AUMENTO DA PRECISÃO DE MODELOS VOLUMÉTRICOS ATRAVÉS DO USO DA TRANSFORMAÇÃO DE BOX E COX

José A. Aleixo da Silva[1]
Sebastião Amaral Machado[2]
Bruce E. Borders[3]
Robert L. Bailey[3]

RESUMO - O uso da família de transformações de dados proposta por Box e Cox (1964) foi usada no modelo $V_i = \beta_0 + \beta_1 DAP + \beta_2 H + \epsilon_i$ que foi comparado com os modelos de Schumacher e Hall e o da variável combinada (Spurr), através de uma análise de variância, considerando as porcentagens residuais resultantes de cada modelo testado. O uso de tal transformação, além de aumentar significativamente a precisão do modelo, também resultou no menor valor das porcentagens residuais.

PALAVRAS-CHAVES: Transformação de Box e Cox, modelos volumétricos.

INCREASING PRECISION IN VOLUMETRIC MODELS BY USING THE BOX AND COX TRANSFORMATION

ABSTRACT - The family of transformations proposed by Box and Cox (1964) was applied to the model $V_i = \beta_0 + \beta_1 DAP + \beta_2 H + \epsilon_i$. This model was compared with the combined variable model and the model proposed by Schumacher and Hall. The transformation significantly improved the precision on the model.

KEY-WORDS: The transformation of Box and Cox, volumetric equations.

INTRODUÇÃO

O uso de modelos de regressão linear é muito comum na estimativa de volumes individuais de árvores ou de povoamentos florestais. O emprego de técnicas de regressão linear envolve certos requisitos tais como: distribuição normal dos erros, variância constante para a distribuição dos erros, independência dos erros e não existência de colinearidade entre as variáveis independentes no caso de regressão linear múltipla (Kelly e Beltz, 1987). m inventários florestais, quando se amostram pequenas e grandes árvores, torna-se muito comum a variância não ser constante porque pequenas árvores tendem a variar menos em volume que maiores árvores. O uso do método dos mínimos quadrados ponderados é uma tentativa para homogeneizar a variância dos erros em tais amostras (Cunia, 1964; Jacobs e Monteith, 1981).

A aplicação de logarítmos em ambos os membros do modelo:

$$V_i = \beta_0 \cdot DAP^{\beta_1} \cdot H^{\beta_2} \cdot \epsilon_i$$

é uma tentativa de normalizar a distribuição dos erros e permitir que se possa utilizar o método linear dos mínimos quadrados na estimativa dos parâmetros do modelo (Silva e Bailey, 1991). Outras transformações como: raiz quadrada, quadrática, exponencial e recíproca também são comumente usadas (Dolby, 1963). Como cada amostra de árvores pode apresentar diferentes distribuições de erros, o problema da escolha da transformação ideal para a amostra torna-se complexo. Box e Cox (1964) desenvolveram uma família de transformações para qualquer variável dependente positiva que tem forma:

$$(Y^{\lambda} - 1)/\lambda, \text{ para } \lambda \neq 0$$

$$W = \ln Y, \text{ para } \lambda = 0$$

1 Departaento de Engenharia Florestal, UFRPE, 52.171-900, Recife, PE.
2 Departamento de Silvicultura e Manejo, UFPR, CP. 2959, 80.035-010 - Curitiba - PR.
3 School of Forest Resources, University of Georgia, Athens-Ga, USA.

que depende diretamente do parâmetro λ para qualquer conjunto de dados, bem como do vetor de parâmetros β para o modelo a ser ajustado.

$$W = X\beta + \in$$

em que $W = (W_1, W_2, ..., W_n)$.

O valor de λ que maximize:

$$K_{max}\lambda = -\frac{n}{2} . \ln \left(\frac{SQR}{n}\right) + (\lambda - 1)\sum_{i=1}^{n} \ln Y_i \quad (1)$$

onde:

SQR = soma de quadrados dos resíduos;

n = número de observações;

ln = logarítmo neperiano.

$K_{max\lambda}$ = é o estimador de máxima verossimilhança de λ (Silva e Bailey, 1991). Este tipo de transformação foi citado pela primeira vez no meio florestal por Shreuder e Swank (1971) e usado com sucesso por (Silva, 1986; Meng e Tsai, 1986; Kelly e Beltz, 1987; Silva e Bailey, 1991; Silva et al., 1992). Em sua maneira tradicional de desenvolvimento, tal família de transformações exige que seja executada um número de análises de regressões lineares igual ao número de valores de λ's a serem adotados, além do traçado de um gráfico para identificar o melhor estimador de λ que maximiza (1). Portanto pode ser um processo demorado dependendo da amplitude adotada para λ.

Neste trabalho não se usa a maneira tradicional de estimar λ e sim através de um modelo não linear que corresponde ao modelo:

$$V_i = \beta_0 + \beta_1 DAP + \beta_2 H + \in_i \; ..... \; (2)$$

transformado pelo método de Box e Cox (1964) resultando em:

$$V_i = [\lambda\,(\beta_0 + \beta_1 DAP_i + \beta_2 H_i) + 1]^{1/\lambda} + \in_i$$

MATERIAL E MÉTODOS

Neste trabalho foram utilizadas 189 árvores de *Pinus elliottii* provenientes de plantações no nordeste do planalto de Santa Catarina e sudeste do primeiro planalto do Paraná, pertencentes à CONFLORESTA. O volume total com casca de cada árvore foi determinado pelo método de Smalian. A distribuição das árvores, por classes de diâmetro e altura, se encontra na Tabela 1.

Para comparar com modelo (2) em sua forma original e transformada, escolheu-se os modelos de Schumacher e Hall (1933) e o da variável combinada (Spurr, 1952) que têm se distinguido entre os inúmeros modelos usados em inventários florestais. Portanto, quatro modelos foram testados:

a) $V_i = \beta_0 + \beta_1 DAP_i + \beta_2 H_i + \in_i$

Aplicando a transformação de Box e Cox em tal modelo e considerando o erro como aditivo, resulta em:

$$V_i = [\lambda\,(\beta_0 + \beta_1 DAP_i + \beta_2 H_i) + 1]^{1/\lambda} + \in_i$$

que pode ser escrito como:

b) $V_i = (\theta_0 + \theta_1 \,.\, DAP_i + \theta_2 \,.\, H_i)^{\lambda'} + \in_i$

onde:

$\theta_0 = \lambda \,.\, \beta_0 + 1$

TABELA 1. Distribuição das árvores por classes de diâmetro e altura.

Classes DAP's (cm)	Classes de Alturas (m)						Total
	08 | −10	10 | −12	12 | −14	14 | −16	16 | −18	18 | −20	
08 | −10	1	1					2
10 | −12		5	1				6
12 | −14		7	1				8
14 | −16		8	12	4			24
16 | −18		3	11	7			21
18 | −20		1	10	16	4		31
20 | −22		1	9	13	4		27
22 | −24			6	15	7	1	30
24 | −26			5	11	9	2	27
26 | −28			1	5	4		10
28 | −30					2	1	3
Total	1	27	56	71	30	4	189

$\theta_1 = \lambda \cdot \beta_1$
$\theta_2 = \lambda \cdot \beta_2$
$\lambda' = 1/\lambda$

c) $V_i = \beta_0 + \beta_1 DAP_i H_i + \in_i$

d) $V_i = \beta_0 \cdot DAP^{\beta_1} \cdot H^{\beta_2} \cdot \in_i$

Os modelos (b) e (d) foram ajustados pelo algorítmo de Marquardt (Marquardt, 1963) no pacote estatístico SAS (SAS, 1988). Os resultados dos coeficientes de determinação e erros padrões residuais estão na Tabela 1. Após ajustados os modelos foram calculados para cada equação resultante as porcentagens residuais da seguinte forma.

$$PR_i = \frac{VOL_i - \hat{VOL}_{ij}}{VOL_i} \cdot 100$$

onde:
PR_i = porcentagem residual para árvore i;
VOL_i = volume real da árvore i;
$\hat{VOL}_{ij}$ = volume estimado da árvore i pelo método j.

Com os valores das PR_i's foi realizada uma análise da variância inteiramente ao acaso, onde os tratamentos foram respectivamente as PR_i's calculadas para cada equação resultante, com a finalidade de verificar se existia diferença(s) significativa(s) entre os modelos testados.

RESULTADOS E DISCUSSÃO

Os resultados das análises de regressão linear e não linear estão presentes na Tabela 2.

onde:
R^2 = coeficiente de determinação
EPR = erro padrão residual

Com base em tais resultados, observa-se que o emprego da transformação de Box e Cox aumentou consideravelmente a precisão do modelo linear múltiplo tornando-o semelhante ao de Schumacher e Hall e o da variável combinada. Caso o valor de λ não fosse significativamente diferente de 1,0 (um), isto é, este valor estivesse presente no intervalo de confiança para a estimativa de λ, não haveria necessidade de transformar o modelo (1). Caso o valor 0 (zero) estivesse contido no intervalo de confiança a transformação que deveria ser empregada seria ln Y_i.

Após realizados os cálculos das porcentagens residuais para cada equação resultante, foram obtidos os seguintes resultados contidos na Tabela 3.

TABELA 3. Valores das diferenças agregadas para as equações resultantes

Modelo	Médias das diferenças agregadas
Linear múltiplo	5.792 a
Transformado	-0.153 b
Schumacher e Hall	-0.276 b
Variável combinada	-0.567 b

Médias unidas pela mesma letra não diferem entre si pelo teste de Tukey a nível de 0.1 de probabilidades.

TABELA 2. Valores dos coeficientes de determinações e erros padrões residuais para os modelos testados.

Modelos	Coeficientes	R^2	EPR
Linear múltiplo	β_0 = 0.33179193	0.9371	0.027253
	β_1 = 0.01933864		
	β_2 = 0,01313155		
Transformado	Θ_1 = -0.04868927	0.99856	0.013114
	Θ_2 = 0.02009783		
	Θ_3 = 0.01633837		
	γ' = 2.79818229		
Variável combinada	β_0 = 0.00485448	0.9847	0.013392
	β_1 = 0.00003892		
Schumacher e Hall	β_0 = 0.00004101	0.9857	0.012981
	β_1 = 1.86964461		
	β_2 = 1.13347353		

Neste caso observa-se que mesmo não diferindo dos modelos de Schumacher e Hall e da variável combinada, o modelo transformado foi o que apresentou um menor valor médio das porcentagens residuais, o que certamente o torna mais preciso na estimativa dos volumes para a amostra usada neste estudo, graças ao uso da transformação de Box e Cox.

A equação resultante é:

$$V_i = [-0{,}04868927 + 0{,}02009783\ DAP_i + 0{,}016133837\ H_i]^{2{,}79818229}$$

Tal transformação permite ao pesquisador indicar se os dados da amostra necessitam ou não serem transformados, bem como qual o tipo de transformação a ser usada, pois dependendo do valor de λ várias transformações são consideradas. Este tipo de transformação pode ser usado em qualquer conjunto de dados onde a variável dependente é sempre positiva, sendo portanto uma ótima opção em outros ramos da ciência florestal.

REFERÊNCIAS BIBLIOGRÁFICAS

BOX, G.E.; COX, D.R. An analysis of transformations. **Journal of the Royal Statisical Society,** London, B-26, p.211-243, 1964.

CUNIA, T. Weighted least squares method and construction of volume tables. **Forest Science**, Washington, v.10, p.180-191, 1964.

DOLBY, J.L. A quick method for choosing a transformation. **Technometrics**, Richmond, v.5, p.317-325, 1963.

JACOBS, M.W.; MONTEITH, D.B. Feasibility of developing regional Weight tables. **Journal of Forestry,** Washington, v.76, p.676-677, 1981.

KELLY, J.F.; BELTZ, R.C. A comparison of tree volume estimation models for forest inventory. **USDA Forest Service. Research Paper SO-233,** St. Paul, 1987. 9p.

MARQUARDT, D.W. An algorithm for least-squares estimation of non-linear parameters. **Journal of the society for industrial and Applied Mahematics,** Philadelphia, v.2, n.2, p.431-441,1963.

MENG, C.H.; TSAI, W.Y. Selection of weight for weighted regression models for forest inventory. **Canadian Jorunal of Forst Research,** Otawa, v.16, p.671-673, 1986.

SCHREUDER, H.T.; SWANK, W.T. A comparison of several statistical models in forest biomass and surface area estimation. In: INTERNATIONAL UNION OF FOREST RESEARCH ORGANIZATIONS, SECTON 25: Yield and Growth, Maine: University of Maine, 1971. p.125-136.

SCHUMACHER, F.X. Yield, staud, and volume tables for Douglas fir in California. California: California Agricultural Experiment Station, 1930.(Bulletin, 491).

SAS Institute INC. SAS/STAT USERS GUIDE 4th Ed. CARY, 1989a. v.2, p.846 (Verniou, 6).

SILVA, J.A.A. da. **Dynamics of stand structure in fertilized slash pine plantation.** Georgia: University of Georgia/School of Forest Resources, 1986. 139p. (PhD. Dissertation).

SILVA, J.A.A. da; BAILEY, R.L. Uso de transformações normalizadoras no ajuste de modelos volumétricos. **Revista Árvore**, Viçosa, v.15, n.12, p.199-206, 1991.

SILVA, J.A.A. da; BORDERS, B.E.; BRISTER, G.H. 1992. A tree volume equation based on two lower stem diameter for Caribbean Pinea in Sri Lanka. **Commonwealth Forestry Review**, London, v.7, n.12, p.114-116, 1992.

SPURR, S.H. **Forestry inventory**. New York, Ronald Press, 1952. 476p.