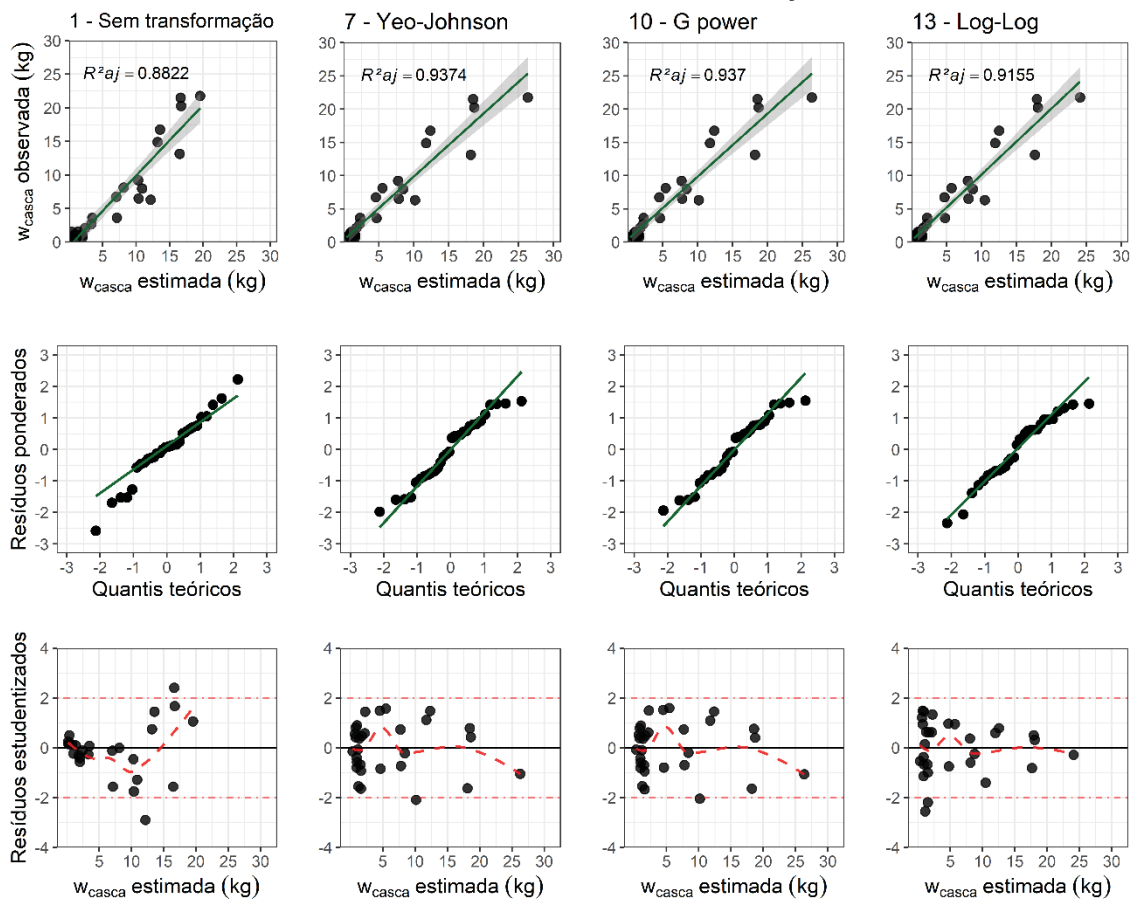


FONTE: O autor (2022).

Em contrapartida, diante da abordagem alométrica da  $w_{casca}$  em função do diâmetro, as estruturas residuais das modelagens demonstraram normalidade para todas as estratégias em estudo. Portanto, pela análise do gráfico Q-Q, inferiu-se que os quantis teóricos se distribuíram linearmente, demonstrando adequação à distribuição normal dos resíduos (Figura 30).

Nessa perspectiva, apenas as dispersões dos resíduos estudentizados evidenciaram diferença entre as estratégias, demonstrando variâncias não constantes para a estratégia 1 (sem transformação). A heterocedasticidade associada ao ajuste dessa estratégia foi comprovada pelo teste de Breusch-Pagan a 1% de probabilidade (Tabela 25).

FIGURA 28 - BIOMASSA DA CASCA OBSERVADA VS ESTIMADA, GRÁFICOS Q-Q E DISTRIBUIÇÃO RESIDUAL DAS MELHORES ESTRATÉGIAS DE MODELAGEM DE BIOMASSA DA CASCA DE *Mimosa scrabella* EM FUNÇÃO DO DIÂMETRO.

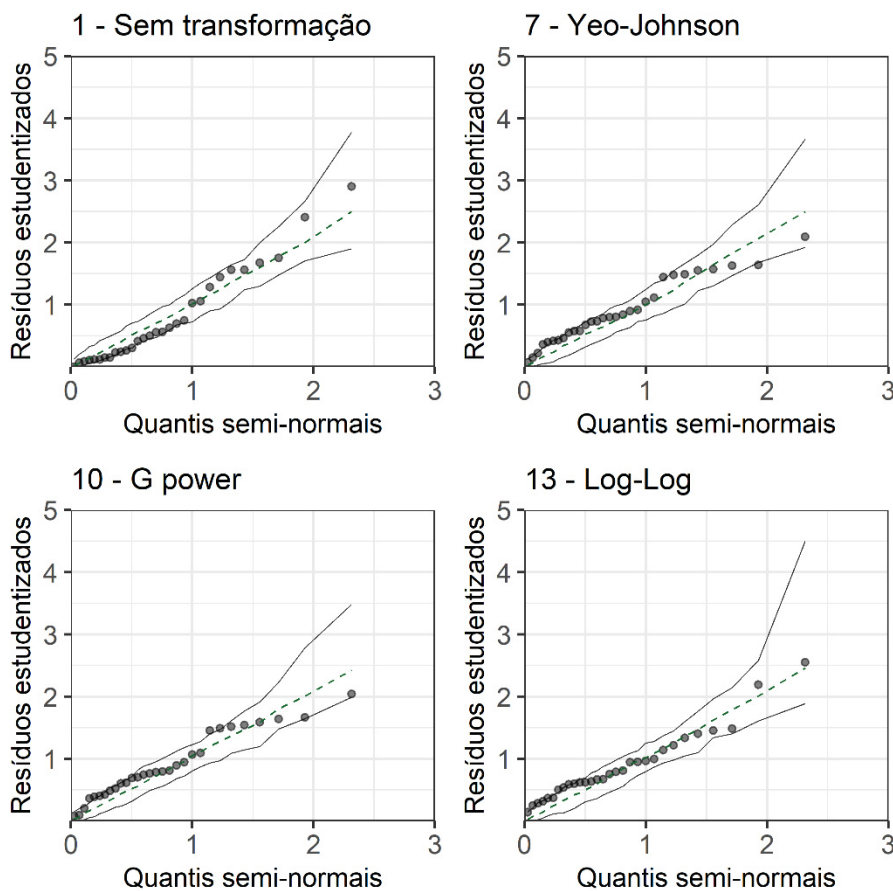


FONTE: O autor (2022).

Observa-se pela distribuição dos resíduos estudentizados que apenas a estratégia 1 apresentou dispersões não constantes e heterocedásticas. Em suma, essa violação foi corrigida quando aplicadas as transformações Yeo-Johnson, G power e Log-Log (Figura 16). Enfatiza-se, ainda, que a aplicação dessas transformações também promoveu melhorias às distribuições dos resíduos, tornando-os mais propensos a distribuição gaussiana, conforme observa-se na distribuição dos quantis teóricos dos gráficos Q-Q (Figura 30).

Contudo, conforme verificado nos gráficos hnp, a modelagem da estratégia 1 (sem transformação) apresentou qualidade preditiva aceitável, equiparando-se com as estratégias de modelagem envolvendo transformações. Destaca-se, porém, que todas as estratégias selecionadas (Yeo-Johnson, Gpower, Log-Log) resultaram em resíduos próximos aos limites dos envelopes de confiança (Figura 31).

FIGURA 29 - GRÁFICOS HNP (HALF-NORMAL PLOT) DAS MODELAGENS DA BIOMASSA DA CASCA DE ÁRVORES DE *Mimosa scrabella* EM FUNÇÃO DO DIÂMETRO.



FONTE: O autor (2022).

Salienta-se que os coeficientes de regressão gerados por cada estratégia, foram significativos a 1% pelo teste  $t$  em todas as ocasiões de modelagens (Tabela 26). Com base nessa análise, assume-se, portanto, que as equações geradas possuem coeficientes de regressão coerentes para a análise alométrica da  $w_{cascas}$ . Todavia, a significância dos coeficientes de regressão gerados pela modelagem da estratégia 1 não foi suscetível à confiabilidade

estatística, pois não houve o cumprimento da premissa de homocedasticidade das variâncias (Tabela 25).

TABELA 26 - COEFICIENTES DE REGRESSÃO E PARÂMETROS DE TRANSFORMAÇÃO DAS MELHORES ESTRATÉGIAS DE MODELAGEM DA BIOMASSA DA CASCA DE *Mimosa scrabella* EM FUNÇÃO DA ALTURA E DO DIÂMETRO.

Modelo	Estratégia	Transformação	$\beta_0$	SE $\beta_0$ %	$\beta_1$	SE $\beta_1$ %	$\lambda$
$w_{cascas}$ (kg) em função da altura	1	Sem transformação	-13,876503**	18,3894	1,5598161**	12,3861	
	4	Box-Cox	-3,1560537**	9,6038	0,3283957**	6,9733	-0,1833
	8	Glog	-1,8749737**	14,9655	0,2948265**	7,1907	
	9	Dual power	-3,3635313**	9,6625	0,3502630**	7,0233	0,2029
$w_{cascas}$ (kg) em função do diâmetro	1	Sem transformação	-6,9853035**	13,9536	1,1523030**	6,7716	
	7	Yeo-Johnson	-0,6176588**	20,4158	0,2104580**	4,7943	0,1427
	10	G power	-1,0780509**	17,9862	0,3227620**	4,8085	0,2215
	13	Log-Log	-4,8550872**	6,9886	2,5630377**	5,6328	

em que  $\beta_0$  e  $\beta_1$  são os coeficientes de regressão do modelo; \*\* significativo ao nível 1% de probabilidade, \* significativo ao nível de 5% de probabilidade; ns não significativo; SE $\beta_n$ % é o erro padrão percentual dos coeficientes estimados; e  $\lambda$  é o coeficiente de transformação.

FONTE: O autor (2022).

## 5.10 ESTRATÉGIAS DE MODELAGEM DA BIOMASSA DA RAIZ DE ÁRVORES DE *Mimosa scrabella*

Nas modelagens destinadas a predição da biomassa das raízes ( $w_{raízes}$ ), foi necessário recorrer ao emprego de transformações, haja vista que diante dos ajustes com as variáveis em escala original pelo menos um dos pressupostos da regressão linear foi violado. Assim, conforme o esperado, as modelagens com variáveis transformadas promoveram satisfatoriamente a adequação de normalidade dos resíduos e homocedasticidade das variâncias.

Entretanto, destaca-se que, nos modelos estruturados com a altura, tanto a normalidade residual quanto a homogeneidade das variâncias foram violadas no ajuste da estratégia 1 (sem transformação). Por outro lado, nos modelos em que se aplicou o diâmetro, apenas a homogeneidade das variâncias não foi atendida para essa estratégia (Tabela 27).

Dessa forma, nas modelagens em que a altura foi empregada como variável explicativa da  $w_{raízes}$ , as estratégias 5, 8 e 11 (Manly, Glog e Log shift, respectivamente) mostraram-se exitosas quanto às correções das suposições estatísticas requeridas. Todavia, as modelagens com essas transformações



promoveram maiores valores de SEE%, demonstrando acentuada perda de precisão nos ajustes (Tabela 27).

TABELA 27 - ESTATÍSTICAS DE REGRESSÃO DAS MELHORES ESTRATÉGIAS DE MODELAGEM DA BIOMASSA DAS RAÍZES DE *Mimosa scrabella* EM FUNÇÃO DA ALTURA E DO DIÂMETRO.

Modelo	Estratégia	Transformação	Retransformação	R <sup>2</sup> aj	SEE%	BIC	p-Li	p-BP
<i>W</i> <sub>raízes</sub> (kg) em função da altura	1	Sem transformação	$y = y$	0,5544	79,69	231,36	0,0023	0,0022
	5	Manly	$y = \frac{\log(\lambda \cdot y' + 1)}{\lambda}$	0,7526	96,62	242,92	0,8197	0,5823
	8	Glog	$y = -\frac{[1 - (\exp^{y'})^2]}{2\exp^{y'}}$	0,7304	89,03	238,02	0,4762	0,7515
	11	Log shift	$y = \exp^{y'} - \lambda$	0,7298	87,21	236,78	0,5731	0,7465
<i>W</i> <sub>raízes</sub> (kg) em função do diâmetro	1	Sem transformação	$y = y$	0,8892	39,74	189,61	0,4783	0,0001
	4	Box-Cox	$y = (\lambda \cdot y' + 1)^{\frac{1}{\lambda}}$	0,9775	23,14	157,17	0,5913	0,2364
	7	Yeo-Johnson	$y = (\lambda \cdot y' + 1)^{\frac{1}{\lambda}} - 1$	0,9790	23,73	158,67	0,3926	0,1446
	9	Dual power	$y = \left[ \sqrt{1 + \lambda^2 y'^2} + \lambda y' \right]^{\frac{1}{\lambda}}$	0,9745	22,99	156,78	0,885	0,4249

em que R<sup>2</sup>aj é o coeficiente de determinação ajustado do modelo; SEE é o erro padrão da estimativa; BIC é o critério de informação Bayesiano; p-BP é o p-valor do teste de homocedasticidade de Breusch-Pagan; e p-Li é o p-valor do teste de normalidade de Lilliefors.

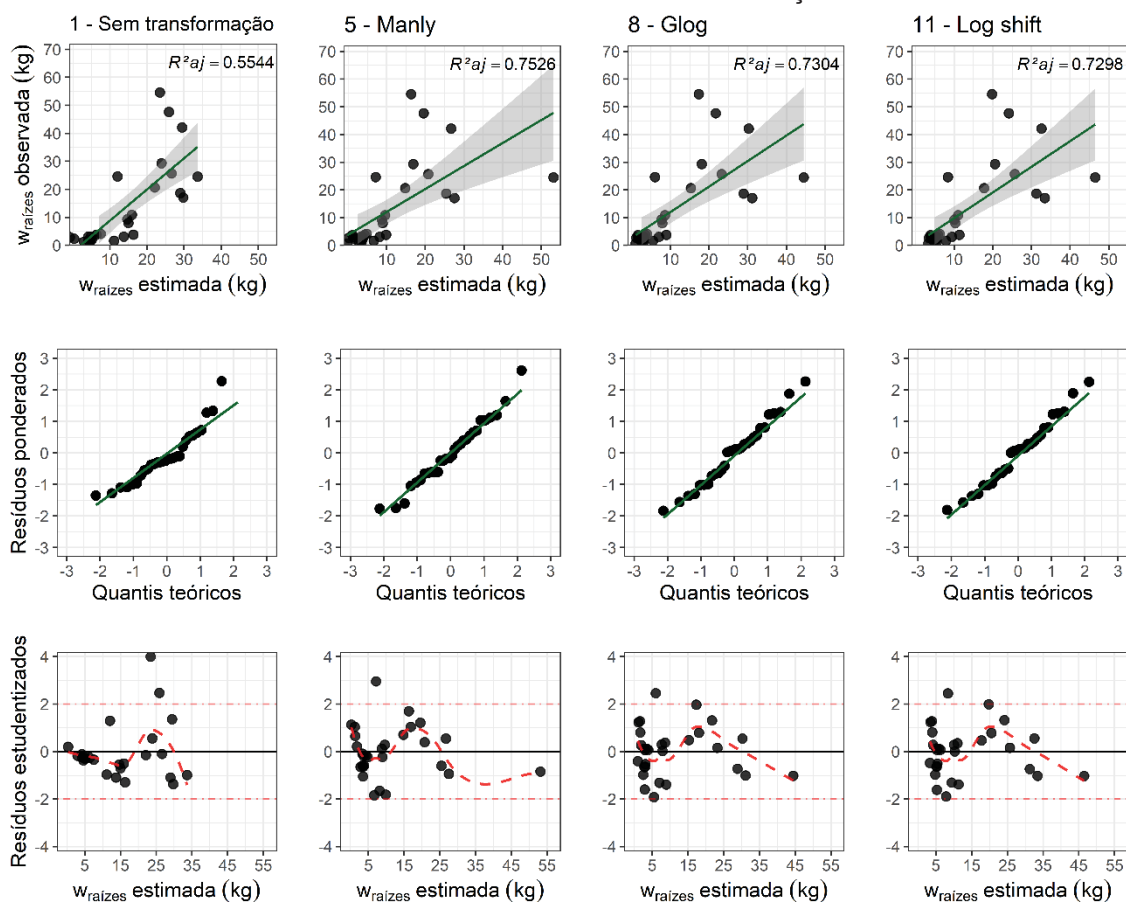
FONTE: O autor (2022).

Por outro lado, nos modelos constituídos com o diâmetro como preditor da *w*<sub>raízes</sub>, seis técnicas de transformação não foram capazes de alcançar a adequação simultânea da homocedasticidade e normalidade dos resíduos, especificamente as estratégias 2, 3, 8, 12, 14 e 15. Salieta-se, portanto, que nesse cenário, as estratégias que envolveram o parâmetro de transformação  $\lambda$  promoveram melhores modelagens, uma vez que atenderam todas as premissas exigidas à análise por regressão linear e melhoraram a precisão e o poder de explicação dos modelos. Dessa forma, sobressaíram-se estatisticamente as estratégias de ajuste 4, 7 e 9, ou seja, aquelas cujas técnicas de transformação foram: Box-Cox, Yeo-Johnson e Dual power, respectivamente (Tabela 27).

Nos modelos lineares simples em que a altura foi usada como variável independente, observou-se que, diante do uso de transformações, os valores de *w*<sub>raízes</sub> foram superestimados. Assim, no esforço das transformações para adequação dos pressupostos, houve um visível aumento na dispersão entre os pares de dados de *w*<sub>raízes</sub> observada e estimada (Figura 32).

Esse comportamento explica o aumento dos SEE das estratégias envolvendo transformações. Apesar disso, com o uso das transformações Manly, Glog e Log shift, além da adequação das premissas de normalidade residual e homogeneidade das variâncias houveram melhoras no poder de explicação dos modelos.

FIGURA 30 - BIOMASSA DA RAIZ OBSERVADA VS ESTIMADA, GRÁFICOS Q-Q E DISTRIBUIÇÃO RESIDUAL DAS MELHORES ESTRATÉGIAS DE MODELAGEM DE BIOMASSA DA RAIZ DE *Mimosa scrabella* EM FUNÇÃO DA ALTURA.



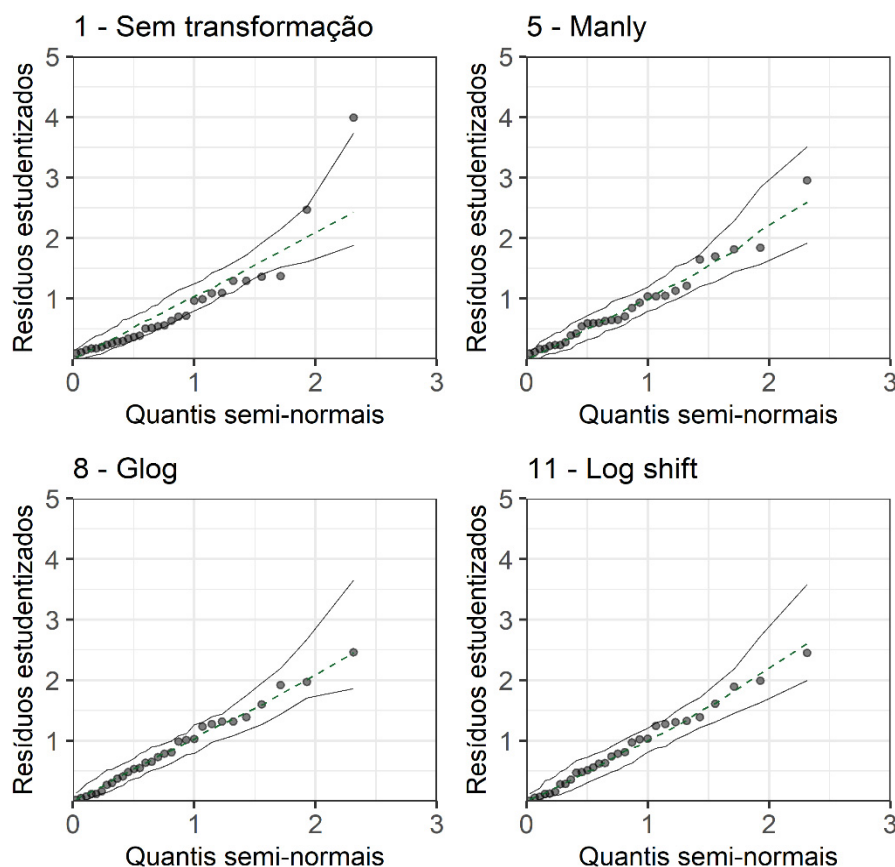
FONTE: O autor (2022).

Destaca-se que as análises gráficas das modelagens demonstraram as mesmas evidências constatadas pelos testes de hipótese descritos na Tabela 27. Nesse sentido, a partir do comportamento não linear dos quantis teóricos positivos no gráfico Q-Q, e a nítida presença de heterocedasticidade das variâncias no gráfico de resíduos estudentizados, infere-se que a modelagem sem transformação (estratégia 1) promove a violação tanto da premissa de normalidade residual quanto de homogeneidade das variâncias (Figura 32).

Entretanto, com base nos gráficos Q-Q e de distribuição de resíduos acerca dos ajustes das estratégias Manly, Glog e Log shift (5, 8 e 11, respectivamente), verifica-se que as premissas violadas no ajuste sem transformações (estratégia 1) são satisfatoriamente corrigidas. Essas inferências gráficas corroboram os resultados dos testes de Lilliefors e Breusch-Pagan descritos detalhados na Tabela 27.

Os gráficos hnp também demonstram essas evidências. Assim, as transformações Manly, Glog e Log shift forneceram melhores qualidades preditivas em relação à modelagem sem transformação (Figura 33). Essas percepções corroboraram com os resultados da Tabela 27 sobre a necessidade de transformar as variáveis para a modelagem da  $w_{raízes}$  em função da altura das árvores.

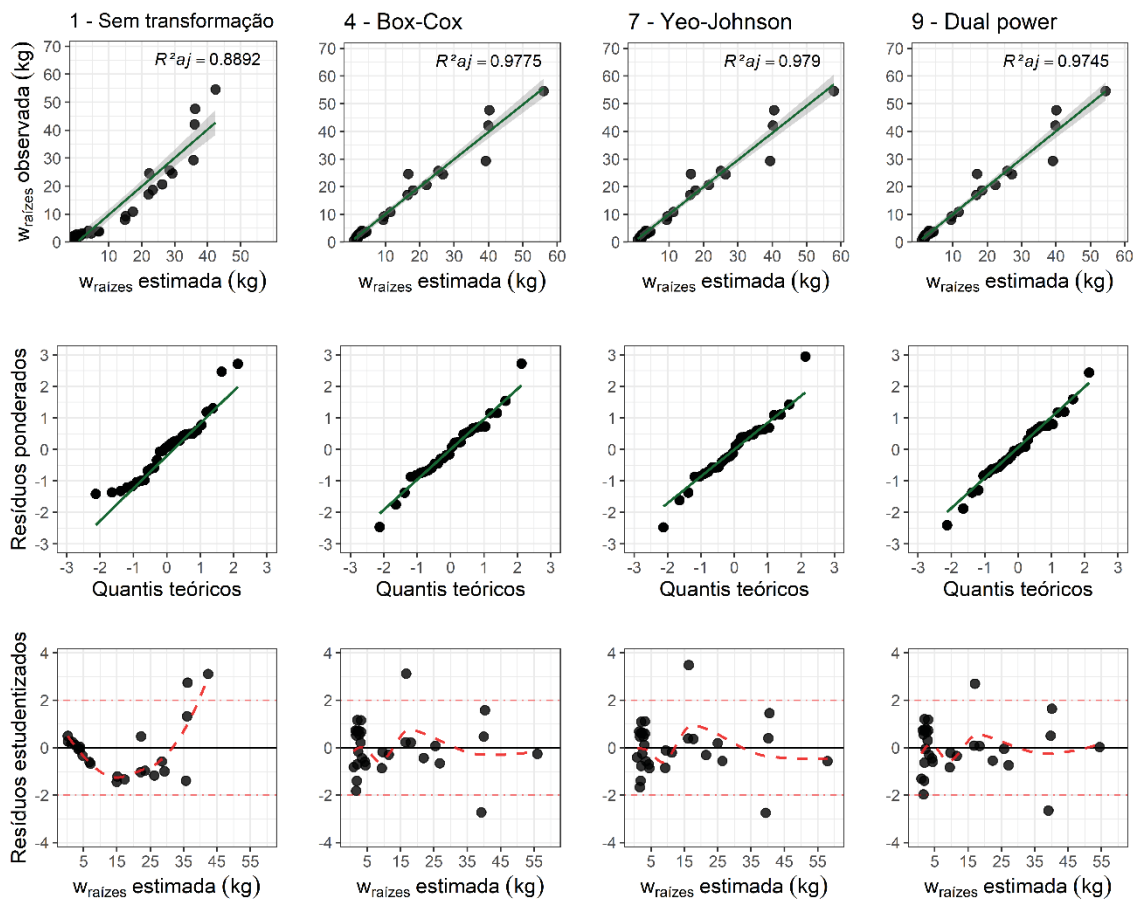
FIGURA 31 - GRÁFICOS HNP (HALF-NORMAL PLOT) DAS MODELAGENS DA BIOMASSA DAS RAÍZES DE *Mimosa scrabella* EM FUNÇÃO DA ALTURA.



FONTE: O autor (2022).

Percepções semelhantes são atribuídas nas regressões na qual considerou-se o diâmetro como preditor da  $w_{raízes}$ , pois houve melhoras promissoras nos termos de erro dos ajustes quando se aplicou transformações às modelagens (Figura 34).

FIGURA 32 - BIOMASSA DA RAIZ OBSERVADA VS ESTIMADA, GRÁFICOS Q-Q E DISTRIBUIÇÃO RESIDUAL DAS MELHORES ESTRATÉGIAS DE MODELAGEM DE BIOMASSA DA RAIZ DE *Mimosa scrabella* EM FUNÇÃO DO DIÂMETRO.

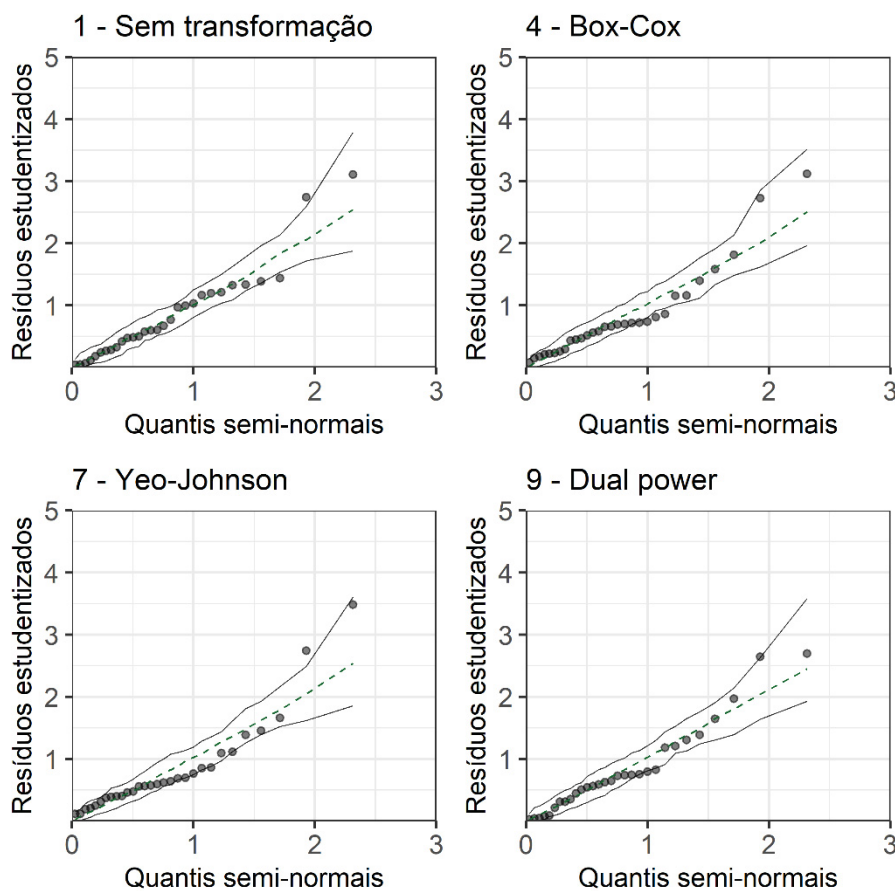


FONTE: O autor (2022).

Dessa forma, a proeminente presença de heterocedasticidade trazida pela modelagem da estratégia 1 foi corrigida com a aplicação das transformações de Box-Cox, Yeo-Johnson e Dual power aos modelos lineares. Assim, as evidentes distribuições inconstantes e enviesadas dos resíduos estudentizados em torno das abscissas foram satisfatoriamente contornadas com o uso dessas técnicas de transformação, conforme evidenciado na Figura 34.

Acerca da eficiência preditiva dessas estratégias, a transformação Dual power (9) aparentou melhor dispersão dos resíduos nos envelopes de confiança na análise dos gráficos hnp (Figura 35). Por outro lado, a ausência de transformação e os métodos Box-Cox e Yeo-Johnson forneceram resultados visualmente semelhantes.

FIGURA 33 - GRÁFICOS HNP (HALF-NORMAL PLOT) DAS MODELAGENS DA BIOMASSA DAS RAÍZES DE ÁRVORES DE *Mimosa scrabella* EM FUNÇÃO DO DIÂMETRO.



FONTE: O autor (2022).

Para todas as estratégias de ajuste selecionadas, os coeficientes de regressão das equações geradas foram significativos de acordo como teste  $t$  a 5% (Tabela 28). Todavia, evidencia-se que a credibilidade desse teste se estende apenas aos ajustes das estratégias de modelagem que envolveram transformações 5, 8, 11, 4, 7 e 9, haja vista que somente nesses cenários as suposições de normalidade residual e homocedasticidade das variâncias foram contempladas.

TABELA 28 - COEFICIENTES DE REGRESSÃO E PARÂMETROS DE TRANSFORMAÇÃO DAS MELHORES ESTRATÉGIAS DE MODELAGEM DA BIOMASSA DAS RAÍZES DE *Mimosa scrabella* EM FUNÇÃO DA ALTURA E DO DIÂMETRO.

Modelo	Estratégia	Transformação	$\beta_0$	SE $\beta_0$ %	$\beta_1$	SE $\beta_1$ %	$\lambda$
$W_{raízes}$ (kg) em função da altura	1	Sem transformação	-26,6047549**	25,1117	3,080548**	16,4224	
	5	Manly	-7,0830978**	19,7428	1,000153**	10,588	-0,7899
	8	Glog	-1,0860222*	38,8666	0,285118**	11,2234	
	11	Log shift	-1,6334117**	25,2049	0,277672**	11,2363	-2,0009
$W_{raízes}$ (kg) em função do diâmetro	1	Sem transformação	-15,86040682**	13,0381	2,531145**	6,5385	
	4	Box-Cox	-1,866822478**	7,7886	0,412953**	2,8090	0,2901
	7	Yeo-Johnson	-0,599850083**	16,2541	0,287249**	2,7154	0,1796
	9	Dual power	-1,452340694**	8,6274	0,334187**	2,9923	0,4267

em que  $\beta_0$  e  $\beta_1$  são os coeficientes de regressão do modelo; \*\* significativo ao nível 1% de probabilidade; \* significativo ao nível de 5% de probabilidade; ns não significativo; SE $\beta_n$ % é o erro padrão percentual dos coeficientes estimados; e  $\lambda$  é o coeficiente de transformação.

FONTE: O autor (2022).

## 6 DISCUSSÃO

Em florestas plantadas ou naturais, as relações entre as variáveis dendrométricas geralmente demonstram padrão não-linear (PAYANDEH, 1983; FELDPAUSCH et al., 2012; KEARSLEY et al., 2013; DUAN et al., 2018). Entretanto, o comportamento linear entre as variáveis modeladas é fundamental para obtenção de bons ajustes em abordagens alométricas envolvendo regressão linear (DRAPER; SMITH, 1998; LONG; ERVIN, 2000; PRETZSCH, 2010; WEISBERG, 2014; DUNN; SMYTH, 2018).

Diante disso, destaca-se que a técnica de transformação, além da adequação de homocedasticidade das variâncias e normalidade residual, pode ser empregada para a correção de eventual má especificação linear entre  $Y$  e  $X$  em estudos dendrométricos (VANCLAY, 1994; PORTÉ; BARTELINK, 2002; PRETZSCH, 2010). Nessa perspectiva, enfatiza-se, que, diante das bases de dados de biomassa, volume e altura, as transformações selecionadas a partir dos pressupostos de normalidade residual e homogeneidade das variâncias também corrigiram a falta de especificação linear entre as variáveis de interesse e explicativa.

Dentre as variáveis dendrométricas mais importantes à ciência florestal, destaca-se a altura das árvores. Essa variável é necessária para estimar volumes, que, por sua vez, são usados para derivar outras variáveis, como

volume comercial, biomassa, conteúdo de carbono e valor econômico (KERSHAW, et al., 2016; MACPHEE et al., 2018; NASCIMENTO et al., 2020). Nesse contexto, observou-se que, em florestas plantadas jovens de *P. oocarpa* aos 5 anos, a modelagem linear simples demonstrou bons ajustes (Tabela 7) e atendeu às suposições paramétricas. Diante disso, infere-se que, para povoamentos jovens, o emprego de transformações das variáveis não seja recomendado, haja vista que não há a violação dos pressupostos. Por outro lado, em povoamentos florestais tardios, a regressão linear não se mostra uma ferramenta satisfatório, visto que não há relação significativa entre altura e diâmetro sob a idade avaliada de 19 anos.

Contudo, a aplicação de técnicas de transformações em modelos hipsométricos, apesar de não ser necessária, agrega ganhos estatísticos às modelagens. Assim, com base nesse propósito, destacaram-se as metodologias de transformações de Manly, Yeo-Johnson e Box-Cox. Porém, devido a limitação relacionada a geração de eventuais erros ao se retransformar a variável para sua escala original somada a perda de interpretação biológica entre as variáveis (PRETZSCH, 2010), não se indica o uso de transformação em modelos hipsométricos quando se pensa, unicamente, em benefícios relacionados às estatísticas do ajuste.

Sob outra perspectiva, a análise volumétrica efetuada no presente estudo, cujas bases de dados representaram florestas plantadas e naturais, demonstrou propensão a violação de pelo menos um dos pressupostos da análise de regressão linear. Enfatiza-se, porém, que esse é o comportamento esperado às variáveis dendrométricas mais importantes (BEHLING et al., 2019; FIORENTIN, 2020; FIORENTIN et al., 2020), o que representa uma preocupação recorrente em estudos florestais.

Diante disso, para adequação da normalidade residual e homogeneidade das variâncias em modelagens volumétricas de povoamentos florestais sob idades avançadas, encoraja-se a aplicação das transformações de base logarítmica, seja simultaneamente nas variáveis resposta e explicativa ou apenas na resposta. Por outro lado, diante de povoamentos jovens, além da transformação logarítmica, pode-se recomendar, também, o uso da transformação potencial de Box-Cox. Para a modelagem do volume comercial em florestas naturais, incentiva-se também a aplicação das transformações

logarítmica e Box-Cox. Todavia, salienta-se que a adoção dessas técnicas não deve ser aplicada precipitadamente, pois, a depender da base de dados, elas nem sempre serão promissoras ao atendimento simultâneo das suposições requeridas.

Com relação as modelagens lineares destinadas às estimativas de biomassa, uma gama variada de estratégias de transformações surtiu os efeitos almejados no que diz respeito a geração de equações apropriadas para fins alométricos. Enfatiza-se que, para estimativas de biomassa de qualquer um dos componentes arbóreos por meio de modelos lineares simples, recomenda-se a adoção do diâmetro em detrimento à altura como variável explicativa, visto que, em todos os casos, demonstrou maior poder explicativo, além de gerar estatísticas mais promissoras.

Nesse segmento, argumenta-se que equações estatisticamente não confiáveis foram publicadas como modelos alométricos. Diante disso, Sileshi (2014) relacionou a esses modelos problemas estatísticos comuns associados à modelagem, dentre os quais destaca-se a estimativa de coeficientes de regressão não confiáveis, que está intrinsecamente relacionada a heterocedasticidade. Essas considerações estatísticas levaram o autor a concluir que modelos alométricos simples baseados em um único preditor (normalmente o diâmetro) devem ser preferidos aos modelos com vários preditores (como diâmetro e altura); assim como defendido também por Picard et al. (2015), Forrester et al. (2017), Santos et al. (2020) e Aneseyee et al. (2021). Por outro lado, há, ao mesmo tempo, um direcionamento de raciocínio na literatura que defende a importância das relações altura-diâmetro como promissoras à alometria de biomassa, conforme enfatizado nos estudos de Temesgen et al. (2011), Poudel et al. (2015) e Temesgen et al. (2015).

Para além dessa polarização de ideias, no presente estudo, as abordagens alométricas da biomassa, por modelagem linear simples, obtiveram melhores estatísticas de ajuste e precisão com o diâmetro como preditor. Assim, em uma perspectiva na qual se visa a adequação das suposições da análise de regressão linear e ganhos em performance estatística, encoraja-se a aplicação da transformação logarítmica tanto em  $\mathcal{Y}$  quanto em  $\mathcal{X}$  para os modelos de predição das biomassas total e dos galhos. Além disso, em uma perspectiva alométrica das biomassas de fuste e raízes, recomenda-se o uso do método de



transformação potencial de Box-Cox. Ademais, para a alometria de biomassas de folhas e casca incentiva-se a adoção, respectiva, das estratégias de Manly e Yeo-Johnson.

Reforça-se que a adequação dos pressupostos de distribuição normal dos erros e de homocedasticidade das variâncias são imprescindíveis para obtenção de estimativas, especialmente em modelagens lineares baseadas sob a abordagem do método de mínimos quadrados ordinários (FOX, 2016). Dessa forma, por influenciar diretamente na estimação dos coeficientes da equação, o pressuposto de homocedasticidade foi amplamente discutido na literatura, de modo que, atualmente, são inúmeros os mecanismos disponíveis para a sua adequação. Dentre esses métodos, destaca-se o uso de matrizes de covariância consistente com heterocedasticidade (LONG; ERVIN, 2000; CRIBARI-NETO; GOIS, 2002), métodos de mínimos quadrados ponderados (BELHING et al., 2018; ZEA-CAMAÑO et al., 2020; COUTINHO et al., 2021), equações não lineares (PICARD et al., 2015; BELHING et al., 2019), técnicas de mineração de dados (SANQUETTA et al., 2015; DEBASTIANI et al., 2019) e até regressões lineares generalizadas (FAHRMEIR; KAUFMANN, 1985; FOX, 2016; DUNN; SMYTH, 2018).

Dentre esses aspectos, diante de modelagens lineares com estratégias de transformações em que se corrige apenas a suposição de normalidade residual, seria possível a conciliação com outra metodologia paramétrica para a posterior adequação da homogeneidade das variâncias em regressão linear. Dessa forma, o ajuste por meio da regressão linear ponderada direcionado apenas a equalização das variâncias na modelagem tornar-se-ia uma alternativa factível como complemento ao uso de transformações.

Contudo, apesar de inúmeras alternativas disponíveis para abordar a violação das suposições paramétricas, sempre o uso de modelos lineares representará a alternativa mais conveniente ao analista de dados em atividades de predições alométricas. Dessa forma, o uso de transformações de variáveis, ao passo em que se direciona à correção das suposições paramétricas exigidas, representa uma possibilidade para a redução de esforços no âmbito da análise de regressão linear.

Nessa perspectiva, conhecimentos sobre a performance de técnicas de transformação em modelagens lineares para predições dendrométricas

representam um avanço importante à ciência florestal. Assim, foram abordadas neste estudo diversas estratégias de transformações para predições das variáveis dendrométricas mais importantes em atividades relacionadas a produção florestal: volume, altura e biomassa total e de seus componentes.

Não obstante, salienta-se que existem vários outros direcionamentos para a difusão dessas estratégias de transformação no âmbito da ciência florestal. Nesse sentido, a inclusão dessa ferramenta em modelagens para outros propósitos dendrométricos possibilitaria agregar novos conhecimentos, o que alcançaria novas opções para a sua implementação. Para tal fim, transformações de variáveis poderiam ser aplicadas, por exemplo, em modelagens de estimação de volumes de sortimento em plantios florestais, ajustes de funções de afilamento, estimativas de estoques de carbono em povoamentos, predição de variáveis morfológica, bem como para outras finalidades envolvendo recursos florestais.

Adicionalmente, ainda se destaca que novas frentes de estudo estariam relacionadas a averiguação dos efeitos da implementação de transformações em modelos lineares múltiplos. Ademais, ressalta-se a possibilidade de comparação do desempenho das transformações com outros métodos de regressão, tais como a regressão linear ponderada e a generalizada. Por último, evidencia-se a viabilidade de desenvolvimentos de novas estratégias de transformação por meio do emprego de técnicas de *machine learning*.

Portanto, essas perspectivas deixam em aberto o desafio, um tanto maior, para a integração de técnicas de transformação com a regressão linear para fins alométricos na ciência florestal. À vista disso, as limitações da análise de regressão linear poderão ser mitigadas, promovendo o aperfeiçoamento e a continuidade do uso de modelos lineares estatisticamente confiável para relações dendrométricas.

## 7 CONCLUSÕES

Na alometria de variáveis dendrométricas, recomenda-se a abordagem de transformação apenas em ocasiões na qual há a violação de pelo menos um dos pressupostos da análise de regressão linear. Diante disso, em estudos da relação altura-diâmetro em florestas plantadas jovens, não se encoraja o uso de estratégias de transformações de variáveis com o intuito apenas de ganhos relacionados às estatísticas da modelagem.

Em regressões lineares para predição volumétrica de florestas plantadas maduras, recomenda-se a aplicação do método de transformação logarítmico, seja simultaneamente na variável dependente e explicativa ou apenas na dependente. Por outro lado, quando se trata de povoamentos jovens e de florestas naturais, incentiva-se, além do método logarítmico, a transformação potencial de Box-Cox.

Diante das modelagens lineares para a alometria da biomassa, as transformações indicadas são a logarítmica quando se visa a predição da biomassa total e de galhos. Além disso, recomenda-se a transformação Box-Cox quando o objetivo é a predição da biomassa do fuste e de raízes, bem como Manly para folhas e Yeo-Johnson para galhos.

## REFERÊNCIAS

- ACZEL, A. D.; SOUNDERPANDIAN, J. **Complete business statistics**. 7. ed. McGraw-Hill/Irwin, 2009. 888 p.
- AKAIKE, H. A new look at the statistical model identification. **IEEE transactions on Automatic Control**, v. 19, n. 6, p. 716-723, 1974.
- ALEXOPOULOS, E. C. Introduction to multivariate regression analysis. **Hippokratia**, v. 14, n. 1, p. 23-28, 2010.
- ALTMAN, D. G. **Practical statistics for medical research**. Chapman & Hall/CRC, 1990. 623 p.
- ALVARES, C. A.; STAPE, J. L.; SENTELHAS, P. C.; GONÇALVES, J. L. M.; SPAROVEK, G. Koppen's climate classification map for Brazil. **Meteorologische Zeitschrift**, v. 22, n. 6, p. 711-728, 2013.
- ANESEYEE, A.; B.; SOROMESSA, T.; ELIAS, E.; FEYISA, G. L. Allometric equations for selected Acacia species (*Vachellia* and *Senegalia* genera) of Ethiopia. **Carbon balance and management**, v. 16, n. 1, p. 1-13, 2021.
- ATKINSON, A. C.; RIANI, M.; CORBELLINI, A. The Box–Cox Transformation: Review and Extensions. **Statistical Science**, v. 36, n. 2, p. 239-255, 2021.
- AVERY, T. E.; BURKHART, H. E. **Forest measurements**. 5. ed. Long Grove: Waveland Press, 2015. 456 p.
- BARTLETT, M. S. Properties of sufficiency and statistical tests. **Proceedings of the Royal Society of London. Series A-Mathematical and Physical Sciences**, v. 160, n. 901, p. 268-282, 1937.
- BASKERVILLE, G. L. Use of logarithmic regression in the estimation of plant biomass. **Canadian Journal of Forest Research**, v. 2, p. 49-53, 1972.
- BEAN, A.; XU, X.; MACEACHERN, S. Transformations and Bayesian density estimation. **Electronic Journal of Statistics**, v. 10, n. 2, p. 3355-3373, 2016.
- BEHLING, A.; PÉLLICO-NETTO, S.; SANQUETTA, C. R.; CORTE, A. P. D.; SIMON, A. A.; RODRIGUES, A. L.; CARON, B. O. Additive and non-additive biomass equations for black wattle. **Floresta e Ambiente**, v. 26, n.4, p. 1-10, 2019.
- BEHLING, A.; PÉLLICO-NETTO, S.; SANQUETTA, C. R.; CORTE, A. P. D.; AFFLECK, D. L. R.; RODRIGUES, A. L.; BEHLING, M. Critical analyses when modeling tree biomass to ensure additivity of its components. **Anais da Academia Brasileira de Ciências**, v. 90, n. 2, p. 1759-1774, 2018.
- BENNET, N. D.; CROKE, B. F. W.; GUARISO, G.; GUILLAUME, J. H. A.; HAMILTON, S. H.; JAKEMEN, A. J.; MARSILI-LIBELLI, S.; NEWHAM, L. T. H.; NORTON, J. P.; PERRIN, C.; PIERCE, S. A.; ROBSON, B.; SEPPELT, R.; VOINOV, A. A.; FATH, B. D.; ANDREASSIAN, V. Characterising performance of

environmental models. **Environmental Modelling & Software**, v. 40, p. 1-20, 2013.

BICKEL, P. J.; DOKSUM, K. A. An analysis of transformations revisited. **Journal of the American Statistical Association**, v. 76, n. 374, p. 296-311, 1981.

BODRO, D. K.; SARTONO, B.; SADIK, K. A simulation study with log, Box-Cox, and dual-power transformation on handling curvilinear relationship in small area estimation. In: **IOP Conference Series: Earth and Environmental Science**, IOP Publishing, 2019. v. 299, n. 012029. p. 1-9.

BOX, G. E. P; COX, D. R. An analysis of transformations. **Journal of the Royal Statistical Society**, v. 26, n. 2, p. 211-243, 1964.

BRASIL. Lei nº 9.985/2000, de 18 de julho de 2000. Regulamenta o art. 225, § 1º, incisos I, II, III e VII da Constituição Federal, institui o Sistema Nacional de Unidades de Conservação da Natureza e dá outras providências. **Diário Oficial da União**: Brasília, DF, 2000. Disponível em [http://www.planalto.gov.br/ccivil\\_03/leis/l9985.htm](http://www.planalto.gov.br/ccivil_03/leis/l9985.htm).

BREUSCH, T. S.; PAGAN, A. R. A simple test for heteroscedasticity and random coeficiente variation. **Econometrica**, v. 47, n. 5, p. 1287-1294, 1979.

BRUCE, A.; BRUCE, P. **Practical statistics for data scientists**. Alta Books, 2019. 320 p.

CAMPOS, J. C. C.; LEITE, H. G. **Mensuração florestal**: perguntas e respostas. Editora UFV, 2017. 636 p.

CARROLL, R. J.; RUPPERT, D. **Transformation and weighting in regression**. Chapman and Hall/CRC, 1988. 264 p.

CASSON, R. J.; FARMER, L. D.M. Understanding and checking the assumptions of linear regression: a primer for medical researchers. **Clinical & experimental ophthalmology**, v. 42, n. 6, p. 590-596, 2014..

CHARNET, R.; FREIRE, C. A. L.; CHARNET, E. M. R.; BONVINO, H. **Análise de Modelos de Regressão linear**: com aplicações. Editora Unicamp, 2008. 353 p.

CHINN, S. Choosing a transformation. **Journal of Applied Statistics**, v. 23, n. 4, p. 395-404, 1996.

CHORTIRAT, T.; CHOMTEE, B.; SINSOMBOONTHONG, J. Comparison of four data transformation methods for Weibull distributed data. **Agriculture and Natural Resources**, v.4 5, n. 2, p. 366-383, 2011.

COHEN, P.; WEST, S. G.; AIKEN, L. S. **Applied multiple regression/correlation analysis for the behavioral sciences**. Psychology press, 2014. 734 p.

COLE, T. G.; EWEL, J. J. Allometric equations for four valuable tropical tree species. **Forest Ecology and Management**, v. 229, n. 3, p. 351-360, 2006.

COUTINHO, V. M.; OLIVEIRA, T. W. G.; FIORENTIN, L. D.; SANQUETTA, M. N. I.; SANQUETTA, C. R.; CORTE, A. P. D. How to estimate black wattle aboveground biomass from heteroscedastic data? **Floresta**, v. 51, n. 1, p. 28-36, 2021.

CRIBARI-NETO, F.; GOIS, M. C. A. Uma análise de Monte Carlo do desempenho de estimadores de matrizes de covariância sob heterocedasticidade de forma desconhecida **Revista Brasileira de Economia**, v. 56, p. 309-334, 2002.

CUNIA, T.; BRIGGS, R. D. Forcing additivity of biomass tables: some empirical results. **Canadian Journal of Forest Research**, v. 14, p. 376-384, 1984.

CYSNEIROS, V. C. **Efeitos dos fatores ambientais na alometria das árvores em diferentes tipologias florestais da mata atlântica no estado do Rio de Janeiro**. 2020. 123 f. Tese (Doutorado em Engenharia Florestal) – Universidade Federal do Paraná, Curitiba, 2020.

CYSNEIROS, V. C.; GAUI, T. D.; SILVEIRA FILHO, T. B.; PELISSARI, A. L.; MACHADO, S. D. A.; DE CARVALHO, D. C.; AMORIM, H. B. Tree volume modeling for forest types in the Atlantic Forest: generic and specific models. **iForest Biogeosciences and Forestry**, v. 13, n. 5, p. 417-425, 2020a.

CYSNEIROS, V. C.; PELISSARI, A. L.; GAUI, T. D.; FIORENTIN, L. D.; CARVALHO, D. C. D.; SILVEIRA FILHO, T. B.; MACHADO, S. D. A. Modeling of tree height–diameter relationships in the Atlantic Forest: effect of forest type on tree allometry. **Canadian Journal of Forest Research**, v. 50, n. 12, p. 1289-1298, 2020b.

CYSNEIROS, V. C.; PELISSARI, A. L.; MACHADO, S. A.; FIGUEIREDO FILHO, A.; SOUZA, L. Modelos genéricos e específicos para estimativa do volume comercial em uma floresta sob concessão na Amazônia. **Scientia Forestalis**, v. 45, n. 114, p. 295-304, 2017.

DEBASTIANI, A. B.; SANQUETTA, C. R.; DALLA CORTE, A. P.; PINTO, N. S.; REX, F. E. Evaluating SAR-optical sensor fusion for aboveground biomass estimation in a Brazilian tropical forest. **Annals of Forest Research**, v. 62, n.2, p. 109-122, 2019.

DINIZ, E. S.; THIELE, J. **Modelos de regressão em R**. 2 ed. Clube de autores, 2021. 152 p.

DRAPER, N. R.; SMITH, H. **Applied regression analysis**. New York: John Wiley & Sons, 1998. 704 p.

DUAN, G.; GAO, Z.; WANG, Q.; FU, L. Comparison of different height–diameter modelling techniques for prediction of site productivity in natural uneven-aged pure stands. **Forests**, v.9, n. 2, p. 1-18, 2018.

DUNN, P. K.; SMYTH, G. K. **Generalized linear models with examples in R**. Springer, 2018. 573 p.

DURBIN, B. P.; HARDIN, J. S.; HAWKINS, D. M.; ROCKE, D. M. A variance-stabilizing transformation for gene-expression microarray data. **Bioinformatics**, v. 18, n. suppl\_1, p. S105-S110, 2002.

EASTAUGH, C.S.; HASENAUER, H. Incorporating management history into forest growth modelling. **iForest Biogeosciences and Forestry**, v. 4, p. 212-217, 2011.

FADINI, S. R. M.; BARBOSA, R. I.; RODE, R.; CORRÊA, V.; FADINI, R. F. Above-ground biomass estimation for a shrubby mistletoe in an Amazonian savanna. **Journal of Tropical Ecology**, v. 36, n. 1, p. 6-12, 2020.

FAHRMEIR, L.; KAUFMANN, H. Consistency and asymptotic normality of the maximum likelihood in Generalized Linear Models. **The Annals of Statistics**, v.13, n.1, p. 342-368, 1985.

FAYOLLE, A.; NGOMANDA, A.; MBASI, M.; BARBIER, N.; BOCKO, Y.; BOYEMBA, F.; COUTERON, P.; FONTON, N.; KAMDEM, N.; KATEMBO, J.; KONDAOULE, H. J.; LOUMETO, J.; MAIDOU, H. M.; MANKOU, G.; MENGUI, T.; MOFACK, G.; MOUNDOUNGA, C.; MOUNDOUNGA, Q.; NGUIMBOUS, L.; NCHAMA, N. N.; OBIANG, D.; ASUE, F. O. M.; PICARD, N.; ROSSI, V.; SENGUELA, Y.; SONKÉ, B.; VIARD, L.; YONGO, O. D.; ZAPFACK, L.; MEDJIBE, V. P. A regional allometry for the Congo basin forests based on the largest ever-destructive sampling. **Forest ecology and management**, v. 430, p. 228-240, 2018.

FELDPAUSCH, T. R.; LLOYD, J.; LEWIS, S. L.; BRIENEN, R. J.; GLOOR, M.; MONTEAGUDO MENDOZA, A.; PHILLIPS, O. L. Tree height integrated into pantropical forest biomass estimates. **Biogeosciences**, v. 9, n. 8, p. 3381-3403, 2012.

FENG, C.; WANG, H.; LU, N.; CHEN, T.; HE, H.; LU, Y.; TU, X. M. Log-transformation and its implications for data analysis. **Biostatistics in Psychiatry**, v. 26, n. 2, p. 105-111, 2014.

FENG, Q.; HANNIG, J.; MARRON, J. S. A note on automatic data transformation. **Stat**, v. 5, n. 1, p. 82-87, 2016.

FIFE, Dustin. The eight steps of data analysis: A graphical framework to promote sound statistical analysis. **Perspectives on Psychological Science**, v. 15, n. 4, p. 1054-1075, 2020.

FINGER, C. A. G. **Biometria florestal**. DCF-UFSM, 2006. 314 p.

FIORENTIN, L. D.; BONAT, W. H.; PELISSARI, A. L.; MACHADO, S. A.; TÉO, S. J. Modelagem marginal conjunta da altura e volume para a *Araucaria angustifolia*. **BIOFIX Scientific Journal**, v. 5, n. 1, p. 121-129, 2020.

FIORENTIN, L. D. **Statistical methods to deal with different aspects of the covariance structure in applied forest biometrics**. 2020. 100 f. Tese (Doutorado em Engenharia Florestal) – Universidade Federal do Paraná, Curitiba, 2020.

FISCHER, C. Comparing the logarithmic transformation and the Box-Cox transformation for individual tree basal area increment models. **Forest Science**, v. 62, n. 3, p. 297-306, 2016.

FORRESTER, D. I.; TACHAUER, I. H. H; ANNIGHOEFER, P.; BARBEITO, I.; PRETZCH, H.; RUIZ-PEINADO, R.; STARK, H.; VACCHIANO, G.; ZLATANOV, T.; CHAKRABORTY, T.; SAHA, S.; SILESHI, G. Generalized biomass and leaf area allometric equations for European tree species incorporating stand structure, tree age and climate. **Forest Ecology and Management**, v. 396, p. 160-175, 2017.

FOX, J. **Applied regression analysis and generalized linear models**. SAGE, 2016. 817 p.

FOX, J.; WEISBERG, S. **An {R} Companion to Applied Regression**. 3. ed. Thousand Oaks CA: SAGE. 2019. Consultado em 27 de janeiro de 2021. Disponível em: <https://socialsciences.mcmaster.ca/jfox/Books/Companion/>.

FRASER, R. A.; LIPSITZ, S. R.; SINHA, D.; FITZMAURICE, G. M.; PAN, Y. Approximate median regression for complex survey data with skewed response. **Biometrics**, v. 72, n. 4, p. 1336-1347, 2016.

GALTON, F. Hereditary talent and character. **Macmillan's Magazine**, v. 12, p. 157-166, 1865.

GARSON, G. D. **Testing statistical assumptions**. Asheboro: Statistical Associates Publishing, 2012. 52 p.

GELMAN, A.; HILL, J. **Data analysis using regression and multilevel/hierarchical models**. Cambridge University Press, 2007. 651 p.

GHOSH, M.; GHOSH, T.; HIROSE, M. Y. Poisson counts, square root transformation and small area estimation. **Sankhya B**, v. 1, p. 1-23, 2021.

GIMENEZ, B. O.; SANTOS, L. T.; GEBARA, J.; CELES, C. H.; DURGANTE, F. M.; LIMA, A. J.; HIGUCHI, N. Tree climbing techniques and volume equations for *Eschweilera* (matá-matá), a hyperdominant genus in the Amazon Forest. **Forests**, v. 8, n. 5, p. 154, 2017.

GROSS, J.; LIGGES, U. **nortest: Tests for normality**. R package version v1.0-4, 2015. 10 p.

HILL, S. A. Statistics. In: HEMMINGGS, H. C.; HOPKINS, P. M. **Foundations of Anesthesia: Basic Sciences for Clinical Practice**, Mosby, 2006. p. 207-2017.



HOYLE, M. H. Transformations: an introduction and a bibliography. **International Statistical Review/Revue Internationale de Statistique**, v. 41. n. 2, p. 203-223, 1973.

HUBER, W.; VON HEYDEBRECK, A.; SUELTMANN, H.; POUSTKA, A.; VINGRON, M. Parameter estimation for the calibration and variance stabilization of microarray data. **Statistical Applications in Genetics and Molecular Biology**, v. 2, n. 1, p. 1-22, 2003.

IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Manual técnico da vegetação brasileira**. 2. ed. Rio de Janeiro, 2012. 271 p.

IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Produção da extração vegetal e da silvicultura 2019**. Rio de Janeiro, v. 34, 2019. 8 p.

KEARSLEY, E.; MOONEN, P. C.; HUFKENS, K.; DOETTERL, S.; LISINGO, J.; BOSELA, F. B.; VERBEECK, H. Model performance of tree height-diameter relationships in the central Congo Basin. **Annals of Forest Science**, v. 74, n. 1, p. 1-13, 2017.

KELMANSKY, D. M.; MARTÍNEZ, E. J.; LEIVA, V. A new variance stabilizing transformation for gene expression data analysis. **Statistical applications in genetics and molecular biology**, v. 12, n. 6, p. 653-666, 2013.

KELMANSKY, D. M.; RICCI, L. A new distribution family for microarray data. **Microarrays**, v. 6, n. 5, p. 1-12, 2017.

KERBER, R. L.; FAGUNDES, S. T. S.; SATILHO, C. O.; TRONCO, K. M. Q.; VENDRUSCOLO, J.; SCCOTI, M. S. V. Flora arbórea na Flona do Jamari, Rondônia. **Research, Society and Development**, v. 10, n. 4, p. 1-19, 2021.

KERSHAW JR.; J. A.; DUCEY, M. J.; BEERS, T. W.; HUSCH, B. **Forest mensuration**. John Wiley & Sons, 2016. 592 p.

KUMARI, K.; YADAV, S. Linear regression analysis study. **Journal of the practice of Cardiovascular Sciences**, v. 4, n. 1, p. 33-36, 2018.

KUMARI, P.; RAMESH, H. Remote sensing image based nearshore bathymetry extraction of Mangaluru coast for planning coastal reservoir. In: SITHARAM, T. G. et al. **Sustainable Water Resource Development Using Coastal Reservoirs**. Butterworth-Heinemann, p. 247-265, 2020.

LHOTKA, J. M.; CUNNINGHAM, R. A.; STRINGER, J. W. Effect of silvicultural gap size on 51 year species recruitment, growth and volume yields in Quercus dominated stands of the Northern Cumberland Plateau, USA. **Forestry: An International Journal of Forest Research**, v. 91, n. 4, p. 451-458, 2018.

LILLIEFORS, H. W. On the Kolmogorov-Smirnov test for normality with mean and variance unknown. **Journal of the American statistical Association**, v. 62, n. 318, p. 399-402, 1967.

LIMA, A. J. N.; SUWA, R.; DE MELLO RIBEIRO, G. H. P.; KAJIMOTO, T., DOS SANTOS, J.; DA SILVA, R. P.; HIGUCHI, N. Allometric models for estimating above-and below-ground biomass in Amazonian forests at São Gabriel da Cachoeira in the upper Rio Negro, Brazil. **Forest Ecology and Management**, v. 277, p. 163-172, 2012.

LONG, J. S.; ERVIN, L. H. Using heteroscedasticity consistent standard errors in the linear regression model. **The American Statistician**, v. 54, n. 3, p. 217-224, 2000.

MACHADO, S. A.; FIGUEIREDO FILHO, A. **Dendrometria**. Guarapuava: UNICENTRO, 2009. 309 p.

MACHADO, S. A.; SOUZA, R. F.; JASKIU, E.; CAVALHEIRO, R. Construction of site curves for native *Mimosa scabrella* stands in the Metropolitan Region of Curitiba. **Cerne**, v. 17, n. 4, p. 489-497, 2011.

MACHADO, S. A.; URBANO, E.; CONCEIÇÃO, M. B. Comparação de métodos de estimativa de volume para *Pinus oocarpa* em diferentes idades e diferente regimes de desbastes. **Boletim de Pesquisas Florestal**, n. 50, p. 81-98, 2005.

MACPHEE, C.; KERSHAW, J.A.; WEISKITTEL, A.R.; GOLDING, J.; LAVIGNE, M.B. Comparison of approaches for estimating individual tree height-diameter relationships in the Acadian Forest region. **Forestry: An international Journal of Forest Research**, v. 91, n. 1, p. 132-146, 2018.

MANIESI, V.; GOMES, F. B.; BASTOS, A. S. Evolução do desmatamento e implicações ao uso e ocupação dos solos no entorno de unidades de conservação da Amazônia: O caso da floresta nacional do Jamari-RO. **Os Desafios da Geografia Física na Fronteira do Conhecimento**, v. 1, p. 6612-6624, 2017.

MANLY, B. F. Exponential data transformations. **Journal of the Royal Statistical Society: Series D (The Statistician)**, v. 25, n. 1, p. 37-42, 1976.

MCDONALD, J. H. **Handbook of biological statistics**. Baltimore: Sparky House Publishing, 2009. 317 p.

MEDINA, L.; CASTRO, P.; KREUTZMANN, A-K.; ROJAS-PERILLA, N.; **Trafo: Estimation, Comparison and Selection of Transformation**. Vienna: R Foundation for Statistical Computing, 2018. Consultado em 20 julho 2021. Disponível em: <https://CRAN.R-project.org/package=trafo>

MEDINA, L.; KREUTZMANN, A. K.; ROJAS-PERILLA, N.; CASTRO, P. The R package *trafo* for transforming linear regression models. **R Journal**, v. 9, n. 2, p. 99-123, 2019.

MONNESS, E. The power-normal distribution: application to forest stands. **Canadian Journal of Forest Research**, v. 41, n. 4, p. 707-714, 2011.

MONTGOMERY, D. C. **Introduction to linear regression analysis**. 5 ed. John Wiley & Sons, 2012. 872 p.

MONTGOMERY, D. C.; PECK, E. A.; VINING, G. G. **Introduction to linear regression analysis**. John Wiley & Sons, 2021. 704 p.

MORAL, R. A.; HINDE, J.; DEMÉTRIO, C. G. B. Half-normal plots and overdispersed models in R: the hnp package. **Journal of Statistical Software**, v. 81, n. 1, p. 1-23, 2017.

MOSER, P.; OLIVEIRA, L. Z. **Regressão linear aplicada à dendrometria: uma introdução e iniciação à linguagem R**. Edifurb, 2017. 152 p.

NASCIMENTO, R. G. M.; VANCLAY, J. K.; FIGUEIREDO FILHO, A.; MACHADO, S. A.; RUSCHEL, A. R.; HIRAMATSU, N. A.; DE FREITAS, L. J. M. The tree height estimated by non-power models on volumetric models provides reliable predictions of wood volume: the Amazon species height modelling issue. **Trees, Forests and People**, v. 2, p. 100028, 2020.

NG'ANDWE, P., CHUNGU, D., YAMBAYAMBA, A. M., CHILAMBWE, A. Modeling the height-diameter relationship of planted *Pinus kesiya* in Zambia. **Forest Ecology and Management**, v. 447, p. 1-11, 2019.

NOGUEIRA, E. M., NELSON, B. W., FEARNSIDE, P. M., FRANCA, M. B., OLIVEIRA, A. C. A. D. Tree height in Brazil's "arc of deforestation": shorter trees in South and Southwest Amazonia imply lower biomass. **Forest Ecology and Management**, v. 255, p. 2963-2972, 2008.

OLIVEIRA, L. Z.; KLITZKE, A. R.; FANTINI, A. C.; ULLER, H. F.; CORREIA, J.; VIBRANS, A. C. Robust volumetric models for supporting the management of secondary forest stands in the Southern Brazilian Atlantic Forest. **Anais da Academia Brasileira de Ciências**, v. 90, p. 1-16, 2018.

OSBORNE, J. W. Improving your data transformations: applying the Box-Cox transformation. **Practical Assessment, Research, and Evaluation**, v. 15, n. 12, p. 1-10, 2010.

OSBORNE, J. W. **Normalizing data transformations**. ERIC Digest, 2002. 6 p.

OSBORNE, J. W. Notes on the use of data transformations. Practical assessment, Research, and Evaluation. **Retrieved May**, v. 10, p. 1-8, 2003.

PARKS, D. R.; ROEDERER, M.; MOORE, W. A. A new "Logicle" display method avoids deceptive effects of logarithmic scaling for low signals and compensated data. Cytometry Part A: **The Journal of the International Society for Analytical Cytology**, v. 69, n. 6, p. 541-551, 2006.

PAYANDEH, B. Some applications of nonlinear regression models in forestry research. **The Forestry Chronicle**, v. 59, n. 5, p. 244-248, 1983.

PÉLLICO NETTO, S. **Inventário Florestal Nacional: florestas nativas do Rio Grande do Sul**. Brasília: IBDF, 1984. 345 p.

PÉLLICO NETTO, S.; BEHLING, A. Data transformation in biological assays. **Pesquisa Florestal Brasileira**, v. 41, p. 1-15, 2021.

PERE, A. Comparison of two methods for transforming height and weight to normality. **Annals of human biology**, v. 27, n. 1, p. 35-45, 2000.

PERRY, M. B. Prediction intervals for the original response when using Manly's exponential transformations. **Quality Engineering**, v. 30, n. 2, p. 195-211, 2018.

PICARD, N.; RUTISHAUSER, E.; PLOTON, P.; NGNOMANDA, A.; HENRY, M. Should tree biomass allometry be restricted to power models? **Forest Ecology and Management**, v. 353, p. 156-163, 2015.

PICARD, N.; SAINT-ANDRE, L.; HENRY, M. **Manual for building tree volume and biomass allometric equations: from field measurement to prediction**. Roma: FAO, Rome and Centre de Coopération Internationale en Recherche Agronomique pour le Développement, Montpellier, 2012. 215 p.

PINO, F. A. A questão da não normalidade: uma revisão. **Revista de economia agrícola**, v. 61, n. 2, p. 17-33, 2014.

PORTÉ, A.; BARTELINK, H. H. Modelling mixed forest growth: a review of models for forest management. **Ecological Modelling**, v. 150, n. 1-2, p. 141-188, 2002.

R CORE TEAM. R: **A language and environment for statistical computing**. Vienna: R Foundation for Statistical Computing, 2021. Consultado em 20 julho 2021. Disponível em <https://www.R-project.org/>

RAYMAEKERS, J.; ROUSSEEUW, P. J. Transforming variables to central normality. **Machine Learning**, v. 1, p. 1-23, 2021.

ROCKE, D. M.; DURBIN, B. Approximate variance-stabilizing transformations for gene-expression microarray data. **Bioinformatics**, v. 19, n. 8, p. 966-972, 2003.

RODRIGUES, C. A. L.; LANGONI, M. P.; RODRIGUES, M. F. A.; CONTIERO, W. **Inventário turístico: inventário de oferta turística**. Agudos: Secretaria de Esporte, Lazer e Turismo, 2016. 201 p.

ROJAS-PERILLA, N. **The use of data-driven transformations and their applicability in small area estimation**. 2018. 208 f. Tese (Doutorado em Estatística e Economia) - Universidade Livre de Berlim, Berlim, 2018.

ROJAS-PERILLA, N.; PANNIER, S.; SCHMID, T.; TZAVIDIS, N. Data-driven transformations in small area estimation. **Journal of the Royal Statistical Society: Series A (Statistics in Society)**, v. 183, n. 1, p. 121-148, 2020.

RUFINI, A. L.; SCOLFORO, J. R. S., OLIVEIRA, A. D. D., MELLO, J. M. D. Equações volumétricas para o cerrado sensu stricto, em Minas Gerais. **Cerne**, v. 16, p. 1-11, 2010.

SAKIA, R. M. The Box-Cox transformation technique: a review. **Journal of the Royal Statistical Society**, v. 41, n. 2, p. 169-178, 1992.

SANQUETTA, C. R.; WOJCIECHOWSKI, J.; DALLA CORTE, A. P.; BEHLING, A.; PÉLICO NETTO, S.; RODRIGUES, A. L.; SANQUETTA, M. Comparison of data mining and allometric model in estimation of tree biomass. **BMC Bioinformatics**, v. 16, n. 1, p. 1-9, 2015.

SANTOS, A. T.; DOMENE, V. D.; MATTOS, P. P.; BRAZ, E. M.; CANETTI, A. NAKAJIMA, N. Y.; ROSOT, N. C. Equações de volume para espécies de Cerrado em Formosa, GO. **Pesquisa Florestal Brasileira**, v. 41, p. 1-8, 2021a.

SANTOS, M. F.; FIGUEIREDO FILHO, A.; GAMA, J. R. V.; RETSLAFF, F. A. de S.; COSTA, D. L. da. Species-specific equations: greater precision in commercial volume estimation in managed forests in the amazon. **Cerne**, v. 26, n. 3, p. 315-330, 2020.

SANTOS, M. F.; GAMA, J. R. V.; COSTA, D. L.; FIGUEIREDO-FILHO, A.; GOMES, K. M. A.; SOUSA LEITE, R.; DIAS, A. N. Generic and specific volumetric equations for individual trees in the Tapajos National Forest, Eastern Amazon. **Revista Brasileira de Ciências Agrárias**, v. 16, n. 1, p. 1-11, 2021b.

SCHMIDT, A. F.; FINAN, C. Linear regression and the normality assumption. **Journal of Clinical Epidemiology**, v. 98, p. 146-151, 2018.

SCHUMACHER, F. X., HALL, F. S. Logarithmic expression of timber-tree volume. **Journal of Agricultural Research**, v. 47, p. 719-734, 1933.

SCHWARZ, G. Estimating the dimension of a model. **Annals of Statistics**, v. 6, p. 461-464, 1978.

SCOLFORO, J. R. S. **Inventário florestal de Minas Gerais**: equações de volume, peso de matéria seca e carbono para diferentes fisionomias da flora nativa. 1 ed. Editora UFLA, 2008. 216 p.

SERINALDI, F.; GRIMALDI, S.; ABDOLHOSSEINI, M.; CORONA, P.; CIMINI, D. Testing copula regression against benchmark models for point and interval estimation of tree wood volume in beech stands. **European Journal of Forest Research**, v. 131, n. 5, p. 1313-1326, 2012.

SHEKIN, D. J. **Handbook of parametric and nonparametric statistical procedures**. Chapman and Hall/CRC, 2011. 1851 p.

SILESHI, G. W. A critical review of forest biomass estimation models, common mistakes and corrective measures. **Forest Ecology and Management**, v. 329, p. 237-254, 2014.

SILVA, L. C. R. D.; MACHADO, S. D. A.; GALVÃO, F.; FIGUEIREDO-FILHO, A. Evolução estrutural de bracatingais da região metropolitana de Curitiba, Paraná, Brasil. **Ciência Florestal**, v. 26, n. 1, p. 171-179, 2016.

SOLEYMANI, S. **Exact Box-Cox analysis**. 2018. 88 f. Tese (Doutorado em Filosofia) - Universidade do Ontário Ocidental, Ontário, 2018.

SOUZA, R. F.; MACHADO, A. S.; FIGUEIREDO-FILHO, A. Modelagem da biomassa total e da lenha por unidade de área para bracatingais nativos. **Revista Árvore**, v. 38, n. 1, p. 53-62, 2014.

SPEED, T. **Statistical analysis of gene expression microarray data**. Chapman and Hall/CRC, 2003. 218 p.

TEMESGEN, H.; AFFLECK, D.; POUDEL, K.; GRAY, A.; SESSIONS, J. A review of the challenges and opportunities in estimating above ground forest biomass using tree level models. **Scandinavian Journal of Forest Research**, v. 30, n. 4, p. 326-335, 2015.

TEMESGEN, H.; MONLEON, V.; WEISKITTEL, A.; WILSON, D. Sampling strategies for efficient estimation of tree foliage biomass. **Forest Science**, v. 57, n. 4, p. 326-335, 2015.

TERRA, M. D. C. N. S.; PRADO-JÚNIOR, J. A.; SOUZA, C. R.; PINTO, L. O. R.; SILVEIRA, E. M. O.; CORDEIRO, N. G.; CIRNE-SILVA, T. M.; MANTOVANI, V. A.; SCOLFORO, J. R. S.; MELLO, J. M. Tree species dominance in neotropical savanna aboveground biomass and productivity. **Forest Ecology and Management**, v. 496, p. 119430, 2021.

THAINES, F.; BRAZ, E. M.; MATTOS, P. P.; THAINES, A. A. R. Equações para estimativa de volume de madeira para a região da bacia do Rio Ituxi, Lábrea, AM. **Pesquisa Florestal Brasileira**, v. 30, n. 64, p. 283-289, 2010.

TSAI, A. C., LIU, M., SIMAK, M., & CHENG, P. E. On hyperbolic transformations to normality. **Computational Statistics & Data Analysis**, v. 115, p. 250-266, 2017.

TUKEY, J. W. **Exploratory data analysis**. Addison-Wesley Publishing Company, 1977. 711 p.

URBANO, E.; MACHADO, S. A.; FIGUEIREDO-FILHO, A.; KOEHLER, H. S. Modelagem da biomassa aérea em bracatingais nativos da região metropolitana de Curitiba. **Floresta**, v. 38, n. 2, p. 361-372, 2008.

URBANO, E.; MACHADO, S. A.; FIGUEIREDO-FILHO, A.; SANQUETTA, C. R. Modeling and dynamics of growth and yield of tree species in *Mimosa scabrella* stands. **Anais da Academia Brasileira de Ciências**, v. 93, n. 1, p. 1-13, 2021.

VANCLAY, J. K. **Modelling forest growth and yield: applications to Mixed Tropical Forests**. Wallingford: CAB International, 1994. 634 p.

VIBRANS, A. C.; MOSER, P.; OLIVEIRA, L. Z.; MAÇANEIRO, J. P. Generic and specific stem volume models for three subtropical forest types in southern Brazil. **Annals of Forest Science**, v. 72, p. 865-874, 2015.

WEISBERG, S. **Applied linear regression**. John Wiley & Sons, 2014. 370 p.

WICKHAM, H. **dplyr: Tidy Messy Data**. Vienna: R Foundation for Statistical Computing, 2021. Consultado em 20 julho 2021. Disponível em <https://dplyr.tidyverse.org/>

WICKHAM, H. **ggplot2: Elegant Graphics for Data Analysis**. New York: Springer-Verlag, 2016. Consultado em 20 julho 2021. Disponível em <http://ggplot2.tidyverse.org>

YANG, K; TU, J.; CHEN, T. Homoscedasticity: an overlooked critical assumption for linear regression. **General Psychiatry**, v. 32, n. 5, p. 1-5, 2019.

YANG, Z. A modified family of power transformations. **Economics Letters**, v. 92, n. 1, p. 14-19, 2006.

YANG, Z. L.; LIU, Y. D.; ZHU, X. S.; CHEN, X.; ZHANG, J. Removing forecasting errors with white gaussian noise after square root transformation. **Journal of Forecasting**, v. 35, n. 8, p. 741-750, 2016.

YEO, I. K.; JOHNSON, R. A. A new family of power transformations to improve normality or symmetry. **Biometrika**, v. 87, n. 4, p. 954-959, 2000.

YUAN, D. S.; IRIZARRY, R. A. High-resolution spatial normalization for microarrays containing embedded technical replicates. **Bioinformatics**, v. 22, n. 24, p. 3054-3060, 2006.

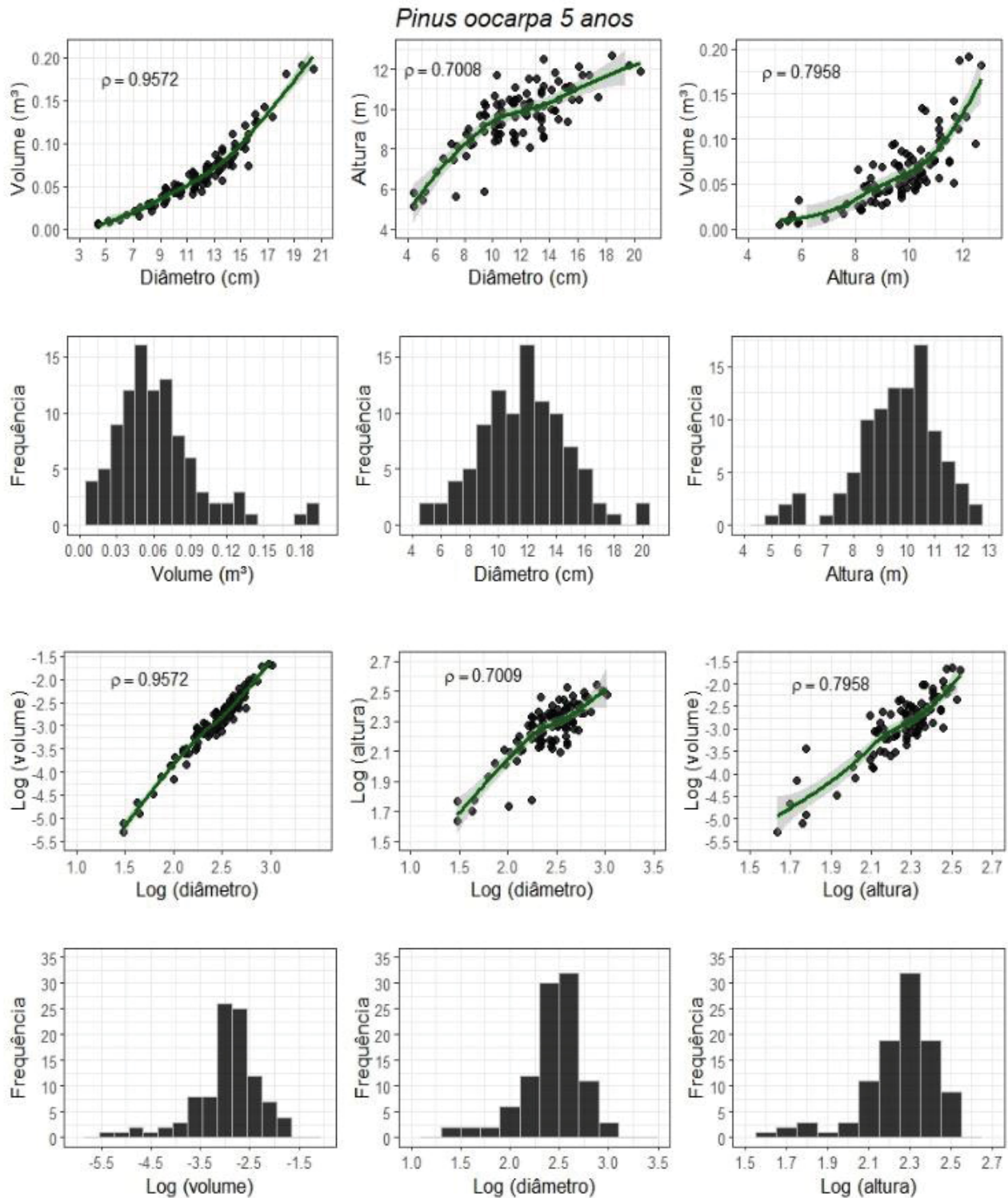
ZEA-CAMAÑO, J. D.; SOTO, J. R.; ARCE, J. E.; PELISSARI, A. L.; BEHLING, A.; ORSO, G. A.; EISFELD, R. D. L. Improving the modeling of the height-diameter relationship of tree species with high growth variability: robust regression analysis of *Ochroma pyramidale* (Balsa-Tree). **Forests**, v. 11, n. 3, p. 1-22, 2020.

ZELL, J. Climate sensitive tree growth functions and the role of transformations. **Forests**, v. 9, n. 382, p. 1-20, 2018.

ZHU, X.; MELNYKOV, V. Manly transformation in finite mixture modeling. **Computational Statistics & Data Analysis**, v. 121, p. 190-208, 2018.

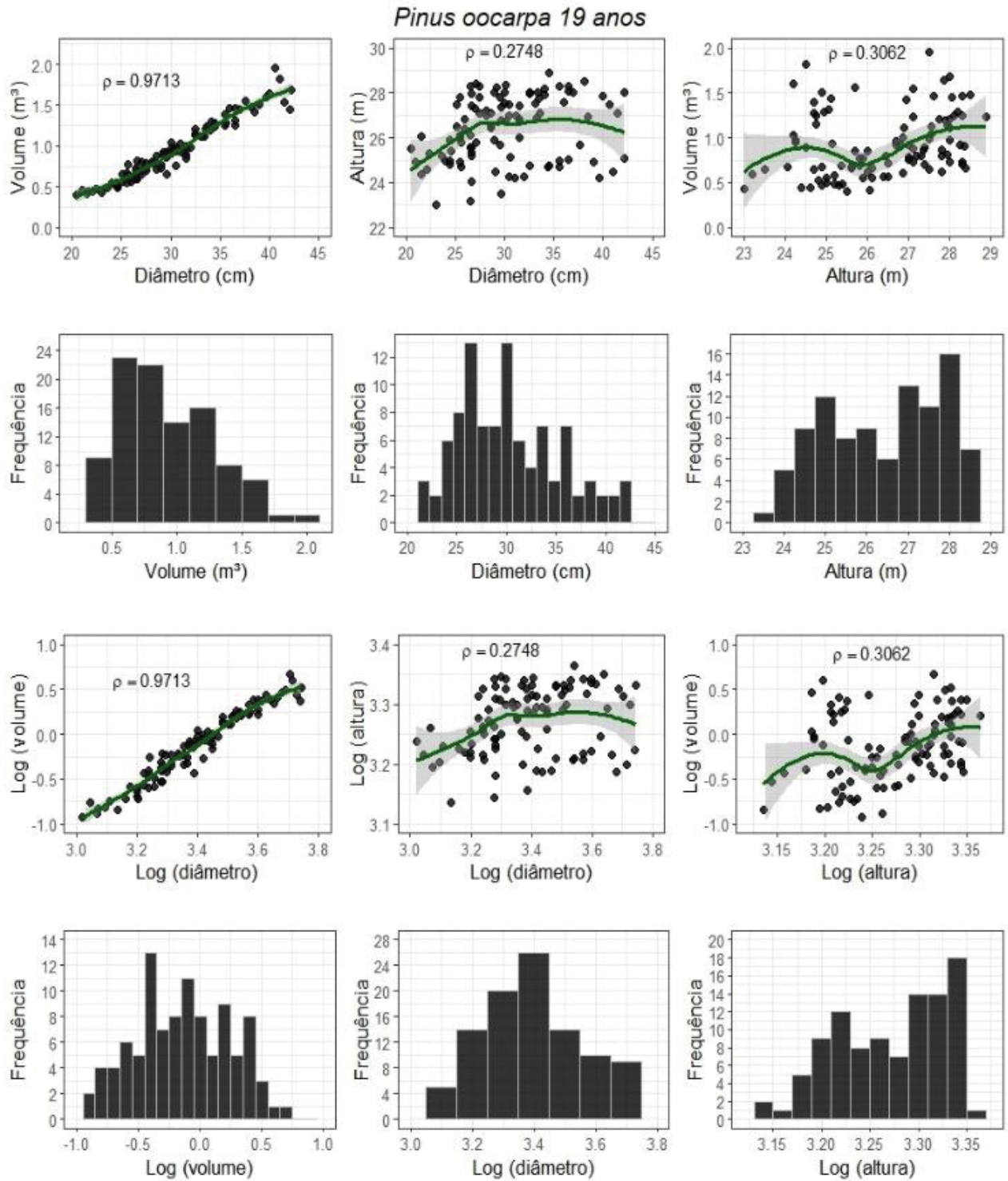
ZUEHLKE, T. W. Estimation of a type 2 Tobit model with generalized Box-Cox transformation. **Applied Economics**, v. 53, n. 17, p. 1952-1975, 2021.

**APÊNDICE 1 - HISTOGRAMAS DE FREQUÊNCIA E RELAÇÕES DE DISPERSÃO ENTRE AS VARIÁVEIS DENDROMÉTRICAS VOLUME, DIÂMETRO E ALTURA EM ESCALA ORIGINAL E SOB EFEITO DA TRANSFORMAÇÃO LOGARÍTMICA (Log).**



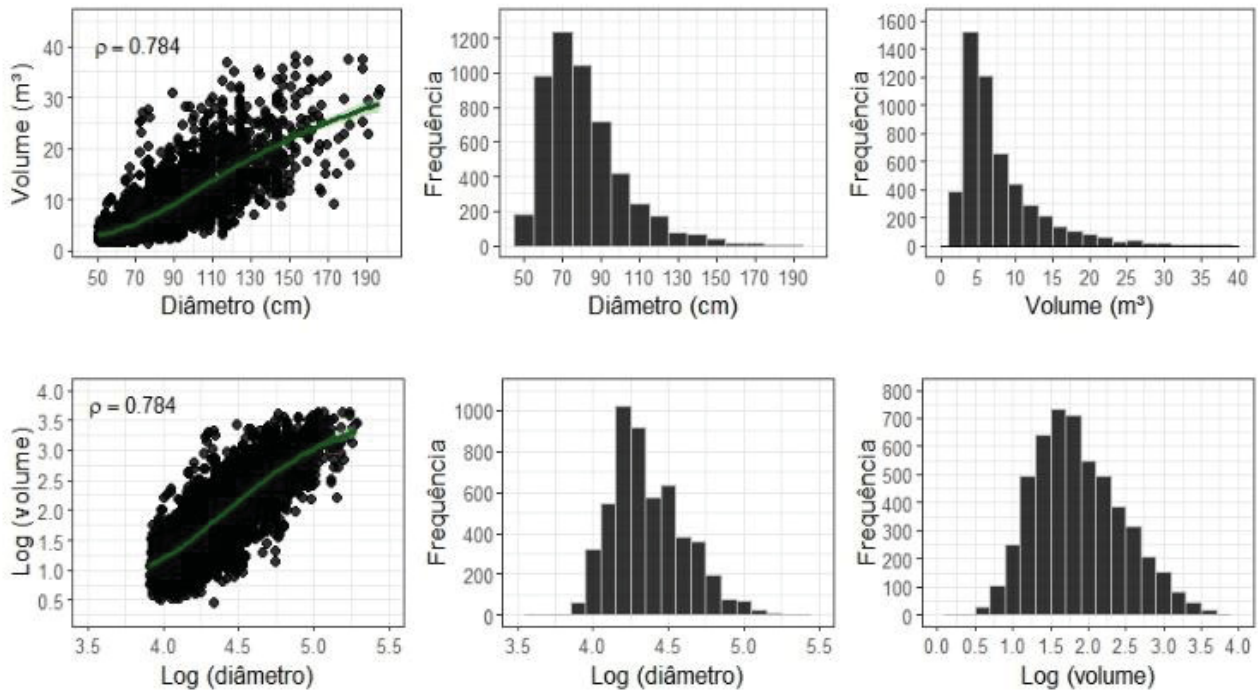
FONTE: O autor (2022).





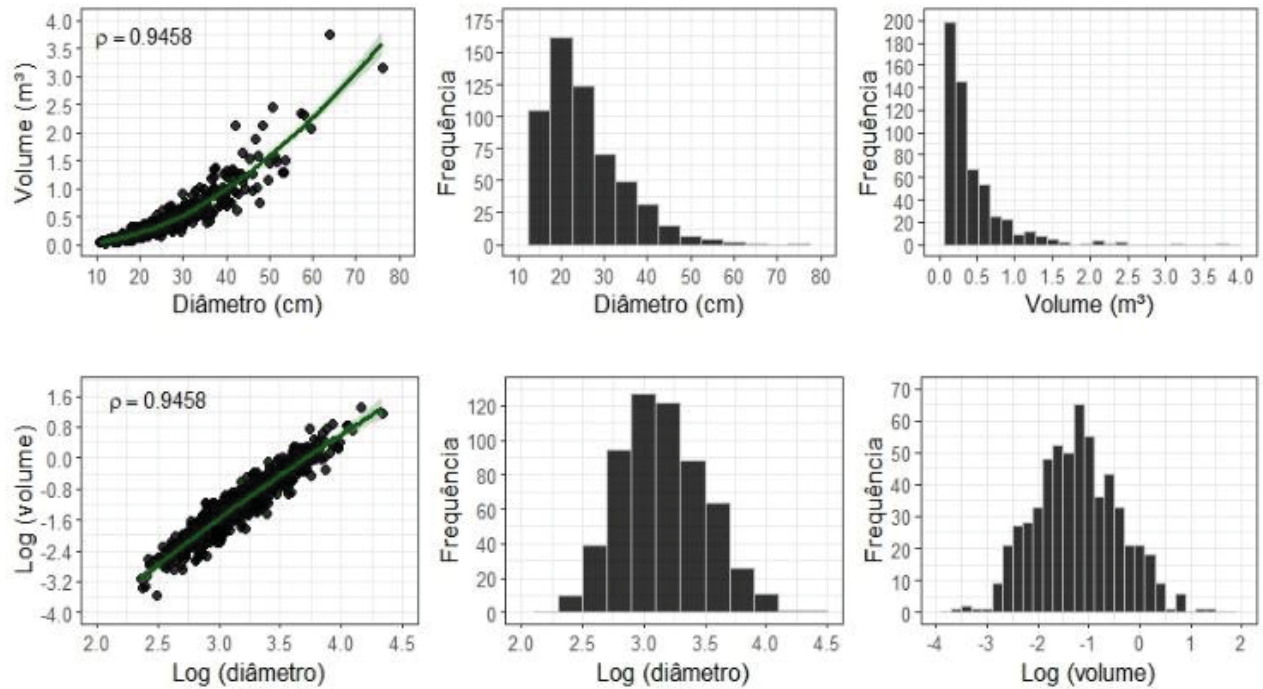
FONTE: O autor (2022).

**APÊNDICE 2 – HISTOGRAMAS DE FREQUÊNCIA E RELAÇÕES DE DISPERSÃO ENTRE AS VARIÁVEIS DENDROMÉTRICAS VOLUME E DIÂMETRO EM ESCALA ORIGINAL E SOB EFEITO DA TRANSFORMAÇÃO LOGARÍTMICA (Log). BASE DE DADOS ADVINDA DE UMA FLORESTA NATURAL DA AMAZÔNIA.**



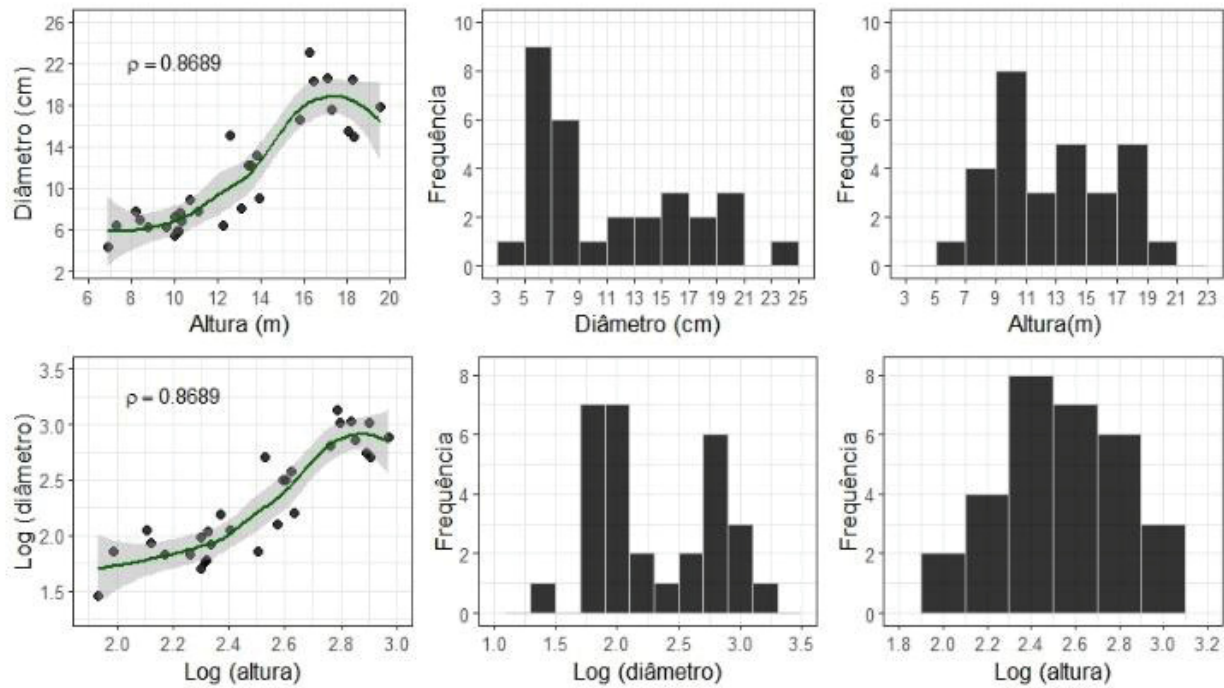
FONTE: O autor (2022).

**APÊNDICE 3 – HISTOGRAMAS DE FREQUÊNCIA E RELAÇÕES DE DISPERSÃO ENTRE AS VARIÁVEIS DEDROMÉTRICAS DIÂMETRO E VOLUME EM ESCALA ORIGINAL E SOB EFEITO DA TRANSFORMAÇÃO LOGARÍTMICA (Log). BASE DE DADOS ADVINDA DE UMA FLORESTA NATURAL DA MATA ATLÂNTICA.**



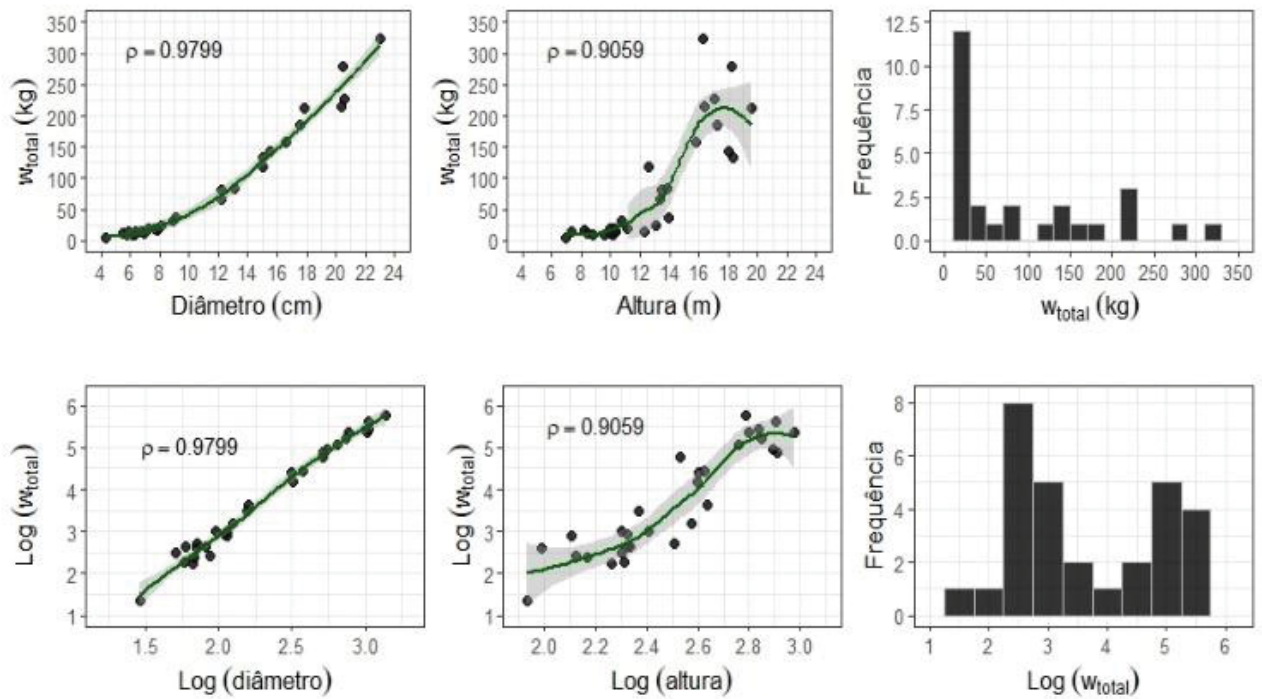
FONTE: O autor (2022).

**APÊNDICE 4 – GRÁFICOS DE DISPERSÃO E HISTOGRAMAS DE FREQUÊNCIA DAS VARIÁVEIS DENDROMÉTRICAS DIÂMETRO E ALTURA EM ESCALA ORIGINAL E SOB EFEITO DA TRANSFORMAÇÃO LOGARÍTMICA (Log). BASE DE DADOS ADVINDA DE UM POVOAMENTO DE *Mimosa scrabella*.**



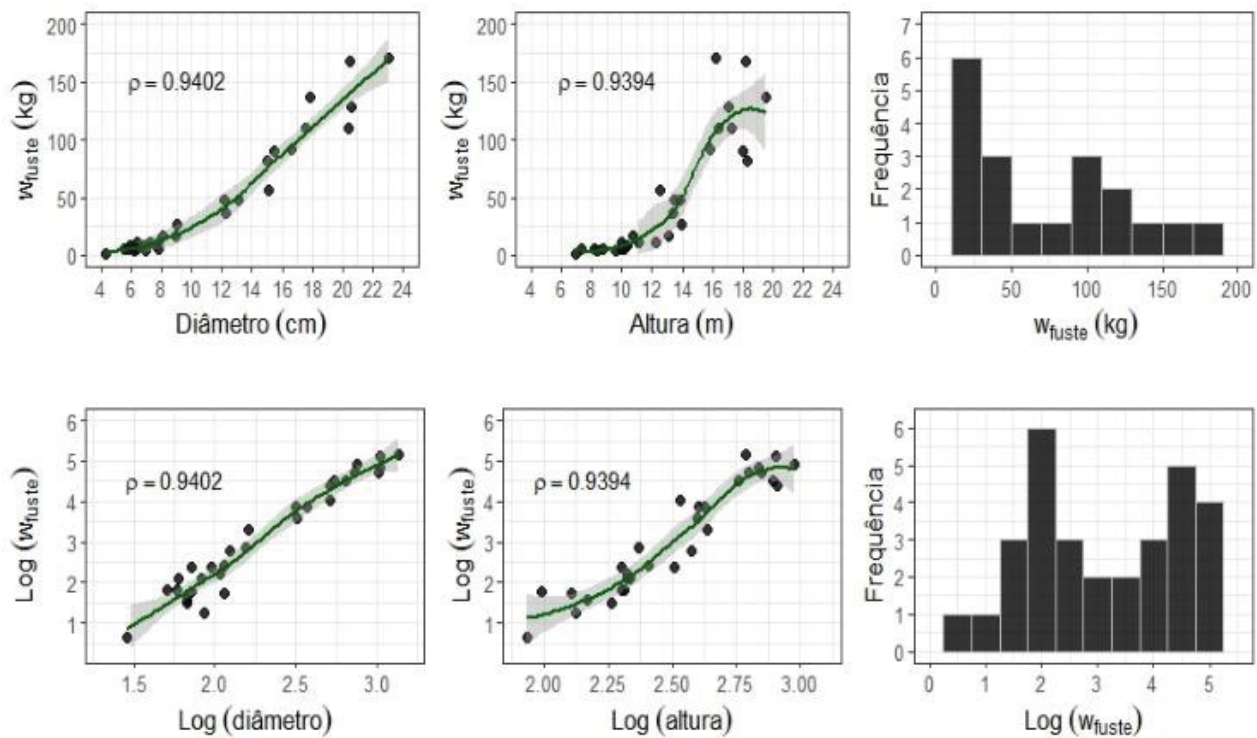
FONTE: O autor (2022).

**APÊNDICE 5 – HISTOGRAMA DE FREQUÊNCIA DA VARIÁVEL RESPOSTA BIOMASSA TOTAL E SUA DISPERSÃO EM FUNÇÃO DAS COVARIÁVEIS DIÂMETRO E ALTURA. VARIÁVEIS REPRESENTADAS EM SUA ESCALA ORIGINAL E SOB EFEITO DA TRANSFORMAÇÃO LOGARÍTMICA (Log). BASE DE DADOS ADVINDA DE UM POVOAMENTO DE *Mimosa scrabella*.**



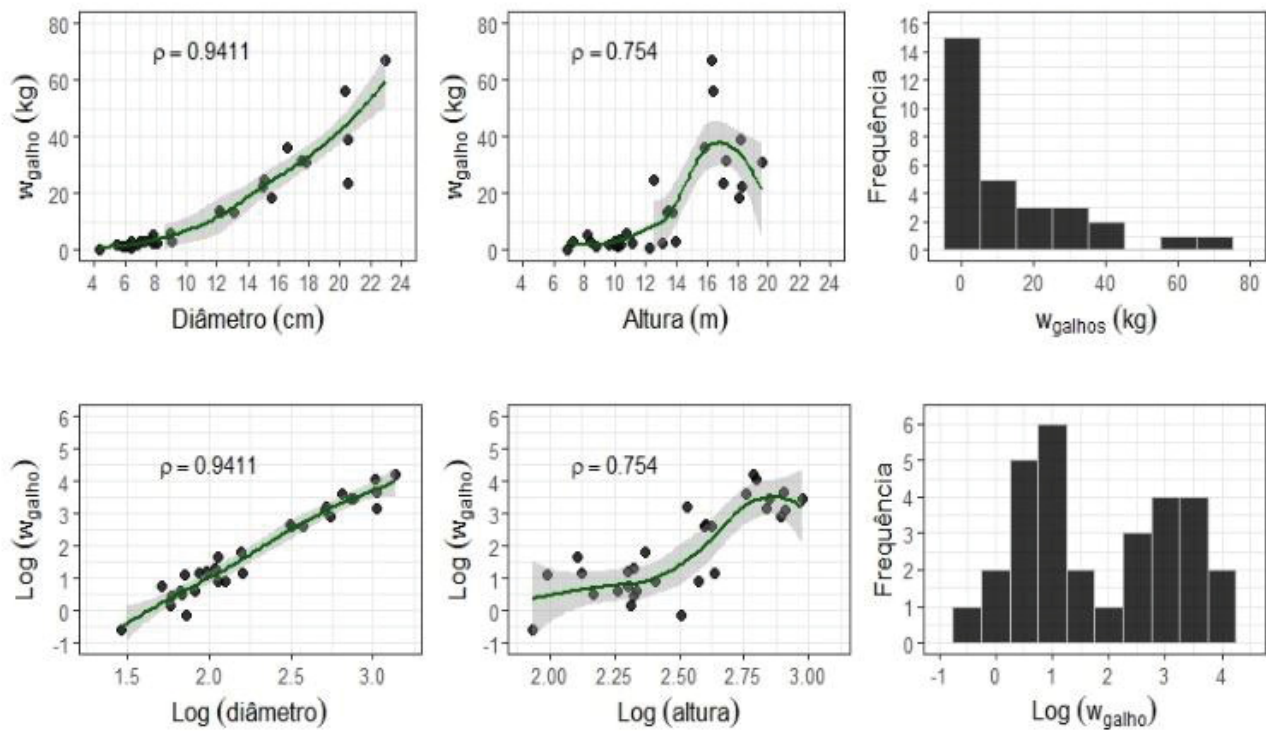
FONTE: O autor (2022).

**APÊNDICE 6 – HISTOGRAMA DE FREQUÊNCIA DA VARIÁVEL RESPOSTA BIOMASSA DOS FUSTES E SUA DISPERSÃO EM FUNÇÃO DAS COVARIÁVEIS DIÂMETRO E ALTURA. VARIÁVEIS REPRESENTADAS EM SUA ESCALA ORIGINAL E SOB EFEITO DA TRANSFORMAÇÃO LOGARÍTMICA (Log). BASE DE DADOS ADVINDA DE UM POVOAMENTO DE *Mimosa scrabella*.**



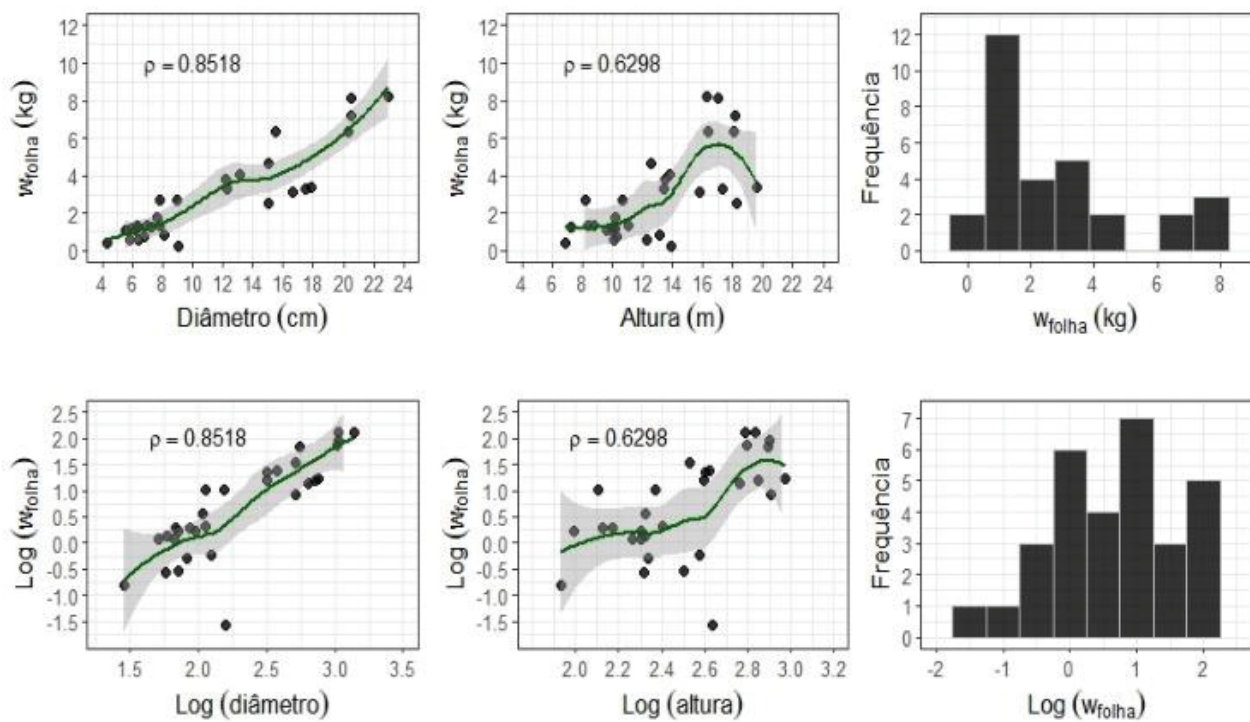
FONTE: O autor (2022).

**APÊNDICE 7 – HISTOGRAMA DE FREQUÊNCIA DA VARIÁVEL RESPOSTA BIOMASSA DOS GALHOS E SUA DISPERSÃO EM FUNÇÃO DAS COVARIÁVEIS DIÂMETRO E ALTURA. VARIÁVEIS REPRESENTADAS EM SUA ESCALA ORIGINAL E SOB EFEITO DA TRANSFORMAÇÃO LOGARÍTMICA (Log). BASE DE DADOS ADVINDA DE UM POVOAMENTO DE *Mimosa scrabella*.**



FONTE: O autor (2022).

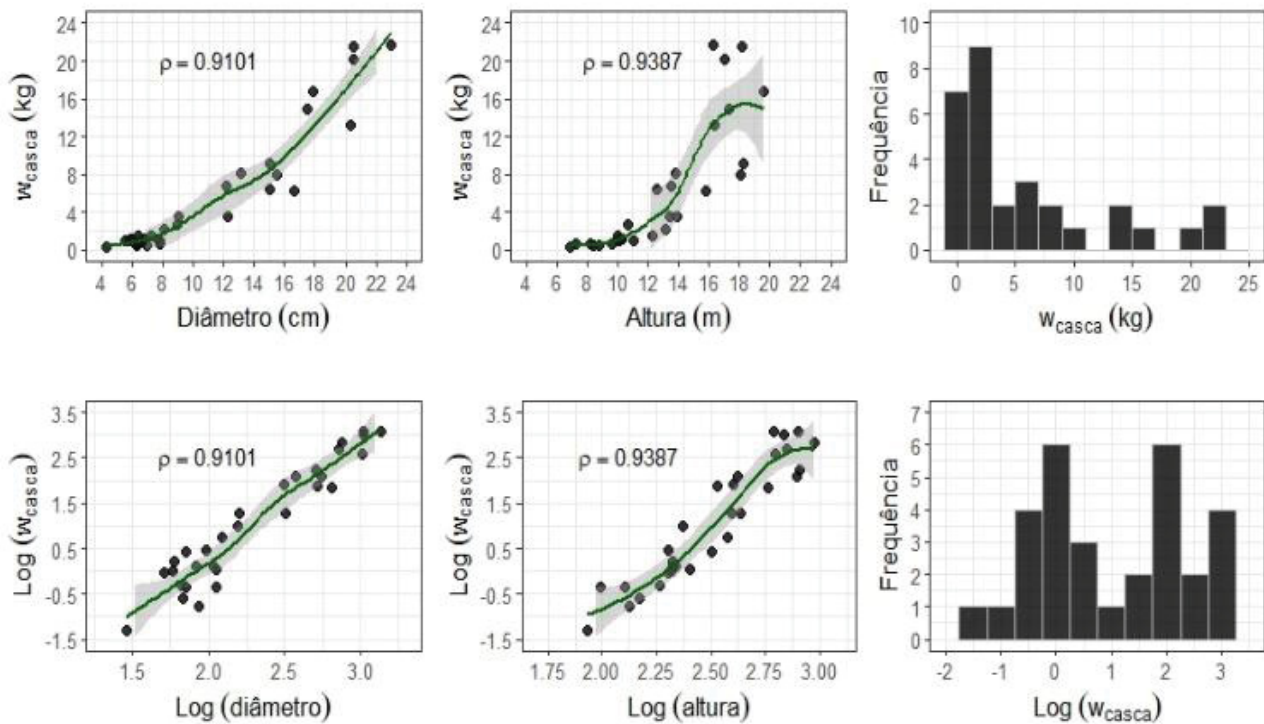
**APÊNDICE 8 – HISTOGRAMA DE FREQUÊNCIA DA VARIÁVEL RESPOSTA BIOMASSA DAS FOLHAS E SUA DISPERSÃO EM FUNÇÃO DAS COVARIÁVEIS DIÂMETRO E ALTURA. VARIÁVEIS REPRESENTADAS EM SUA ESCALA ORIGINAL E SOB EFEITO DA TRANSFORMAÇÃO LOGARÍTMICA (Log). BASE DE DADOS ADVINDA DE UM POVOAMENTO DE *Mimosa scrabella*.**



FONTE: O autor (2022).



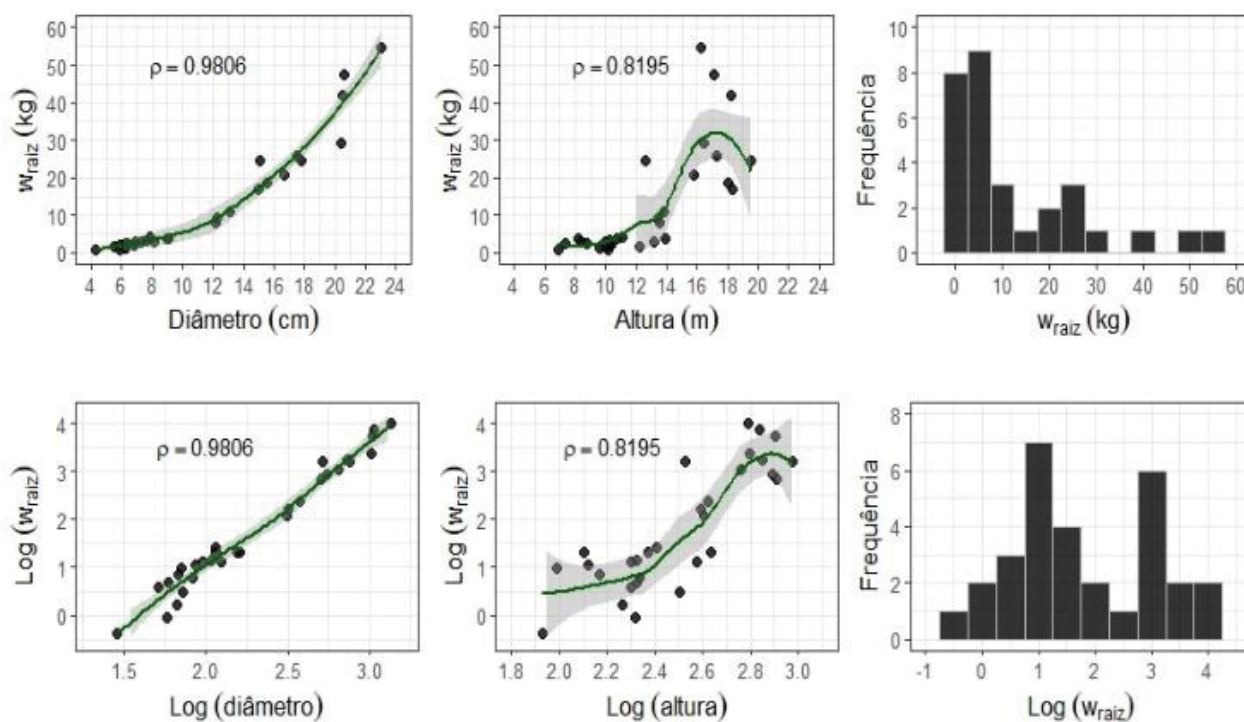
**APÊNDICE 9 – HISTOGRAMA DE FREQUÊNCIA DA VARIÁVEL RESPOSTA BIOMASSA DAS CASCAS E SUA DISPERSÃO EM FUNÇÃO DAS COVARIÁVEIS DIÂMETRO E ALTURA. VARIÁVEIS REPRESENTADAS EM SUA ESCALA ORIGINAL E SOB EFEITO DA TRANSFORMAÇÃO LOGARÍTMICA (Log). BASE DE DADOS ADVINDA DE UM POVOAMENTO DE *Mimosa scrabella*.**



FONTE: O autor (2022).

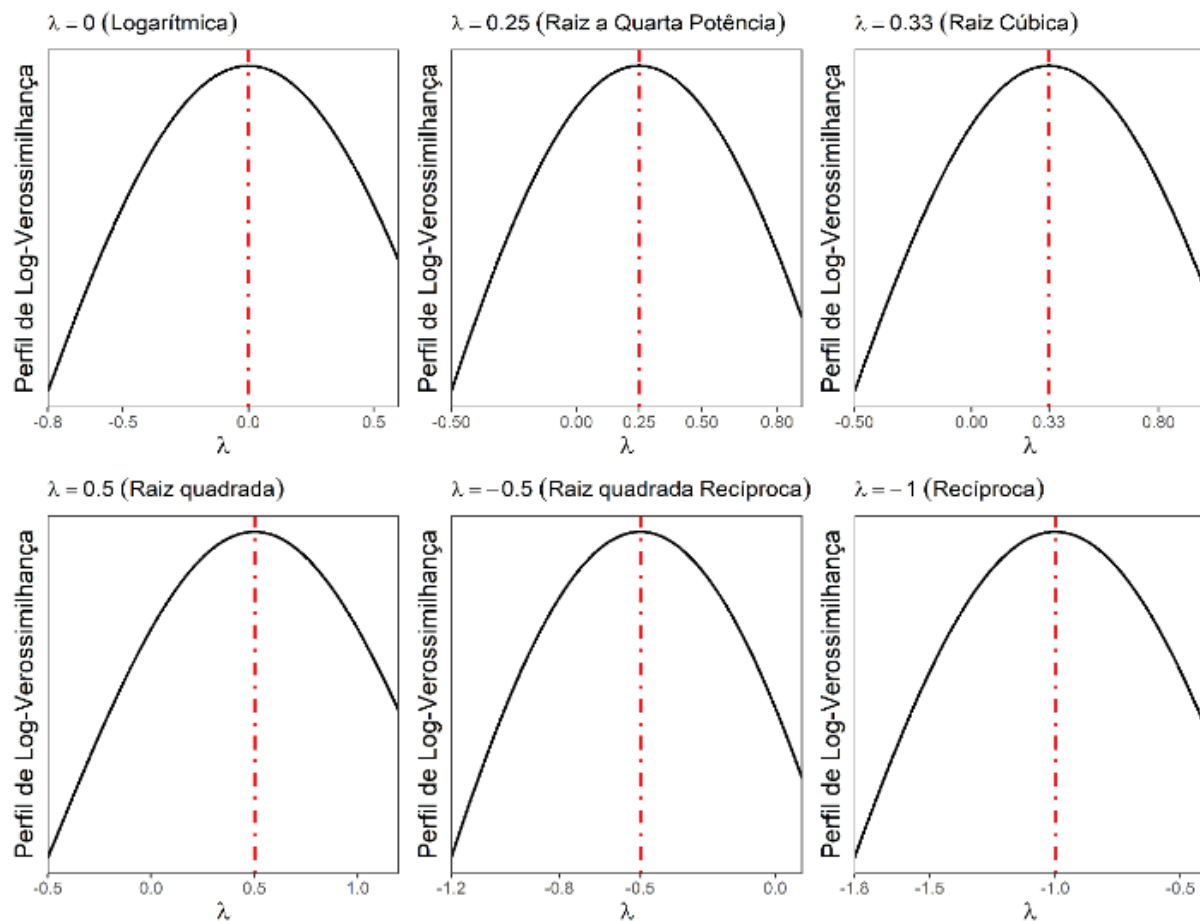
**APÊNDICE 10 – HISTOGRAMA DE FREQUÊNCIA DA VARIÁVEL  
RESPOSTA BIOMASSA DAS RAÍZES E SUA DISPERSÃO EM FUNÇÃO  
DAS COVARIÁVEIS DIÂMETRO E ALTURA. VARIÁVEIS REPRESENTADAS  
EM SUA ESCALA ORIGINAL E SOB EFEITO DA TRANSFORMAÇÃO  
LOGARÍTMICA (Log). BASE DE DADOS ADVINDA DE UM POVOAMENTO**

**DE *Mimosa scrabella***



FONTE: O autor (2022).

**APÊNDICE 11 – DEMONSTRAÇÃO DOS PERFIS DE MÁXIMA  
VEROSSIMILHANÇA DE LAMBDA ( $\lambda$ ) PARA OS CASOS ESPECIAIS DE  
TRANSFORMAÇÕES POTENCIAIS DE BOX-COX.**



FONTE: O autor (2022).