

## AVALIAÇÃO DAS CONDICIONANTES DE REGRESSÃO NA ESTIMATIVA DE DIÂMETRO DE COPA PARA *Araucaria angustifolia*

### EVALUATION OF REGRESSION CONSTRAINTS ON CROWN DIAMETER ESTIMATION FOR *Araucaria angustifolia*

Gerson Santos Lisboa<sup>1</sup>, Dirceu Lúcio Carneiro de Miranda<sup>2</sup>, Thiago Floriani Stepka<sup>3</sup>, Solon Jonas Longhi<sup>4</sup>, Clebson Lima Cerqueira<sup>5</sup>, Luciano Cavalcante de Jesus França<sup>6</sup>, Edson Luis Serpe<sup>7</sup>

<sup>1</sup> Universidade Federal do Sul da Bahia, Itabuna, Bahia, Brasil – [gerson.lisboa@gmail.com](mailto:gerson.lisboa@gmail.com)

<sup>2</sup> Universidade Federal de Mato Grosso, Sinop, Mato Grosso, Brasil – [mirandaufpr@yahoo.com.br](mailto:mirandaufpr@yahoo.com.br)

<sup>3</sup> Universidade do Estado de Santa Catarina, Lages, Santa Catarina, Brasil – [tfstepka@yahoo.com.br](mailto:tfstepka@yahoo.com.br)

<sup>4</sup> Universidade Federal de Santa Maria, Santa Maria, Rio Grande do Sul, Brasil – [longhi.solon@gmail.com](mailto:longhi.solon@gmail.com)

<sup>5</sup> Universidade Federal do Paraná, Curitiba, Paraná, Brasil – [clebson.mensuracao@gmail.com](mailto:clebson.mensuracao@gmail.com)

<sup>6</sup> Universidade Federal de Lavras, Lavras, Minas Gerais, Brasil – [lucianodejesus@florestal.eng.br](mailto:lucianodejesus@florestal.eng.br)

<sup>7</sup> Lwarcel Celulose, Lençóis Paulista, São Paulo, Brasil – [serpe.edson@gmail.com](mailto:serpe.edson@gmail.com)

#### RESUMO

A pesquisa analisou as condicionantes de regressão na estimativa do diâmetro de copa para a *Araucaria angustifolia* em um fragmento de Floresta Ombrófila Mista localizada na Floresta Nacional de Irati, Paraná. Mediu-se a circunferência a 1,3 m do solo e o diâmetro da copa, considerando sua projeção no solo em duas direções (Norte/Sul e Leste/Oeste). Foram mensurados 189 diâmetros de copas, sendo analisadas 173 indivíduos de Araucária após a exclusão de *outliers*. As árvores foram selecionadas aleatoriamente e apresentavam-se isoladas e no estrato superior da floresta. Os dados foram processados com auxílio do *software Statistical Analysis System* versão 9.2®. O coeficiente de correlação de Pearson apresentou o valor de 0,78, indicando correlação forte e positiva. O coeficiente de assimetria apresentou valor de -0,2556, indicando assimetria fraca à esquerda e coeficiente de curtose de 0,2896 (platicúrtica), com curva mais achatada que a distribuição Normal teórica. Após verificada a tendência normal dos dados, foi utilizado o modelo linear. O teste  $\chi^2$  de Bartlett indicou heterogeneidade de variância a 95% de probabilidade. O teste de Shapiro-Wilk indicou que as amostras não provêm de uma população normal. O teste de Durbin-Watson provou que existe correlação positiva em série. Não foram atendidas as premissas da análise de regressão, necessitando transformar a variável dependente. A transformação da variável dependente pelo método Box-Cox foi utilizada após constatada a heterogeneidade de variância, o qual encontrou a potência ( $\lambda=0$ ), ou seja, a melhor transformação da variável dependente foi logaritmo neperiano de Y.

**PALAVRAS-CHAVE:** Análise de variância, Homocedasticidade, Independência, Normalidade, Box-Cox.

#### ABSTRACT

This research analyzed the regression constraints in the crown diameter estimation for *Araucaria angustifolia* in a fragment of Mixed Ombrophilous Forest located in National Forest of Irati, Paraná State. Circumference at 1.3 m above the ground and crown diameter were measured considering their projection in the soil in two directions (North / South and East / West). 189 crown diameters were measured, in which 173 crown diameters of Araucaria were analyzed. The trees were randomly selected, isolated and located in the upper forest stratum. Data were processed using the Software Statistical Analysis System version 9.2®. The Pearson correlation coefficient presented a value of 0.78, indicating a strong and positive correlation. The asymmetry coefficient presented a value of -0.2556, indicating weak asymmetry to the left and coefficient of kurtosis 0.2896 (platycurtic), with curve flatter than the Normal distribution. After verifying the normal trend of the data, the linear model was used. The Bartlett  $\chi^2$  test indicated heterogeneity of variance at a 95% probability. The Shapiro-Wilk's test indicated the samples did not come from a normal population. The Durbin-Watson's test proved there is a positive correlation in series. The assumptions of regression analysis were not met, necessitating transformation of the dependent variable. The transformation of the dependent variable by Box-Cox's method was used after verifying the variance heterogeneity, which found the power ( $\lambda = 0$ ), that is, the best transformation of dependent variable was logarithm of Y.

**KEYWORDS:** Analysis of variance, Homoscedasticity, Independence, Normality, Box-Cox.

## INTRODUÇÃO

A Floresta Ombrófila Mista é um dos mais exuberantes ecossistemas do Brasil e abriga uma das poucas coníferas de ocorrência subtropical no hemisfério Sul do continente americano: a *Araucaria angustifolia* (SANQUETTA, 2005). Essa tipologia florestal também é conhecida popularmente como Floresta com Araucária e/ ou Mata dos Pinheirais. O termo Floresta Ombrófila Mista foi inicialmente utilizado por Veloso & Goés Filho (1982) e adotado pelo IBGE (1992) na classificação fisionômico-ecológica da vegetação brasileira.

Este ecossistema sempre exerceu grande importância no desenvolvimento da região Sul do país, considerando-se a grande quantidade de empresas madeireiras que existiam, gerando empregos e riquezas a inúmeras cidades nos três estados do Sul do Brasil. Em contrapartida, pouca atenção foi dada para a conservação das florestas nativas, ou mesmo o seu manejo florestal sustentável, a fim de garantir que este recurso se perpetuasse para as gerações futuras.

Segundo Schneider & Finger (2000), o aumento da produção florestal ocorreu por meio do sacrifício dos ecossistemas e, apesar da exploração da madeira ter proporcionado bens econômicos momentâneos, provocou a conversão do uso da terra, empobrecendo a diversidade biológica e a sociedade, consequência da ocupação desordenada desse outrora rico ecossistema florestal. Acrescentaram ainda que o manejo florestal bem aplicado é uma das formas de uso das áreas florestais, por manter as principais funções ambientais da biodiversidade.

Um dos problemas frequentemente encontrados na prática é descrever e prever fenômenos observados. Isso pode ser resolvido pela construção de um modelo matemático que relacione as variáveis envolvidas no fenômeno, podendo este modelo ser utilizado para fins de predição.

De acordo com Gotelli & Ellison (2011), a regressão é utilizada para analisar relações entre variáveis contínuas. A regressão descreve a relação entre uma variável preditora, representada graficamente no eixo X, e uma variável resposta, representada no eixo Y. Neste contexto, as técnicas de modelagem apresentam alternativas eficientes que fornecem subsídios para a otimização de informações advindas do campo, tornando possível a integração destes dados, sendo que a técnica da regressão busca obter equações matemáticas a partir das medições efetuadas em campo.

Segundo Brena et al. (1978) e Schneider et al. (2009), a

análise de regressão tem sido usada com ênfase na solução de grande parte dos problemas florestais, especialmente quando se pretende obter estimativas de parâmetros da floresta, utilizando-se de relações biométricas que possibilitam obter valores estimados de forma indireta por meio de equações de regressão.

Para Scolforo (2005), o usuário de regressão tem normalmente dois objetivos: o primeiro consiste em encontrar um modelo para representar a relação funcional entre a variável dependente (Y) e independente (X); o segundo consiste em testar hipóteses entre a variável dependente e uma ou mais variáveis independentes.

Para que esta relação funcional seja ajustada, é necessária uma amostra, onde tanto os valores de X como os valores de Y sejam determinados. Pode-se, então, utilizando o método de ajuste linear ou não-linear, obter a estimativa dos parâmetros da regressão ( $B'_{is}$ ).

O ajuste destes modelos busca fazer previsões e com isto otimizar parte dos problemas encontrados na prática do Manejo Florestal, obtendo estimativas de parâmetros da floresta, possibilitando estimativas indiretas por meio de técnicas de regressão. Segundo Schneider et al. (2009), três aspectos devem ser considerados no uso de equações de regressão: o primeiro consiste em distribuir a amostra por classe da variável independente, a fim de obter boa precisão no ajuste das equações; o segundo são relacionados à escolha do modelo de regressão; o terceiro, refere-se à verificação das condicionantes de regressão, como, homogeneidade de variância, normalidade e independência.

A fonte de energia de uma árvore provém da radiação solar, que é transformada pelo processo da fotossíntese em energia química; dessa forma, pode-se considerar que a copa é a porção da árvore que influencia na interceptação da energia fotossintética e, consequentemente, no crescimento da árvore. O diâmetro da copa pode ser considerado como variável básica na estimativa de outras variáveis das árvores (SANQUETTA et al., 2014).

Variáveis, como a altura total da árvore, a altura de copa e o diâmetro de copa, podem ser facilmente correlacionadas com o diâmetro a 1,30 m do solo (DAP) (ORELLANA & KOEHLER, 2008; SANQUETTA et al., 2011; COSTA, 2011).

Neste trabalho, objetivou-se apresentar algumas estatísticas referentes à análise de regressão, dando ênfase às condicionantes de regressão. Alguns aspectos apresentados foram: análise exploratória e visual dos dados, Boxplot, coeficiente de correlação de Pearson,

teste de  $X^2$  de Bartlett, assimetria, curtose, normalidade, independência e transformação Box-Cox.

## MATERIAL E MÉTODOS

### Área de Estudo

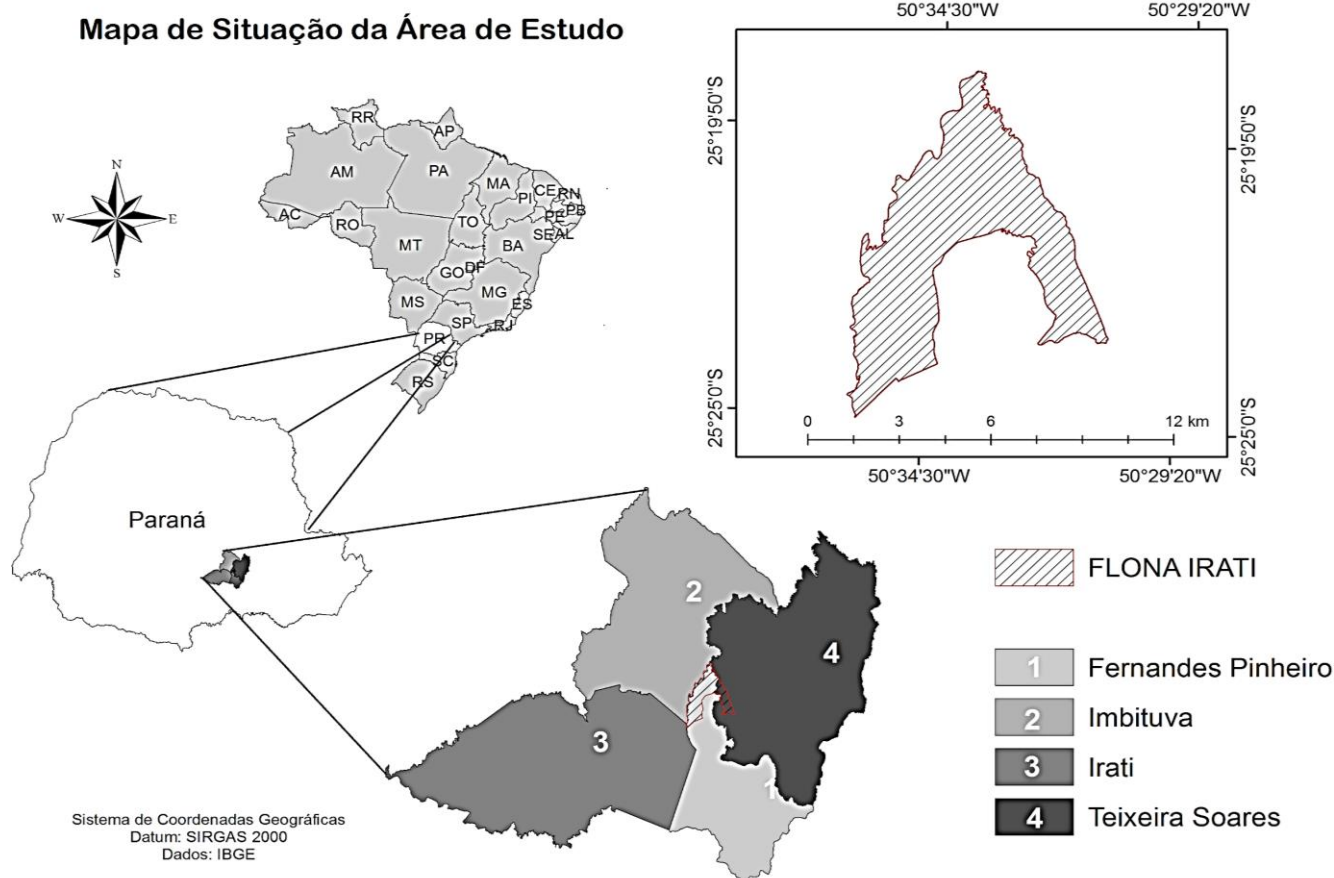
Os dados para o desenvolvimento deste trabalho foram coletados na Floresta Nacional (FLONA) de Irati, na microrregião Centro-Sul do Paraná (Figura 1), no Segundo Planalto Paranaense. A FLONA de Irati localiza-se entre as coordenadas geográficas 25°25' de latitude Sul, 50°36' de longitude Oeste, 25°17' de latitude Sul e 50°30' de longitude Oeste (Figura 2).

Encontra-se a uma altitude média de 812 m acima do nível do mar, a 10 km do centro da cidade de Irati, Paraná e a 150 km da capital do estado, Curitiba, Paraná. Segundo Mazza (2006) e conforme classificação de Köppen, o clima da região pertence ao tipo Cfb, subtropical úmido mesotérmico, caracterizado por verões frescos, geadas severas e frequentes, sem estação seca.

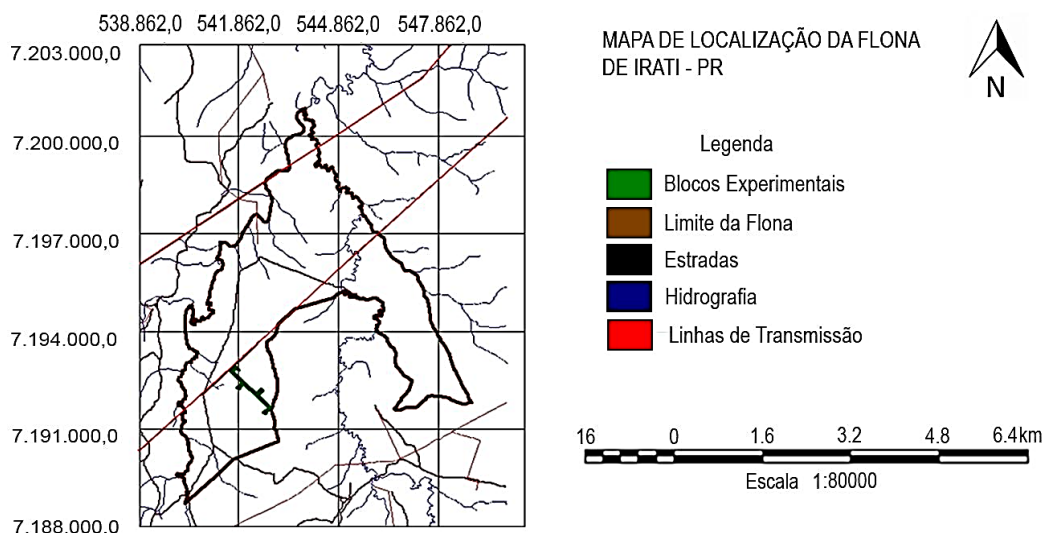
### Coleta e Processamento de Dados

O processo consistiu em medir o diâmetro da copa considerando sua projeção no solo, com auxílio de trena e bússola. Foram mensuradas 189 árvores de Araucária de forma aleatória e que estivessem no estrato superior dentro dos limites das parcelas permanentes existentes no local. Procedeu-se a medição de dois diâmetros de copa. O diâmetro  $D_1$  foi medido seguindo a orientação Norte/Sul, e o diâmetro  $D_2$  seguindo a orientação Leste/Oeste, para que todas as árvores tivessem a mesma metodologia de medição. Além dos diâmetros de copa também foram medidas as circunferências a 1,30 m (CAP) com uso de fita métrica comum, posteriormente dividido pelo valor de  $\pi$  para obtenção do diâmetro a altura do peito (DAP).

Com isso, utilizou-se a média aritmética dos diâmetros de copa mensurados para serem usados como variáveis dependentes na estimativa dos DAP's. Os dados foram processados com auxílio do *software Statistical Analysis System* versão 9.2<sup>®</sup>.



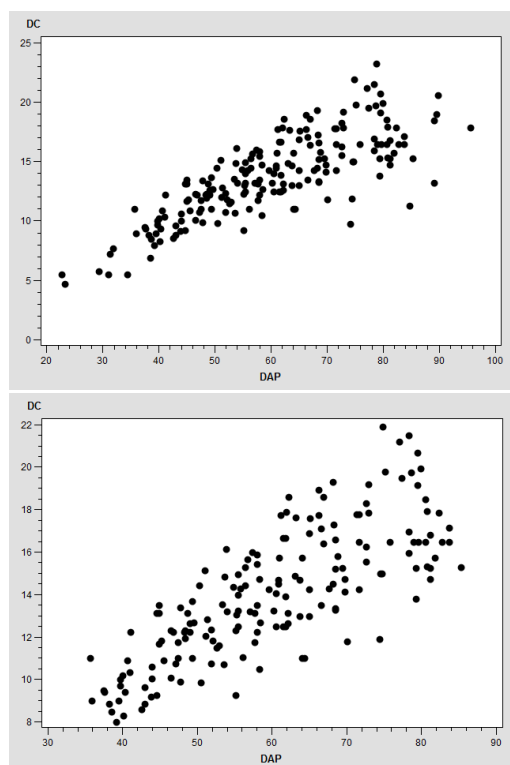
**Figura 1.** Localização da Microrregião de Irati - PR.



**Figura 2.** Localização da FLONA de Irati - PR (adaptado de DISPERATI et al., 2002).

### Análise exploratória e ajuste do modelo matemático

A análise exploratória dos dados foi empregada visando obter informações a respeito da estrutura dos dados, suas tendências, variações e possibilidade de detecção de *outliers* (Figura 3). Essas análises auxiliaram sobremaneira na escolha dos modelos a serem utilizados.



**Figura 3.** a) Distribuição dos dados diâmetro de copa (DC) e DAP com valores extremos; b) distribuição dos dados sem os valores extremos.

### Coeficiente de correlação de Pearson

O coeficiente de correlação permitiu verificar a correlação entre variáveis independentes e entre variável independente e dependente. Do ponto de vista de eficiência, é interessante que exista uma alta correlação entre a variável independente e a dependente e uma baixa ou inexistente correlação entre as variáveis independentes.

O conhecimento da correlação entre as variáveis é importante para a modelagem de uma regressão. A correlação linear simples foi utilizada para verificar a existência de relação entre as duas variáveis envolvidas (DC e DAP). A medida de correlação entre as duas variáveis é dada pelo coeficiente de correlação  $\rho_{xy}$ , definido pela equação (1):

$$\rho = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X) \cdot \text{var}(Y)}} \quad (1)$$

Em que:  $\bar{X}$  = média de X;  $\bar{Y}$  = média de Y;  $n$  = número de observações; e  $x_1, x_2, \dots, x_n$  e  $y_1, y_2, \dots, y_n$  = valores medidos de ambas as variáveis.

A análise de correlação indica a relação entre 2 variáveis lineares e os valores sempre serão entre -1 e +1, ou seja, mede o grau dessa correlação e o sentido (positiva ou negativa). O sinal indica a direção, se a correlação é positiva ou negativa, e o tamanho da variável indica a força da correlação. Hipóteses:  $H_0: \rho = 0$  (não há correlação linear) e  $H_1: \rho \neq 0$  (há correlação linear).

O teste para o coeficiente de correlação entre duas variáveis é realizado por meio do teste "t". Hipóteses:  $H_0: \rho = 0$  (não há correlação linear);  $H_1: \rho \neq 0$  ou  $p > 0$  ou  $p < 0$

(há correlação linear).

A estatística calculada dá-se da seguinte forma, conforme a equação (2):

$$t_c = \frac{r_{xy}\sqrt{n-2}}{\sqrt{1-r_{xy}^2}} \quad (2)$$

Em que:  $t_{tab}$  = valor da tabela de t de Student, com  $\alpha$  e  $v = n-2$  graus de liberdade.

Se a probabilidade desse resultado ter ocorrido for muito pequena, pode-se concluir que o resultado observado é estatisticamente relevante. Essa probabilidade também é chamada de *p-valor* ou valor p. Consequentemente, o nível de confiança  $\alpha$  é igual a  $1 - p\text{-valor}$  (LAPPONI, 2005).

Normalmente é usado um "ponto de corte" para o *p-valor* ou para o nível de confiança desejado para definir se a hipótese nula deve ser rejeitada ou não. Se o *p-valor* for menor que esse "ponto de corte", a hipótese nula é rejeitada, caso contrário, a hipótese nula não é rejeitada.

Desta forma, conclui-se que ao nível de confiança  $(1 - \alpha)$  % calculado for menor que o tabelado, que não há correlação linear se uma das três situações a seguir for identificada:

- $(-t_{tab} < t_c < t_{tab})$ ;
- $(t_c < t_{tab})$  e
- $(t_c > t_{tab})$  respectivamente, de acordo com as possíveis hipóteses de  $H_1$ .

### Detecção de outliers

Tukey (1977) propôs a utilização de um gráfico chamado *boxplot* que resume as seguintes medidas estatísticas: mediana, quartis inferior e superior, valores mínimos e máximos. Nesta pesquisa, o *boxplot* foi utilizado para avaliar a distribuição dos dados, sendo formado pelo primeiro e terceiro quartil, além da mediana.

Os quartis são valores que dividem o conjunto de dados em quatro partes de um mesmo conjunto de dados, sendo ordenados em ordem crescente. O cálculo inicia obtendo-se o valor da mediana, que divide o conjunto de dados em duas partes.

### Homogeneidade de variâncias

Foi avaliada a homocedasticidade das Variâncias por

meio do teste Qui-quadrado ( $X^2$ ) de Bartlett, o qual compara o valor do  $X^2_{calc}$  com o valor tabelado para  $(n-1)$  graus de liberdade (classes de DAP, estabelecidas pelo método proposto por Sturges) ao nível  $\alpha$  de probabilidade estabelecida ( $p\text{-valor} \leq 0,05$ ). A variável diâmetro à altura do peito foi distribuída em oito classes de frequência (Tabela 1).

**Tabela 1.** Classes de DAP definidas pelo método de Sturges.

Intervalo de Classe	Frequência	Frequência Acumulada	Percentual
35,0 f 41,3	16	16	9,25
41,3 f 47,6	19	35	10,98
47,6 f 53,9	25	60	14,45
53,9 f 60,2	27	87	15,61
60,2 f 66,5	28	115	16,18
66,5 f 72,8	24	139	13,87
72,8 f 79,1	15	154	8,67
79,1 f 85,4	19	173	10,98
<b>Total</b>	<b>173</b>		<b>100</b>

A transformação dos dados para a variável dependente é uma das possíveis formas de contornar o problema de dados que não obedecem aos pressupostos da análise de variância. A transformação Box-Cox automaticamente identifica uma transformação a partir de uma família de transformações  $Y' = Y^\lambda$  de potência de Y, conforme (SCHNEIDER et al., 2009). A família de transformações de potência é dada por:  $(\lambda = 2 \rightarrow Y' = Y^2)$ ;  $(\lambda = 0,5 \rightarrow Y' = \sqrt{Y})$ . A transformação Box-Cox consiste em transformar os dados com a expressão conforme a equação (3):

$$y' = \frac{y^\lambda - 1}{\lambda} \quad (3)$$

Em que:  $\lambda$  = parâmetro a ser estimado a partir dos dados da amostra da variável Y; se  $\lambda = 0$ , a equação se reduz a  $y' = \log(y)$ . Em que:  $\log$  = logaritmo neperiano.

Obtido o valor de  $\lambda$  (lâmbda), encontra-se os valores dos dados transformados conforme a equação (5) e utiliza-se estes dados transformados para efetuar a análise de regressão.

### Assimetria e curtose

As medidas de assimetria e curtose, complementaram

as medidas de posição e de dispersão e proporcionam uma descrição e compreensão mais completa das distribuições de frequências. Assimetria é o grau de desvio, afastamento da simetria ou grau de deformação de uma distribuição de frequências (COSTA NETO, 2002). O coeficiente de assimetria mede o grau de deformação da distribuição, sendo expresso conforme a equação (4):

$$C.A. = \frac{\bar{X} - Mo}{S} \quad (4)$$

Em que: C.A.= coeficiente de assimetria;  $\bar{X}$  = média;  $Mo$ = moda; e  $S$ = desvio padrão.

A interpretação dos dados dá-se da seguinte forma (COSTA NETO, 2002):

- Simetria para o intervalo de valores de -0,2 à 0,2;
- Assimetria fraca positiva para o intervalo de valores de 0,2 a 1,0;
- Assimetria forte positiva para valores > 1,0;
- Assimetria fraca negativa para o intervalo de valores de -0,2 a -1,0; e
- Assimetria forte negativa para valores < -1,0.

A curtose mede o grau de achatamento de uma curva em relação a curva normal, sendo esta tomada como padrão. Uma curva pode ser classificada como (COSTA NETO, 2002):

- Platicúrtica, quando a curva é mais achatada do que o normal ( $\sigma$  ou  $S$  são grandes);
- Mesocúrtica, quando a curva é normal ( $\sigma$  ou  $S$  são intermediários); e
- Leptocúrtica, quando a curva é mais alta do que o normal ( $\sigma$  ou  $S$  são pequenos).

O coeficiente de curtose é expresso conforme a equação (5):

$$K = \frac{Q3-Q1}{2(D9-D1)} \quad (5)$$

Em que:  $K$ = coeficiente de curtose;  $Q_1$ = primeiro quartil;  $Q_3$ = terceiro quartil;  $D_1$ = primeiro decil; e  $D_9$ = nono decil.

A interpretação dos dados dá-se da seguinte maneira (COSTA NETO, 2002):

- $K < 0,263$  curva leptocúrtica;

- $K = 0,263$  curva mesocúrtica; e
- $K > 0,263$  curva platicúrtica.

### Normalidade na distribuição dos resíduos

Para atender os supostos de normalidade, para uma variável contínua, a frequência dos resíduos deve seguir a distribuição normal, a qual garante a aplicação do teste de significância para esta distribuição (SCHNEIDER et al., 2009).

É necessário verificar se a distribuição das frequências observadas se ajusta ao modelo teórico pré-determinado. Segundo Prodan (1968), a normalidade pode ser verificada com o qui-quadrado. Segundo Schneider et al. (2009), o teste de aderência de Kolmogorov-Smirnov, assim como o teste de Shapiro-Wilk podem ser utilizados para verificar a normalidade dos resíduos.

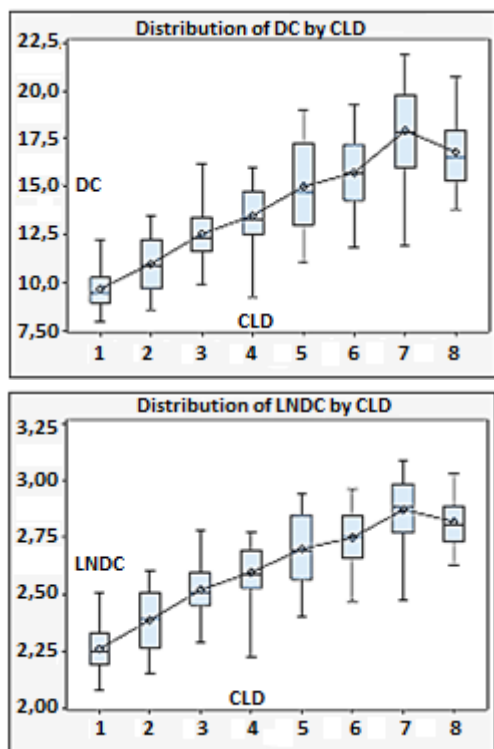
### Independência dos resíduos

Para Scolforo (2005), os resíduos gerados por uma equação não devem apresentar autocorrelação, ou seja, devem ser independentes. A autocorrelação indica que existe dependência entre os resíduos ou que os resíduos são autocorrelacionados entre si. Neste caso, as estimativas dos mínimos quadrados dos parâmetros não apresentam variância mínima, além do erro padrão ser distorcido e desta forma, as estimativas se tornam ineficientes e os intervalos de confiança gerados são incorretos. Nesta pesquisa foi utilizado o teste de Durbin-Watson para verificar a independência dos resíduos.

### RESULTADOS E DISCUSSÃO

A análise exploratória dos dados foi utilizada para detectar possíveis *outliers* e verificar a tendência normal dos dados. Após esta verificação, foram eliminados 16 pares de dados e utilizado o modelo de linear, ou seja,  $DC = b_0 + b_1 * DAP$ .

O coeficiente de correlação de Pearson apresentou valor de 0,78 ( $0 < p_{xy} < 1$ ), indicando correlação forte e positiva. O valor apresentado foi verificado por meio do teste “t” de significância da correlação linear entre as duas variáveis. O teste apresentou  $t_{cal} > t_{tab}$ , rejeitando a hipótese de nulidade, ou seja, comprovando a correlação linear entre DAP e DC. O gráfico *boxplot*, também não apresentou valores discrepantes, conforme pode ser observado na Figura 4.



**Figura 4.** a) Boxplot para as variáveis DC= diâmetro de copa e CLD= classes de DAP; e b) Boxplot para as variáveis LNDC= logaritmo neperiano do diâmetro de copa e CLD.

O resultado do teste do  $\chi^2$  de Bartlett indicou ocorrência de heterocedasticidade, com o valor do  $\chi^2_{\text{calc}}$  maior que o  $\chi^2_{\text{tab}}$  a 95% de probabilidade, conforme apresentado na Tabela 2.

**Tabela 2.** Teste de Bartlett, para variável dependente diâmetro de copa.

G.L.	$\chi^2_{\text{calc}}$	$\chi^2_{\text{tab}}$ 95% Prob.
n-1= 7	20,39	14,07

Em que:  $\chi^2$  calculado >  $\chi^2$  tabelada significa que há heterogeneidade de variâncias.

A apresentação dos resultados da Tabela 2, indica que as variâncias dos dados não são constantes, necessitando utilizar transformações para a variável dependente (Y) por meio do método Box-Cox.

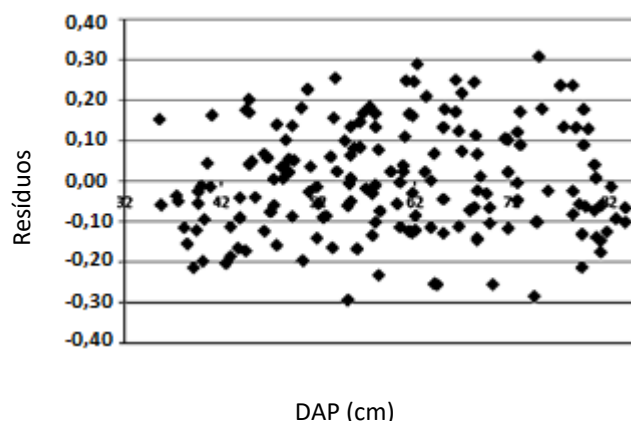
O coeficiente de assimetria de Pearson apresentou valor de -0,2556, ou seja, para esta intensidade negativa a assimetria é fraca e à esquerda devido ao sinal negativo. O coeficiente negativo é interpretado da seguinte forma:  $\bar{X}$  60,03 <  $M_d$  60,16 <  $M_o$  63,35 ( $\bar{X}$ = média aritmética;  $M_d$ = mediana;  $M_o$ = moda).

O coeficiente de curtose apresentou valor de 0,2806, ou seja, para esta intensidade a curva é platicúrtica, onde

a curva é mais achatada que o normal.

Nenhuma análise de regressão deve ser aceita antes que se faça uma análise gráfica prévia dos resíduos. Esta análise indica se o ajuste foi bom em toda a linha da estimativa, se não ocorre tendenciosidades. Os resíduos são independentes se a variância for homogênea para os valores estimados.

Na verificação do modelo linear, o teste de Bartlett aponta heterogeneidade de variâncias, onde o resultado apresenta  $\chi^2$  calculado >  $\chi^2$  tabelado. Na análise gráfica dos resíduos não apresenta tendência na distribuição dos resíduos, conforme pode ser visualizado na Figura 5.



**Figura 5.** Tendência da distribuição de resíduos com variâncias heterogêneas.

Em relação ao teste de White, o mesmo resultou em um valor de  $\chi^2$  de 17,18, significativo ( $p= 0,0002$ ), portanto, os dados de diâmetro de copa apresentam heterogeneidade de variância, devendo ser transformados.

O resultado do teste de Shapiro-Wilk indicou que as amostras não provêm de uma população normal, ou seja, não atendem a condição de normalidade com o valor de  $W_{\text{calc}}$  maior que o  $W_{\text{tab}}$  a 95% de probabilidade, conforme apresentado na Tabela 3.

**Tabela 3.** Teste de Shapiro-Wilk para variável dependente diâmetro de copa.

G.L.	$W_{\text{calc}}$	$W_{\text{tab}}$ 95% Prob.
n-1= 7	0,98	0,818

Em que: G.L.= graus de Liberdade; W calculado > W tabelado significa que a amostra não provém de uma população normal.

No teste de Durbin-Watson estabeleceu-se o valor máximo de aceitação ao nível  $\alpha$  0,05 de probabilidade, para 173 pares de observações, considerando apenas o coeficiente b1 e b2. O resultado do teste apresentou um



valor  $\alpha = 0,1103$  para o coeficiente  $b_0$ ,  $\alpha = 0,0001$  para o coeficiente  $b_1$  e  $\alpha = 0,0088$  para o coeficiente  $b_2$ , o que neste caso, somente o coeficiente  $b_0$  ultrapassa o valor máximo estabelecido de  $\alpha = 0,05$ , devendo ser retirado do modelo.

A prova de significância rejeitou a hipótese de nulidade, ou seja, existe correlação em série. Quando os  $E_i$  (erros) são dependentes, apresentam uma correlação em série, neste caso, o método dos mínimos quadrados pode não dar as melhores estimativas. Para evitar a correlação em série, deveriam ser obtidas mais observações de DAP e DC, ou seja, deve-se aumentar o tamanho da amostra, mas essa prática nem sempre é possível ou nem sempre altera os resultados.

O modelo linear resultou na seguinte equação:  $DC = 2,98825 + 0,18382 * DAP$ , em que:  $Y_i$  = diâmetro de copa; DAP = diâmetro a altura do peito. A Tabela 4 mostra as estatísticas do modelo.

**Tabela 4.** Análise estatística do modelo de regressão linear.

N	R <sup>2</sup>	R <sup>2</sup> ajust.	CV%	F
173	0,6092	0,6070	13,68	266,62

Em que: N= número de indivíduos; R<sup>2</sup>= coeficiente de determinação; R<sup>2</sup> ajust.= coeficiente de determinação ajustado; CV%= coeficiente de variação; e F= distribuição F.

Constatada a heterogeneidade de variância, utilizou-se a transformação dos dados pelo método *Box-Cox*, encontrando a potência ( $\lambda=0$ ) para a variável dependente, ou seja, a melhor transformação da variável é por meio do logaritmo neperiano da variável Y.

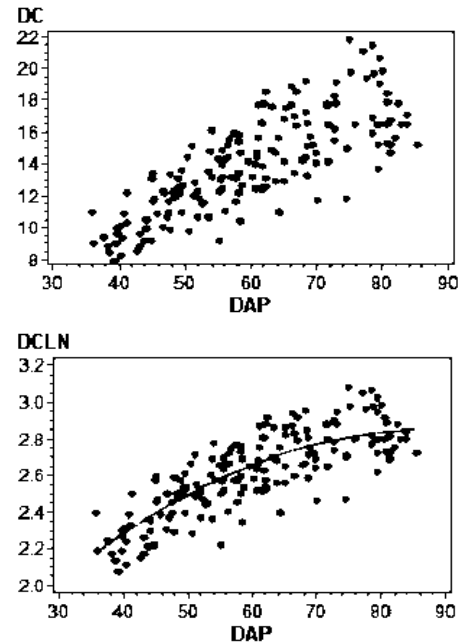
A transformação da variável dependente buscou reduzir a heterogeneidade de variância, conforme a Figura 6 a e b. Conforme pode ser verificado na Figura 6 b, o modelo a ser utilizado passa a ser o modelo polinomial do segundo grau, escrito da seguinte forma:  $LNDC = b_0 + b_1 * DAP + b_2 * DAP^2$ .

Após utilizada a transformação *Box-Cox*, o problema de heterogeneidade de variâncias foi eliminado, resultando em  $X^2$  calculado <  $X^2$  tabelado, conforme é apresentado na Tabela 5.

**Tabela 5.** Teste de Bartlett para a variável dependente  $\ln DC$ .

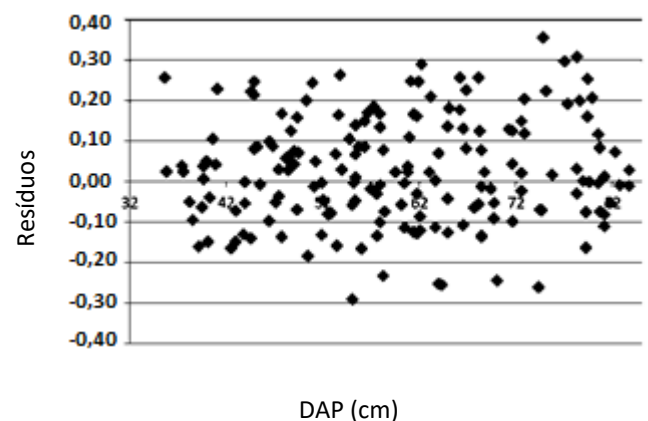
G.L.	$X^2_{calc}$	$X^2_{tab}$ 95% Prob.
n-1= 7	6,91	14,07

Em que: G.L.= graus de liberdade;  $X^2$  calculado <  $X^2$  tabelado significa que há homogeneidade de variâncias.



**Figura 6.** a) Distribuição dos dados não transformados e b) distribuição dos dados transformados pelo método *Box-Cox*.

Na verificação do modelo polinomial do segundo grau, o teste de Bartlett aponta homogeneidade de variâncias, onde o resultado apresenta  $X^2$  calculado <  $X^2$  tabelado. Na análise gráfica dos resíduos não apresenta tendência na distribuição dos resíduos, conforme pode ser visualizado na Figura 7.



**Figura 7.** Tendência da distribuição de resíduos com variâncias heterogêneas.

O teste de White resultou em um valor de  $X^2$  de 5,72, significativo ( $p = 0,2211$ ), portanto, os dados de diâmetro de copa apresentam homogeneidade de variância, ou seja, a transformação *Box-Cox* resolveu o problema de heterogeneidade.

O resultado do teste de Shapiro-Wilk-W indicou que as amostras provêm de uma população normal, ou seja,



atendem a condição de normalidade com o valor de  $W_{calc}$  maior que o  $W_{tab}$  a 95% de probabilidade, conforme apresentado na Tabela 6.

**Tabela 6.** Teste de Shapiro-Wilk para a variável dependente diâmetro de copa transformada.

G.L.	$W_{calc}$	$W_{tab}$ 95% Prob.
n-1= 7	0,97	0,979

Em que: G.L.= graus de liberdade; W calculado < W tabelado, significa que a amostra provém de uma população normal; e n= número de classes.

O teste de Durbin-Watson após transformação da variável dependente, apresentou um valor  $\alpha = 0,0092$ , o que não ultrapassou os valores estabelecidos de  $\alpha = 0,05$ . A prova de significância aceitou a hipótese de nulidade, ou seja, não existe correlação em série, indicando que a transformação utilizada foi suficiente para eliminar a correlação em série.

O modelo polinomial do segundo grau resultou na equação:  $LNDC = 0,944 + 0,04327 \cdot DAP - 0,00024523 \cdot DAP^2$ , em que:  $Y_i$ = diâmetro de copa; DAP= diâmetro a altura do peito e  $DAP^2$ = diâmetro a altura do peito ao quadrado. A Tabela 7 mostra as estatísticas do modelo.

**Tabela 7.** Análise estatística do modelo de regressão transformado.

N	$R^2$	$R^2$ ajust.	CV%	F
173	0,6582	0,6542	5,02	163,68

Em que: N= número de indivíduos;  $R^2$ = coeficiente de determinação;  $R^2$  ajust.= coeficiente de determinação ajustado; CV%= coeficiente de variação; e F= distribuição F.

## CONCLUSÕES

É comum, em estudos de regressão, autores simplesmente escolherem modelos matemáticos e aplicá-los em seus estudos, sem ao menos verificar a tendência natural que este conjunto de dados segue. Esta prática é um grave problema e pode influenciar as estatísticas de precisão de uma equação.

Em muitos estudos biológicos (variáveis relacionadas ao estudo de florestas), os dados nem sempre apresentam homogeneidade de variância. No entanto, esta heterogeneidade pode ser removida pelo uso de equações ponderadas, transformações logarítmicas ou transformações Box-Cox.

Geralmente, uma variável contínua, como DAP, CAP, altura total, altura comercial, volume, área basal, apresenta distribuição normal. Porém, para cumprir o

pressuposto de normalidade, testes de significância devem ser utilizados.

Na tentativa de ajustar um conjunto de dados por meio de uma equação linear quando possui tendência polinomial, os resíduos podem apresentar-se de forma tendenciosa, mal distribuídos e correlacionados, fazendo com que o modelo não atenda a suposição de independência.

A relação diâmetro de copa e DAP apresentou heterogeneidade na análise de variância, sendo removida por meio da transformação Box-Cox. A transformação da variável dependente trouxe melhora nas estatísticas de precisão do modelo, coeficiente de determinação, coeficiente de determinação ajustado, erro padrão da estimativa e coeficiente de variação.

## AGRADECIMENTOS

Ao professor Attilio Antonio Disperati (*in memoriam*), ao Programa de Pós-Graduação em Engenharia Florestal da UFSM e aos amigos Wilian José Kuchla, Ademar Luiz Chiquetto e Jonatan dos Santos Lisboa, na valiosa colaboração na coleta de dados.

## REFERÊNCIAS

- BRENA, D.A. et al. Metodologia para verificação das condicionantes da análise de regressão. *Floresta*, v.9, n.2, p.25-45, 1978.
- COSTA, E.A. *Influência de variáveis dendrométricas e morfométricas da copa no incremento periódico de Araucaria angustifolia* (Bertol.) Kuntze, Lages, SC. 2011. 148p. (Dissertação de mestrado).
- DISPERATI, A.A. et al. Delineação visual de copas de árvores em fotografias aéreas de alta resolução de uma área com vegetação natural de Araucária: resultados parciais. *5º Seminário de Atualização em Sensoriamento Remoto e Sistemas de Informações Geográficas Aplicados à Engenharia Florestal*, p. 241-242, 2002.
- IBGE. *Manual técnico da vegetação brasileira*. Rio de Janeiro: Série Manuais Técnicos em Geociências, 1992.
- LISBOA, G.S. *Fotografias aéreas de escala grande e imagem IKONOS-2 no mapeamento de copas de Araucaria angustifolia*. 2009. 104p. (Dissertação de mestrado).
- MAZZA, C.A.S. *Caracterização ambiental da paisagem da Microregião Colonial de Irati e zoneamento ambiental da Floresta Nacional de Irati*, PR. São Carlos: UFSCar, 2006.
- PRODAN, M. *Forest biometrics*. New York: Pergamon Press, 1968.
- GOTTELI, N.J.; ELLISON, A.M. *Princípios de estatística em*

**ecologia**. Porto Alegre: Artmed, 2011.

LAPPONI, J.C. **Estatística usando Excel**. 4. ed. Rio de Janeiro: Elsevier, 2005.

ORELLANA, E.; KOEHLER, B.A. Relações morfométricas de *Ocotea odorifera* (Vell) Rohwer. **Revista Acadêmica: Ciências Agrárias e Ambientais**, v.6, n.2, p.229-237, 2008.

SANQUETTA, C.R. **Perspectivas da recuperação e do manejo sustentável das florestas de araucária**. 2005. Disponível em: <http://www.comciencia.br/reportagens/2005/08/09.shtm>

SANQUETTA C.R. et al. Crown area and trunk diameter relationship for tree species at a Mixed Araucaria natural forest in the mid-southern Parana state, Brazil. **Floresta**, v.41, n.1, p.63-72, 2011.

SANQUETTA, R.C. et al. Equações para estimativa do diâmetro de copa para acácia-negra. **Floresta e Ambiente**, v.21, n.2, p.192-205, 2014.

SCOLFORO, J.R.S. **Biometria florestal**. Lavras: UFLA/FAEPE, 2005.

SCHNEIDER, P.R.; FINGER, C.A.G. **Manejo sustentado de florestas inequidâneas heterogêneas**. Santa Maria: UFSM, 2000.

SCHNEIDER, P.R. et al. **Análise de regressão aplicada à engenharia florestal**. 2. ed. Santa Maria: Editora FACOS, 2009.

TUKEY, J.W. **Exploratory data analysis**. Addison-Wesley, 1977.

VELOSO, H.P.; GÓES-FILHO, L. **Fitogeografia brasileira, classificação fisionômica ecológica da vegetação neotropical**. Salvador: Ministério de Minas e Energia-Projeto RADAMBRASIL, 1982.