



Estratégias para transformação de variáveis na modelagem de povoamentos e florestas naturais

Transformation strategies for variables in modeling stands and natural forests

J. S. Rocha^{1*}; A. L. Pelissari¹; L. D. Fiorentin¹, L. R. Lansanova², V. C. Cysneiros³; C. K. Rodrigues¹

¹Programa de Pós-Graduação em Engenharia Florestal, Universidade Federal do Paraná, 80210-170, Curitiba-PR, Brasil

²Departamento de Pesquisa, Instituto Federal de Educação, Ciência e Tecnologia de Mato Grosso, 78320-000, Juína-MT, Brasil

³Departamento de Agricultura, Biodiversidade e Florestas, Universidade Federal de Santa Catarina, 89520-000, Curitibanos-SC, Brasil

*jobert.job.rocha@gmail.com

(Recebido em 16 de junho de 2023; aceito em 26 de julho de 2023)

A regressão linear tem sido amplamente utilizada em diversas áreas de conhecimento devido sua praticidade. Na ciência florestal, essa ferramenta é fundamental para modelagem de variáveis e relações dendrométricas. No entanto, para a adequada aplicação, algumas premissas devem ser atendidas, dentre as quais destacam-se a normalidade dos resíduos e a homogeneidade das variâncias, as quais são frequentemente violadas. Nessa perspectiva, as estratégias de transformação de variáveis são alternativas promissoras para corretas inferências estatísticas. Diante disso, objetivou-se avaliar diferentes estratégias de transformação em modelos de regressão linear para predição da altura e do volume por árvore em povoamentos e florestas naturais. Para a modelagem da relação altura-diâmetro, foram empregados dados de *Pinus oocarpa* em idades de 5 e 19 anos. Para a modelagem do volume, foram utilizadas bases de dados de florestas naturais da Amazônia e da Mata Atlântica, além dos povoamentos de *P. oocarpa*. 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 corrigiram simultaneamente as premissas de normalidade dos resíduos e homocedasticidade foram avaliadas estatisticamente para a seleção das estratégias mais adequadas. Assim, orienta-se a transformação em situações na qual pelo menos um dos pressupostos é violado. Nessas ocasiões, em estudos da relação altura-diâmetro, recomenda-se as estratégias de Manly e Box-Cox, ao passo que, em modelagens volumétricas, indicam-se a aplicação do logaritmo e de Box-Cox. Palavras-chave: alometria, relação altura-diâmetro, volumetria.

Linear regression has been widely used in several areas of knowledge due to its practicality. In forest science, this tool is essential for modeling tree variables and relationships. However, for proper application, some assumptions must be met, among which are the normality of residuals and homogeneity of variances, which are often violated. In this perspective, variable transformation strategies represent promising alternatives for correct statistical inferences. Therefore, this study aimed to evaluate different transformation strategies in linear regression models for predicting tree height and volume in stands and natural forests. To model the height-diameter relationship, data from *Pinus oocarpa* stands that were 5 and 19 years-old were used. For volume modeling, natural forest databases from Amazon and Atlantic Forests were used, as well as *P. oocarpa* stands. 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 corrected the assumptions of normality of residuals and homoscedasticity were statistically evaluated to select the most appropriate strategies. Thus, transformation is indicated in situations in which at least one of the assumptions is violated. On these occasions, in studies of height-diameter relationship, Manly and Box-Cox strategies are recommended, while in volume modeling, log and Box-Cox are indicated.

Keywords: allometry, height-diameter relationship, volume.

1. INTRODUÇÃO

As estimativas precisas de variáveis dendrométricas são uma das questões mais relevantes ao manejo adequado das florestas plantadas e naturais [1, 2]. Nessa perspectiva, a técnica de

regressão linear configura-se como uma ferramenta imprescindível, a qual pode ser empregada para prever o comportamento e a relação de determinados atributos de interesse a partir de uma ou mais variáveis explicativas [3, 4]. Além disso, os modelos de regressão linear fornecem informações consistentes para a tomada de decisão, sobretudo no que diz respeito à produção e ao uso sustentável das florestas [2, 5, 6]. No entanto, para a sua aplicação adequada, algumas premissas precisam ser rigorosamente atendidas. Apesar de fundamentais, tais premissas são comumente negligenciadas no ajuste de modelos.

Diante disso, em modelagens lineares de variáveis dendrométricas, comumente são violados alguns importantes pressupostos da regressão linear, como a normalidade dos resíduos e a homogeneidade das variâncias, também denominada de homocedasticidade. Contudo, o atendimento desses dois critérios é fundamental para obter equações cujos coeficientes sejam estimados com eficiência e forneçam estatísticas confiáveis relacionadas a testes estatísticos, como F e t [3, 4]. Desse modo, destacam-se as estratégias de transformação de variáveis como meios para corrigir a violação desses pressupostos para corretas inferências estatísticas [7].

A transformação de variáveis representa uma prática importante no tratamento de dados às análises estatísticas posteriores, necessária principalmente quando os resíduos gerados não são aderentes às suposições da regressão linear [8]. Contudo, verifica-se uma notável resistência quanto a sua adoção, pois são raros os trabalhos descritos na literatura florestal que abordam essa temática [9, 10]. Não obstante, a ausência de informações sobre o uso de transformação de variáveis contribui para a aplicação indiscriminada de transformações específicas, como, por exemplo, a logarítmica, sendo essa a principal utilizada [5, 11, 12]. Todavia, a literatura estatística apresenta diversos outros métodos, como as transformações Box-Cox, Manly, Bickel-Doksum, Yeo-Johnson, Glog, Dual power, G power, Log shift, entre outras [8], as quais até o momento não foram abordadas em modelagens de dados de povoamentos e florestas naturais.

Nesse contexto, o presente estudo tem como objetivo avaliar estratégias de transformação de variáveis mais adequadas para a modelagem de relações altura-diâmetro e volume comercial em povoamentos e florestas naturais. Desse modo, almeja-se que os pressupostos de normalidade dos resíduos e homocedasticidade sejam simultaneamente contemplados para corretas inferências estatísticas acerca da precisão e eficiência dos modelos.

2. MATERIAL E MÉTODOS

2.1 Bases de dados

Dados de dois povoamentos de *Pinus oocarpa* Schiede ex Schltdl., sendo um jovem de 5 anos e outro maduro de 19 anos (Tabela 1), foram utilizados para a modelagem da relação altura-diâmetro e do volume. Esses povoamentos são provenientes da região Sudoeste do estado de São Paulo, cujo clima é Aw (Köppen) [13], com inverno seco, temperatura média anual de 19 a 21 °C e índice pluviométrico anual de 1.340 mm. A base de dados consistiu em 100 árvores amostradas aleatoriamente em cada povoamento, em que foram mensurados os diâmetros a 1,3 m do solo (D) e as alturas totais (H). Para as modelagens volumétricas, a cubagem das árvores consistiu na medição dos diâmetros nas posições de 0,1 m; 1,3 m e em intervalos de 2 m até o diâmetro mínimo de 3 cm, de modo que os volumes totais individuais (v) foram obtidos pelo método de Smalian [14].

Para a modelagem do volume comercial (v_m), foram utilizadas bases de dados de florestas naturais da Amazônia e da Mata Atlântica. Os dados referentes à Floresta Amazônica (Tabela 2) foram oriundos de uma área de concessão na Floresta Nacional do Jamari, estado de Rondônia. O clima da região é do tipo Am (Köppen) [13]. Os períodos de inverno e seca são bem definidos, com média anual de precipitação de 2.400 mm e temperatura média anual de 25 °C. As variáveis utilizadas nas modelagens foram obtidas a partir da cubagem de 5.204 árvores distribuídas entre 33 espécies, das quais coletaram-se os diâmetros nas alturas 0,1 m, 1 m; 1,3 m; 2 m e a cada 1 m até a última seção do tronco, caracterizando a altura comercial das árvores. O processo de cubagem foi executado conforme o método de Smalian.

Tabela 1. Estatística descritiva das variáveis em povoamentos de *Pinus oocarpa* para modelagem da relação altura-diâmetro e do volume.

Variável	Idade (anos)							
	5				19			
	Mínimo	Média	Máximo	Desvio padrão	Mínimo	Média	Máximo	Desvio Padrão
<i>D</i> (cm)	4,40	11,75	20,40	3,11	20,50	30,16	42,20	5,20
<i>H</i> (m)	5,15	9,63	12,68	1,55	23,00	26,38	28,90	1,44
<i>v</i> (m ³)	0,0050	0,0640	0,1912	0,0367	0,3982	0,933	1,9615	0,3582

Tabela 2. Estatística descritiva das variáveis em florestas naturais da Amazônia e Mata Atlântica para modelagem do volume comercial.

Tipologia Florestal	<i>D</i> (cm)				<i>v_m</i> (m ³)			
	Mínimo	Média	Máximo	Desvio padrão	Mínimo	Média	Máximo	Desvio padrão
Amazônia	50,00	81,14	197,00	21,46	1,6031	7,7907	38,039	5,3645
Mata Atlântica	10,66	25,04	76,11	9,30	0,0282	0,4128	3,752	0,4090

A base de dados que representa o domínio da Mata Atlântica (Tabela 2) foi obtida a partir da cubagem por métodos não-destrutivos de árvores em diferentes fragmentos florestais no estado do Rio de Janeiro. Segundo a classificação de Köppen, o clima da região é predominantemente Aw e Cwa, cuja precipitação anual oscila de 1.000 a 2.200 mm e a temperatura média anual de 12 a 24 °C [13]. Ao todo, foram cubadas 583 árvores distribuídas em 84 espécies, das quais foram coletados os diâmetros nas alturas 0,3 m; 0,5 m; 1 m; 1,3 m; 2 m e a cada 1 m até a última seção do tronco, referente ao ponto de inversão morfológica da árvore.

2.2 Estratégias de transformação de variáveis

Nas modelagens da relação altura-diâmetro e volume, utilizou-se modelos lineares simples tendo o diâmetro medido a 1,3 m (*D*) como variável explicativa, uma vez que essa é a principal variável mensurada em levantamentos florestais e, por vezes, a única disponível para predição, principalmente em florestas naturais. Nos métodos que utilizaram um parâmetro de transformação lambda (λ), as transformações foram aplicadas na variável resposta. O parâmetro de transformação λ foi otimizado por meio do pacote trafo versão 1.0.1 [8] do programa computacional R [15], visando maximizar a verossimilhança dos ajustes. Ao todo, foram testadas 14 estratégias de transformações, as quais estão descritas na sequência.

Transformação logarítmica (Log e Log-Log): essa estratégia foi abordada de duas formas, sendo aplicada apenas na variável resposta (1) e, também, nas variáveis resposta (*Y*) altura (*H*) ou volume (*v*) e explicativa diâmetro a 1,3 m do solo (*D*) simultaneamente (2). O procedimento de retransformação à escala original foi efetuado a partir da exponenciação dos valores transformados.

$$\log(Y) = \beta_0 + \beta_1 D \quad (1)$$

$$\log(Y) = \beta_0 + \beta_1 \log(D) \quad (2)$$

em que *Y* é a variável resposta altura (m) ou volume (m³) e *D* é a variável explicativa diâmetro a 1,3 m do solo (cm).

Transformação recíproca: essa estratégia também foi aplicada sob duas abordagens. Primeiro, apenas na variável resposta (3) e, em seguida, tanto na variável resposta quanto na explicativa (4). A retransformação consistiu em obter o inverso da variável transformada.

$$\frac{1}{y} = \beta_0 + \beta_1 D \quad (3)$$

$$\frac{1}{y} = \beta_0 + \beta_1 \frac{1}{D} \quad (4)$$

Transformação de Box-Cox: esse método, proposto por Box e Cox (1964) [16], representa uma classe de transformações potenciais, em que se eleva os valores das variáveis a um parâmetro de transformação λ (5). A finalidade da estimativa de λ é a distribuição normal dos resíduos e a homogeneidade das variâncias.

$$y' = \begin{cases} \frac{y^\lambda - 1}{\lambda} & \text{se } \lambda \neq 0 \\ \log(y) & \text{se } \lambda = 0 \end{cases} \quad (5)$$

em que y' é a variável transformada e λ é o parâmetro de transformação estimado pelo método de máxima verossimilhança.

Nessa perspectiva, quando $\lambda \neq 0$, a modelagem com a transformação Box-Cox foi implementada conforme o modelo (6), com a retransformação obtida de acordo com a expressão (7).

$$\frac{y^\lambda - 1}{\lambda} = \beta_0 + \beta_1 D \quad (6)$$

$$y = (\lambda y' + 1)^{\frac{1}{\lambda}} \quad (7)$$

Transformação de Manly: esse método, proposto por Manly (1976) [17], pertence à família exponencial de transformação (8). Para λ diferentes de zero, o ajuste foi efetuado como o modelo (9). Em contrapartida, a retransformação à escala original foi obtida conforme a expressão (10).

$$y' = \begin{cases} \frac{e^{\lambda y} - 1}{\lambda} & \text{se } \lambda \neq 0 \\ y & \text{se } \lambda = 0 \end{cases} \quad (8)$$

$$\frac{e^{\lambda y} - 1}{\lambda} = \beta_0 + \beta_1 D \quad (9)$$

$$y = \frac{\log(\lambda y' + 1)}{\lambda} \quad (10)$$

Transformação de Bickel-Doksum: proposta por Bickel e Doksum (1981) [18], essa transformação foi aplicada conforme a expressão (11). As modelagens foram executadas com base no modelo (12), com a retransformação da variável resposta obtida de acordo com a expressão (13).

$$\frac{|y|^{\lambda \text{sinal}(y)} - 1}{\lambda} \quad \text{se } \lambda > 0 \quad (11)$$

$$\frac{|y|^{\lambda \text{sinal}(y)} - 1}{\lambda} = \beta_0 + \beta_1 D \quad (12)$$

$$\pm[\text{sinal}(y)(\lambda y + 1)]^{\frac{1}{\lambda}} \quad (13)$$

Transformação de Yeo-Johnson: essa estratégia, concebida por Yeo e Johnson (2000) [19], é definida a partir dos valores das variáveis respostas e dos parâmetros de transformações (λ), sendo aplicada conforme os procedimentos descritos em (14).

$$y' = \begin{cases} \frac{(y + 1)^\lambda - 1}{\lambda} & \text{se } \lambda \neq 0, y \geq 0 \\ \log(y + 1) & \text{se } \lambda = 0, y \geq 0 \\ \frac{(1 - y)^{2-\lambda} - 1}{\lambda - 2} & \text{se } \lambda \neq 2, y < 0 \\ -\log(1 - y) & \text{se } \lambda = 2, y < 0 \end{cases} \quad (14)$$

Enfatiza-se que, devido as características das variáveis avaliadas, o uso da estratégia de transformação Yeo-Johnson foi fundamentado com base em valores de y positivos. Dessa forma, as modelagens foram conduzidas conforme o modelo (15). Todavia, quando $\lambda \neq 0$, a estrutura do modelo ajustado é descrita no modelo (1). A retransformação da variável resposta foi obtida de acordo com a expressão (16).

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

$$y = (\lambda y' + 1)^{\frac{1}{\lambda}} - 1 \quad (16)$$

Transformação Glog: proposta por Durbin et al. (2002) [20], esse método fundamenta-se em uma modificação da transformação de base logarítmica, aplicada conforme a expressão (17).

$$y' = \log(y + \sqrt{y^2 + 1}) \quad (17)$$

Nessa conjectura, os modelos ajustados apresentaram a estrutura descrita em (18). O procedimento de retransformação para a estimativa da variável resposta foi executado de acordo com a expressão (19).

$$\log(y + \sqrt{y^2 + 1}) = \beta_0 + \beta_1 D \quad (18)$$

$$y = -\frac{[1 - (e^{y'})^2]}{2e^{y'}} \quad (19)$$

Transformação Dual power: essa estratégia, concebida por Yang (2006) [21], foi abordada conforme a expressão (20). Sua aplicação consistiu na estrutura de modelo descrita em (21). O procedimento de retransformação da variável resposta foi efetuada conforme a expressão (22).

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

$$\frac{(y^\lambda - y^{-\lambda})}{2\lambda} = \beta_0 + \beta_1 d \quad (21)$$

$$y = \left[\sqrt{1 + \lambda^2 y'^2} + \lambda y' \right]^{\frac{1}{\lambda}} \quad (22)$$

Transformação G power: desenvolvida por Kelmansky et al. (2013) [22], essa estratégia caracterizou-se como uma extensão da transformação Glog e conduzida conforme (23). Nessa perspectiva, quando $\lambda \neq 0$, os ajustes foram efetuados seguindo o modelo (24), ao passo que a retransformação foi obtida pela expressão (25).

$$y' = \begin{cases} \frac{(\mathcal{Y} + \sqrt{y^2 + 1})^\lambda - 1}{\lambda} & \text{se } \lambda \neq 0 \\ \log(\mathcal{Y} + \sqrt{y^2 + 1}) & \text{se } \lambda = 0 \end{cases} \quad (23)$$

$$\frac{(\mathcal{Y} + \sqrt{y^2 + 1})^\lambda - 1}{\lambda} = \beta_0 + \beta_1 D \quad (24)$$

$$y = - \left[\frac{1 - (\lambda y' + 1)^{\frac{2}{\lambda}}}{2(\lambda y' + 1)^{\frac{1}{\lambda}}} \right] \quad (25)$$

Transformação Log shift: esse método configura-se como uma extensão da transformação logarítmica, na qual inclui-se um parâmetro de transformação λ . Sua forma de aplicação está descrita na expressão (26), sendo abordada conforme o modelo (27). A retransformação da variável resposta foi obtida pelo procedimento (28).

$$y' = \log(\mathcal{Y} + \lambda) \quad (26)$$

$$\log(\mathcal{Y} + \lambda) = \beta_0 + \beta_1 D \quad (27)$$

$$y = e^{y'} - \lambda \quad (28)$$

Transformação de raiz quadrada deslocada: proposta por Medina et al. (2019) [8], essa estratégia é considerada uma extensão da transformação de raiz quadrada, atribuindo-se um parâmetro de transformação λ , conforme a expressão (29). As modelagens foram efetuadas com base na expressão (30), cuja retransformação foi obtida conforme (31).

$$y = \sqrt{y + \lambda} \quad (29)$$

$$\sqrt{y + \lambda} = \beta_0 + \beta_1 D \quad (30)$$

$$y = (y'^2 - \lambda) \quad (31)$$

Transformação Raiz quadrada: configura-se como uma das mais tradicionais para tratamento de dados. Sua aplicação consistiu na conversão das variáveis resposta e explicativa pelo método de raiz quadrada. Assim, os modelos seguiram a estrutura (32), com a retransformação obtida por meio do quadrado das variáveis transformadas.

$$\sqrt{y} = \beta_0 + \beta_1 \sqrt{D} \quad (32)$$

2.3 Avaliação dos ajustes

As modelagens com variáveis transformadas que atenderam simultaneamente as premissas de normalidade dos resíduos e homocedasticidade foram avaliadas estatisticamente para a seleção da estratégia mais adequada. A adequação ao pressuposto de normalidade dos resíduos foi avaliada com base no teste de Lilliefors, a 5% de significância, e no gráfico quantil-quantil. Para avaliar a homogeneidade das variâncias, utilizou-se o teste de Breusch-Pagan, a 5% de significância, e a análise gráfica de resíduos estudentizados.

Para a seleção das melhores estratégias de transformação, foram calculadas as estatísticas de coeficiente de determinação ajustado (R^2_{aj}), que expressa a quantidade da variação total explicada pela regressão; erro padrão da estimativa em porcentagem (SEE%), que indica o quanto os modelos erram em média ao estimar a variável dependente; e Critério de Informação Bayesiano (BIC), que avalia a parcimônia do ajuste, promovendo inferências sobre a qualidade do modelo. Essas estatísticas foram recalculadas para as unidades de medida originais das variáveis resposta. Eventuais vieses gerados nas transformações inversas das previsões logarítmicas para a escala

original foram corrigidos pelo fator de correção definido como $CF = (SEE^2/2)$ [5, 12]. Os coeficientes de regressão foram estimados pelo método dos mínimos quadrados ordinários, cuja significância foi avaliada pelo teste t, a 5% de significância.

3. RESULTADOS

Para a modelagem da relação altura-diâmetro em povoamento de *P. oocarpa* aos 5 anos, verificou-se que o tratamento das variáveis para atendimento dos pressupostos da análise de regressão linear não foi necessário. Assim, a homocedasticidade, pelo teste de Breusch-Pagan ($p\text{-BP} > 0,05$), e a normalidade dos resíduos, pelo teste de Lilliefors ($p\text{-Li} > 0,05$), foram atendidas com as variáveis em escala original (Tabela 3). No entanto, as transformações melhoraram as estatísticas de ajuste e precisão. Dentre as estratégias que aprimoraram a qualidade estatística, destacaram-se as transformações de Manly, Yeo-Johnson e Box-Cox. Ressalta-se que os coeficientes de regressão (β_0 e β_1) resultam na estimativa da variável resposta transformada, necessitando retransformá-la à escala original para aplicação preditiva dos modelos. Adicionalmente, os valores apresentados dos parâmetros de transformação (λ) foram os que resultaram na máxima verossimilhança dos ajustes.

Tabela 3. Estatísticas das melhores estratégias de transformação para modelagem da relação altura-diâmetro em povoamento de *Pinus oocarpa* aos 5 anos.

Transformação	R ² aj	SEE%	BIC	p-BP	p-Li	β_0	SE β_0	β_1	SE β_1	λ
Sem transformação	0,601	10,17	291,5	0,190	0,484	5,0728**	0,3848	0,3882**	0,3168	-
Manly	0,627	9,84	284,9	0,551	0,992	4,6535**	1,2077	1,2142**	0,0994	0,123
Yeo-Johnson	0,625	9,86	285,2	0,592	0,982	10,8359 ^{ns}	6,1802	6,2091**	0,5089	2,197
Box-Cox	0,625	9,86	285,3	0,599	0,980	4,9998 ^{ns}	4,1033	4,1223**	0,3378	2,069

R²aj é o coeficiente de determinação ajustado, 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 Beusch-Pagan, p-Li é o p-valor do teste de normalidade de Lilliefors, β_1 é o coeficiente de regressão estimado, ** significativo a 1% de probabilidade pelo teste t, ns não significativo, SE é o erro padrão do coeficiente de regressão e λ é o parâmetro de transformação.

O ajuste com as variáveis em escala original, além de atender as premissas da análise de regressão, forneceu coeficientes de regressão significativos (Tabela 3). Os ajustes relacionados às transformações Yeo-Johnson e Box-Cox foram os únicos a apresentar um de seus coeficientes não-significativos. Entre as estratégias testadas, as transformações recíproca e raiz quadrada deslocada não satisfizeram simultaneamente as premissas de normalidade dos resíduos e homocedasticidade.

Em contrapartida, o ajuste da relação altura-diâmetro para o povoamento de *P. oocarpa* aos 19 anos resultou em baixa relação entre as variáveis altura e diâmetro, com coeficiente de correlação linear de Pearson igual a 0,248. Portanto, o modelo ajustado sob essas condições estima um valor médio das alturas para quaisquer diâmetros, resultando em um coeficiente angular (β_1) estatisticamente não-significativo.

Por conseguinte, nas modelagens volumétricas dos povoamentos de *P. oocarpa*, observou-se que a transformação das variáveis foi necessária para as idades de 5 e 19 anos. Nessa perspectiva, destacaram-se as estratégias de base logarítmica (Tabela 4). Enfatiza-se que as modelagens com variáveis transformadas, além de promoverem a adequação aos pressupostos de normalidade dos erros e homogeneidade das variâncias, melhoraram as estatísticas de ajuste e precisão em ambas as idades.

Tabela 4. Estatísticas das melhores estratégias de transformação para modelagem volumétrica em povoamentos de *Pinus oocarpa* e florestas naturais.

Transformação	R ² aj	SEE%	BIC	p-BP	p-Li	β ₀	SEβ ₀	β ₁	SEβ ₁	λ
<i>Pinus oocarpa</i> – 5 anos										
Sem transformação	0,878	20,16	-575,8	<0,05	0,012	-0,0664**	0,0050	0,0111**	0,0004	-
Box-Cox	0,938	14,33	-644,0	0,083	0,994	-2,3255**	0,0166	0,0566**	0,0014	0,421
Bickel-Doksum	0,938	14,33	-644,0	0,083	0,994	-2,3255**	0,0165	0,0566**	0,0013	0,421
Log-Log	0,934	14,38	-643,4	0,286	0,751	-8,3601**	0,1199	2,2359**	0,0491	-
<i>Pinus oocarpa</i> – 19 anos										
Sem transformação	0,940	9,41	-190,8	<0,05	<0,05	-1,0827**	0,0519	0,0668**	0,0017	-
Log-Log	0,935	9,75	-183,7	0,607	0,058	-7,5643**	0,1789	2,1883**	0,0526	-
Log	0,900	12,12	-140,2	0,079	0,063	-2,2854**	0,0602	0,0711**	0,0019	-
Dual power	0,900	12,12	-140,2	0,079	0,063	-2,2854**	0,0602	0,0711**	0,0019	6,2×10 ⁻⁵
Floresta natural – Amazônia										
Sem transformação	0,630	41,86	27.095	<0,05	<0,05	-8,3143**	0,1768	0,1985**	0,0021	-
Log-Log	0,640	42,75	27.315	<0,05	<0,05	-6,6299**	0,0885	1,9469**	0,0202	-
Box-Cox	0,617	51,02	29.154	<0,05	<0,05	0,1217**	0,0199	0,0217**	0,0002	0,007
Floresta natural – Mata Atlântica										
Sem transformação	0,820	42,09	-368,9	<0,05	<0,05	-0,5839**	0,0207	0,0398**	0,0008	-
Log-Log	0,901	36,08	-548,6	0,190	0,932	-8,3755**	0,0987	2,2613**	0,0311	-
Box-Cox	0,895	40,28	-420,1	0,194	<0,05	-2,4558**	0,0229	0,0605**	0,0009	0,321

Nessa perspectiva, recomenda-se a transformação Box-Cox para modelagem volumétrica em povoamentos jovens com 5 anos, enquanto a transformação Log-Log foi a mais adequada para os povoamentos com 19 anos, por atenderem aos pressupostos de normalidade residual e homogeneidade das variâncias (p-Li e p-BP > 0,05) e não refletirem perdas nas estatísticas de ajuste. Assim, as análises gráficas dos volumes observados e estimados, resíduos estudentizados e gráficos quantil-quantil corroboram os resultados obtidos (Figuras 1 e 2). Destaca-se a eficiência dessas transformações ao corrigir as distorções residuais e linearizar as dispersões dos dados.

Para a modelagem volumétrica em florestas de produção na Amazônia (Tabela 4), as estratégias de transformação não foram capazes de corrigir a ausência de normalidade residual, bem como a heterogeneidade das variâncias (p-Li e p-BP < 0,05). Isto posto, depreende-se que o emprego de transformações na modelagem volumétrica mostrou-se inapropriado. Assim, com base nas análises gráficas, verificam-se distorções nas distribuições residuais, o que evidencia a presença de viés nas estimativas (Figura 3). Todavia, nota-se que a modelagem com transformações logarítmicas (Log-Log) promoveu dispersões residuais mais equalizadas. Por outro lado, a transformação Box-Cox acarretou prejuízos à estrutura residual da regressão, gerando resíduos notavelmente enviesados (Figura 3).

Considerando a modelagem do volume do fuste na Mata Atlântica (Tabela 4), verificou-se que a não normalidade residual e a heterogeneidade das variâncias foram simultaneamente corrigidas apenas com a transformação Log-Log (p-Li e p-BP > 0,05). A transformação Box-Cox, por sua vez, foi capaz de adequar apenas a homogeneização das variâncias (p-BP > 0,05). Ressalta-se, ainda, que o uso dessas estratégias promoveu ganhos à modelagem, resultando em melhores estatísticas de ajuste.

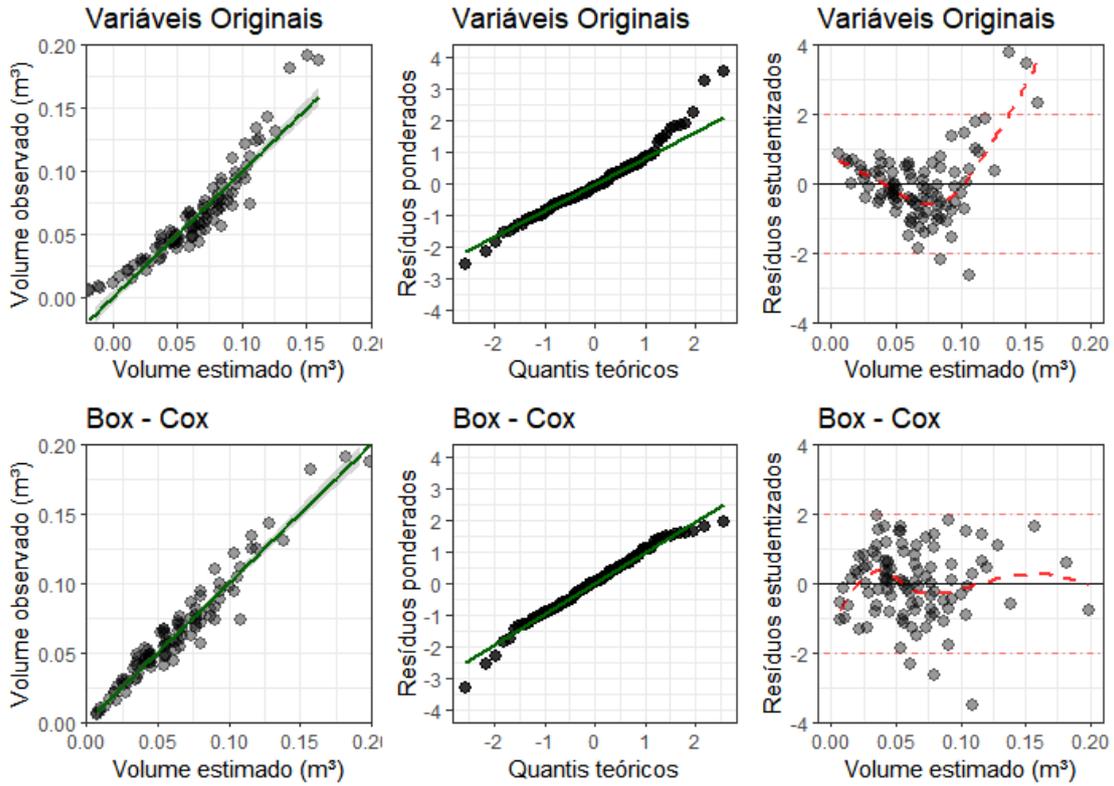


Figura 1. Análises gráficas das melhores estratégias de transformação para modelagem volumétrica em povoamentos de *Pinus oocarpa* aos 5 anos.

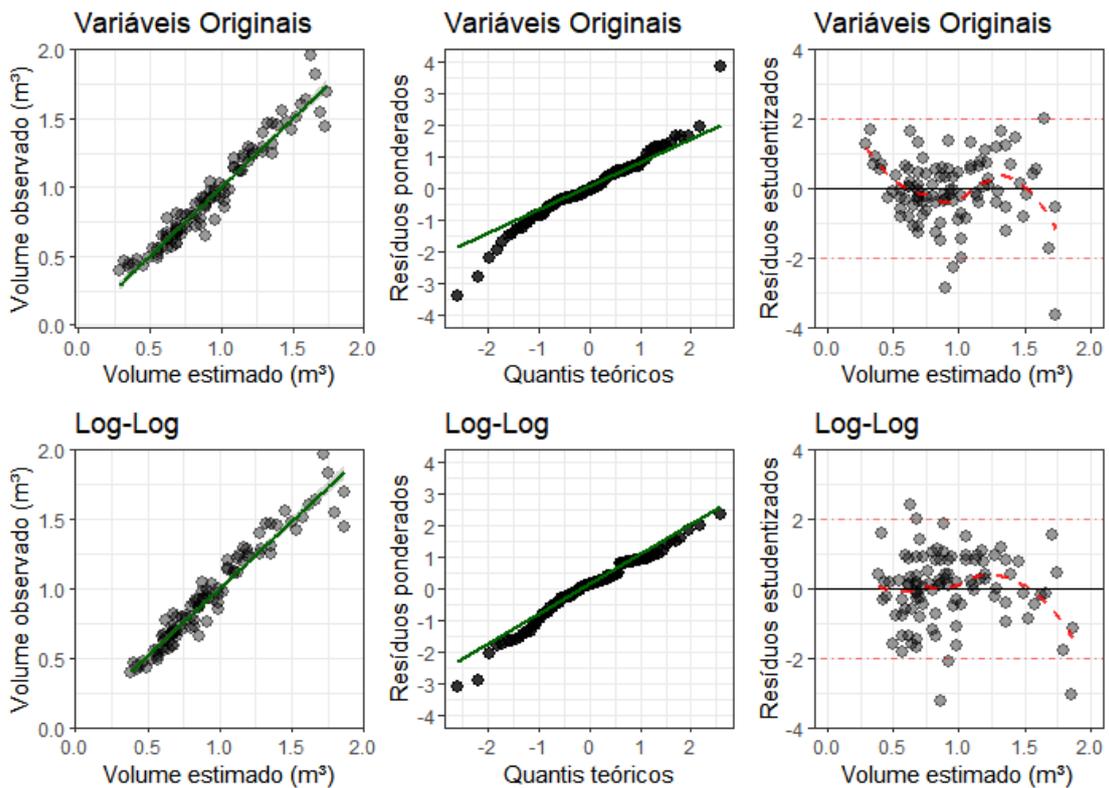


Figura 2. Análises gráficas das melhores estratégias de transformação para modelagem volumétrica em povoamentos de *Pinus oocarpa* aos 19 anos.

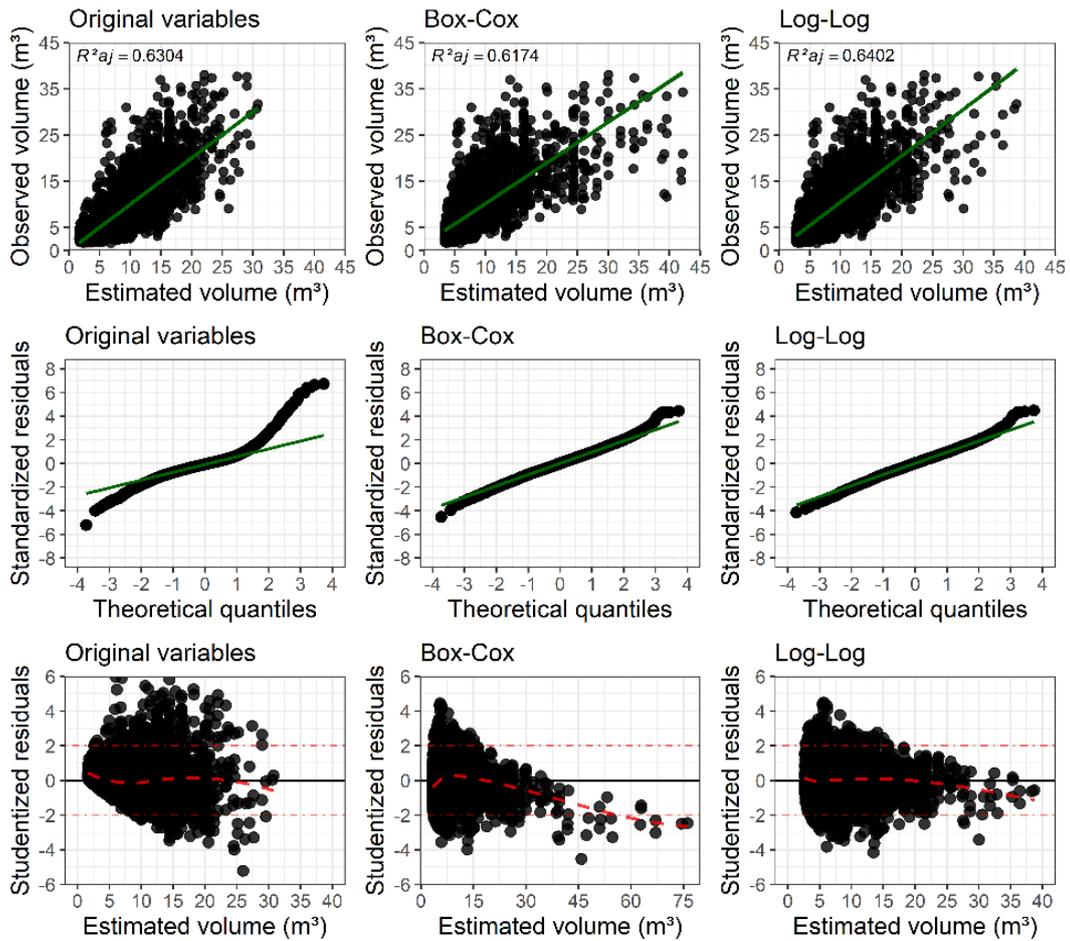


Figura 3. Análises gráficas das melhores estratégias de transformação para modelagem do volume comercial em floresta natural da Amazônia.

Em via de regra, observa-se que as transformações Log-Log e Box-Cox promoveram a linearização requerida para o procedimento de análise de regressão (Figura 4). Entretanto, a transformação Box-Cox, apesar de se configurar uma alternativa para modelagem volumétrica, não confere integralmente o atendimento das pressuposições da análise de regressão linear, devido a habilidade em corrigir apenas a normalidade dos resíduos. Por outro lado, a alternativa Log-Log de transformação mostra-se mais eficiente ao atendimento simultâneo das premissas requeridas.

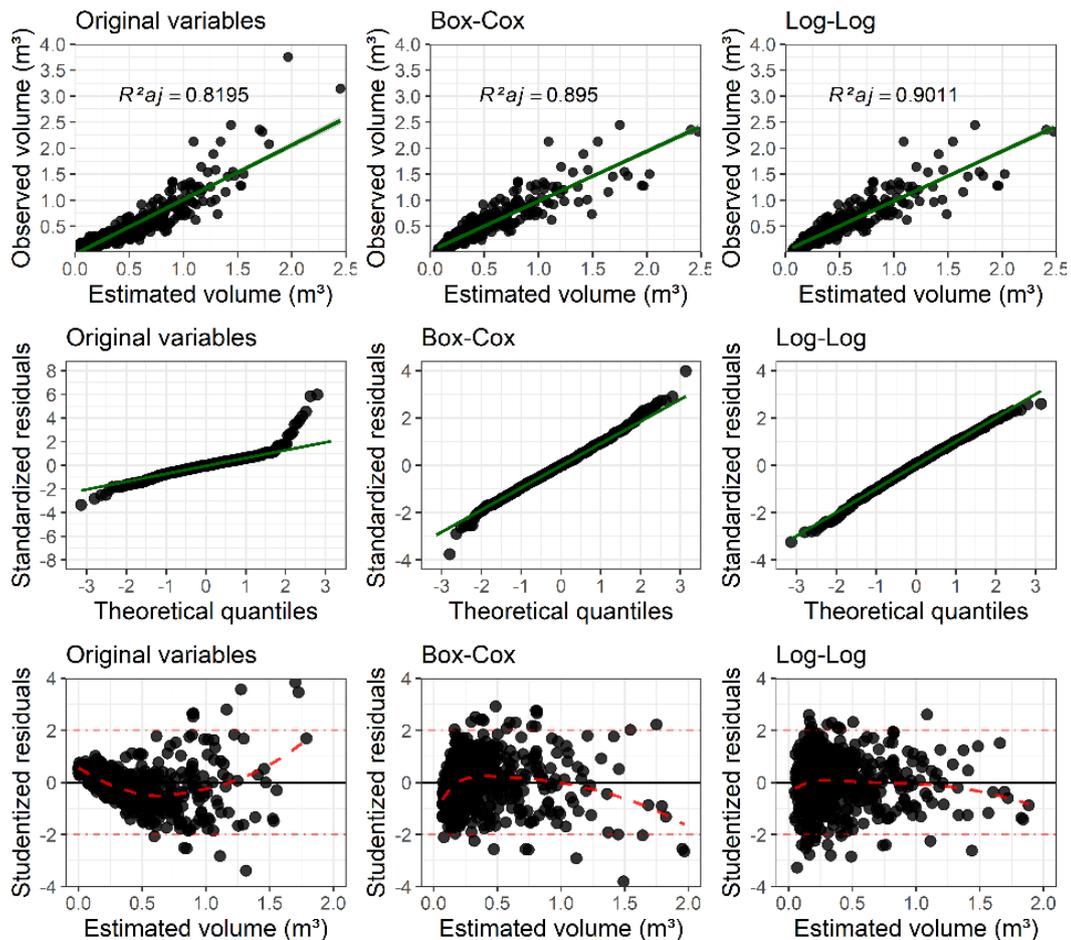


Figura 4. Análises gráficas das melhores estratégias de transformação para modelagem do volume comercial em floresta natural da Mata Atlântica.

4. DISCUSSÃO

Em florestas naturais ou plantadas, as relações entre as variáveis dendrométricas geralmente demonstram padrão não-linear [1, 23]. Entretanto, a suposição da associação linear entre as variáveis é requerida para se obter bons ajustes em abordagens alométricas fundamentadas no uso da regressão linear [4, 24]. Diante disso, além da adequação da normalidade residual e da homogeneidade das variâncias, a transformação de dados pode ser empregada para a correção de eventual má especificação entre as variáveis resposta e explicativa [7]. Nessa perspectiva, enfatiza-se que, diante das bases de dados de volume e altura, as transformações selecionadas a partir dos pressupostos de normalidade residual e homocedasticidade também linearizaram a relação entre as variáveis (Figuras 1 a 4).

Dentre as variáveis dendrométricas mais importantes à ciência florestal destaca-se a altura, requerida para melhoria das estimativas volumétricas de árvores, que, por sua vez, possibilita a prever outras variáveis, como biomassa, estoque de carbono e valor econômico [6, 12]. Nesse contexto, observou-se que a modelagem sem transformação demonstrou bons ajustes e atendeu às suposições da regressão em povoamentos de *P. oocarpa* aos 5 anos (Tabela 3). Diante disso, não se faz necessário o emprego de transformações nas variáveis para povoamentos jovens, visto que não houve a violação dos pressupostos. Contudo, as transformações de Manly, Yeo-Johnson e Box-Cox, apesar de não serem necessárias, destacaram-se por fornecer ganhos estatísticos às modelagens. Por outro lado, em povoamento maduro aos 19 anos, a regressão linear não foi uma ferramenta satisfatória, visto que a ausência de correlação significativa entre altura e diâmetro na idade avaliada prejudica o ajuste do modelo. Esse resultado reflete a dinâmica da relação altura-

diâmetro ao longo do tempo, bem como a influência dos desbastes, na estrutura do povoamento florestal [25].

Sob outra perspectiva, a análise volumétrica efetuada no presente estudo, cujas bases de dados representaram povoamentos e florestas 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 quando se trata de variáveis dendrométricas [1], o que representa uma preocupação recorrente nas pesquisas florestais. Diante disso, Sileshi (2014) [26] relacionou alguns problemas estatísticos comuns associados à modelagem, dentre os quais destaca-se a confiabilidade da estimativa dos coeficientes de regressão.

Dessa forma, para adequação da normalidade residual e homogeneidade das variâncias em modelagens volumétricas de povoamentos florestais sob idades avançadas, recomenda-se a aplicação das transformações de base logarítmica, seja simultaneamente na variável resposta e explicativa ou apenas na variável resposta. Por outro lado, diante de povoamentos jovens, além da transformação logarítmica, recomenda-se o uso da transformação 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 estratégias não deve ser aplicada precipitadamente, pois, a depender da base de dados, nem sempre serão promissoras ao atendimento simultâneo das suposições da regressão linear.

A adequação dos pressupostos de distribuição normal dos erros e homocedasticidade são imprescindíveis para obtenção de estimativas estatisticamente confiáveis, especialmente em modelagens lineares baseadas sob a abordagem do método de mínimos quadrados ordinários [3, 4, 24]. Dessa forma, por influenciar diretamente na estimação dos coeficientes de regressão, o pressuposto de homocedasticidade é amplamente discutido na literatura, de modo que, atualmente, são inúmeros os mecanismos disponíveis para a sua adequação. Dentre esses métodos, destacam-se o uso de matrizes de covariância consistentes com heterocedasticidade [27], métodos de mínimos quadrados ponderados [28] e técnicas de mineração de dados [29].

Contudo, apesar de inúmeras alternativas disponíveis para superar as limitações da regressão linear, o uso de modelos lineares representa a alternativa mais favorável ao analista de dados na predição de variáveis dendrométricas, em razão do conhecimento consolidado e da confiabilidade dos resultados. Dessa forma, as estratégias de transformação de variáveis representam ferramentas operacionalmente viáveis à modelagem aplicada ao manejo de povoamentos e florestas naturais, dado a disponibilidade em pacotes computacionais.

5. CONCLUSÃO

Recomenda-se a abordagem de transformação de variáveis em ocasiões em que há a violação de pelo menos um dos pressupostos da regressão linear em modelos dendrométricos, acarretando modelos corretos quanto às inferências estatísticas. Por outro lado, evidencia-se as transformações de Manly, Yeo-Johnson e Box-Cox para melhoria da qualidade estatística dos modelos de relação altura-diâmetro em povoamentos jovens de *Pinus oocarpa*.

Para a predição do volume das árvores em povoamentos maduros de *Pinus oocarpa*, recomenda-se a aplicação da transformação logarítmica, seja simultaneamente nas variáveis resposta e explicativa ou apenas na variável resposta. Por outro lado, quando se trata da modelagem volumétrica de povoamentos jovens e de florestas naturais da Amazônia e da Mata Atlântica, incentiva-se também a transformação Box-Cox.

6. AGRADECIMENTOS

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela bolsa de mestrado concedida ao primeiro autor.

7. REFERÊNCIAS BIBLIOGRÁFICAS

1. Fiorentin LD, Bonat WH, Pelissari AL, Machado SA, TEO SJ. Modelagem marginal conjunta da altura e volume para a *Araucaria angustifolia*. BIOFIX Scientific J. 2020;5(1):121-9. doi: 10.5380/biofix.v5i1.69745
2. Santos MF, Figueiredo Filho A, Gama JRV, Retslaff FAZ, Costa DL. Species-specific equations: greater precision in commercial volume estimation in managed forests in the Amazon. Cerne 2020;26(3):315-30. doi: 10.1590/01047760202026032741
3. Weisberg S. Applied linear regression. 4 ed. Hoboken (US): John Wiley & Sons; 2014.
4. Montgomery DC, Peck EA, Vining GG. Introduction to linear regression analysis. 3 ed. Hoboken (US): John Wiley & Sons; 2021.
5. Vibrans AC, Moser P, Oliveira LZ, Maçaneiro JP. Generic and specific stem volume models for three subtropical forest types in southern Brazil. Ann For Sci. 2015;72:865-74. Doi: 10.1007/s13595-015-0481-x
6. Nascimento RGM, Vanclay JK, Figueiredo Filho A, Machado as, Ruschel AR, Hiramatsu NA, et al. The tree height estimated by non-power models on volumetric models provides reliable predictions of wood volume: the Amazon species height modelling issue. Trees For People. 2020;2:100028. doi: 10.1016/j.tfp.2020.100028
7. Pretzsch H. Forest dynamics, growth and yield. 1 ed. Berlin: Springer; 2009.
8. Medina L, Kreutzmann AK, Rojas-Perilla N, Castro P. The R package trafo for transforming linear regression models. R Journal. 2019;9(2):99-123.
9. Fischer C. Comparing the logarithmic transformation and the Box-Cox transformation for individual tree basal area increment models. For Sci. 2016;62(3):297-306. doi: 10.5849/forsci.15-135
10. Zell J. Climate sensitive tree growth functions and the role of transformations. Forests. 2018;9(7):382. doi: 10.3390/f9070382
11. Cysneiros VC, Gaudi TD, Silveira Filho TB, Pelissari AL, Machado SDA, Carvalho DC, et al. Tree volume modeling for forest types in the Atlantic Forest: generic and specific models. iForest. 2020;1(5):417-25. doi: 10.3832/ifer3495-013
12. Cysneiros VC, Pelissari AL, Gaudi TD, Fiorentin LD, Carvalho DCD, Silveira Filho TB, et al. Modeling of tree height–diameter relationships in the Atlantic Forest: effect of forest type on tree allometry. Can J For Res. 2020;50(12):1289-98. doi: 10.1139/cjfr-2020-0060
13. Alvares CA, Stape JL, Sentelhas PC, Gonçalves JLM, Sparovek G. Koppen's climate classification map for Brazil. Meteorol Z. 2013;22(6):711-28. doi: 10.1127/0941-2948/2013/0507
14. Machado SA, Figueiredo-Filho A. Dendrometria. 2. ed. Irati (PR): Unicentro; 2014.
15. R Core Team. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2022.
16. Box GEP, Cox DR. An analysis of transformations. J R Stat Soc. 1964;26(2):211-43. doi: 10.1127/0941-2948/2013/0507
17. Manly BF. Exponential data transformations. J R Stat Soc. 1976;25(1):37-42. doi: 10.2307/2988129
18. Bickel PJ, Doksum KA. An analysis of transformations revisited. J Am Stat Assoc. 1981;76(374):296-311. doi: 10.2307/2287831
19. Yeo IK, Johnson RA. A new family of power transformations to improve normality or symmetry. Biometrika. 2000;87(4):954-9. doi: 10.1093/biomet/87.4.954
20. Durbin BP, Hardin JS, Hawkins DM, Rocke DM. A variance-stabilizing transformation for gene-expression microarray data. Bioinformatics 2002;18(1):S105-10. doi: 10.1093/bioinformatics/18.suppl_1.S105
21. Yang Z. A modified family of power transformations. Econ Lett. 2006;92(1):14-9. doi: 10.1016/j.econlet.2006.01.011
22. Kelmansky DM, Martínez EJ, Leiva V. A new variance stabilizing transformation for gene expression data analysis. Stat Appl Genet Mol Biol. 2013;12(6):653-66. doi: 10.1515/sagmb-2012-0030
23. 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. 2018;9(2):63. doi: 10.3390/f9020063
24. Fox J. Applied regression analysis and generalized linear models. 3. ed. Los Angeles (US): SAGE; 2016.
25. Figueiredo Filho A, Dias AN, Kohler SV, Verussa AA, Chiquetto AL. Evolution of the hypsometric relationship in *Araucaria angustifolia* plantations in the mid-south region of Paraná state. CERNE. 2010;16:347-57. doi: 10.1590/S0104-77602010000300010
26. Sileshi GW. A critical review of forest biomass estimation models, common mistakes and corrective measures. For Ecol Manag. 2014;329:237-54. doi: 10.1016/j.foreco.2014.06.026

27. Hayes AF, Cai L. Using heteroskedasticity-consistent standard error estimators in OLS regression: An introduction and software implementation. *Behav Res Methods*. 2007;39:709-72. doi 10.3758/BF03192961
28. Dutcă I, McRoberts RE, Næsset E, Blujdea VNB. Accommodating heteroscedasticity in allometric biomass models. *For Ecol Manag*. 2022;505:119865. doi: 10.1016/j.foreco.2021.119865
29. Sanquetta CR, Wojciechowski J, Dalla Corte AP, Behling A, Péllico Netto S, Rodrigues AL, et al. Comparison of data mining and allometric model in estimation of tree biomass. *BMC Bioinform*. 2015;16:247. doi: 10.1186/s12859-015-0662-5