

Método da Variável Fictícia para Ajuste de Modelos Volumétricos Estáveis e Compatíveis em Povoamentos Florestais

$$[ln(V_1)]D_1 + [ln(V_2)]D_2 + [ln(V_3)]D_3 + [ln(V_4)]D_4 = 8,87276.D_1 - 9,45063.D_2 - 9,07965.D_3 - 9,65695.D_4 + 1,80842.(ln(DAP)).D_1 + 1,81779.(ln(DAP)).D_2 + 1,87621.(ln(DAP)).D_3 + 1,88501.(ln(DAP)).D_4 + 0,667632.(ln(Ht)).D_1 + 0,800847.(ln(Ht)).D_2 + 0,674498.(ln(Ht)).D_3 + 0,808362.(ln(Ht)).D_4$$

$$DFFITS_i = \left(\frac{h_{ii}}{1-h_{ii}} \right) \times \left(\frac{e_i}{\sqrt{S^2(1-h_{ii})}} \right)$$

$$H = \begin{bmatrix} h_{11} & h_{12} & h_{13} & h_{14} & \dots & h_{1n} \\ h_{21} & h_{22} & h_{23} & h_{24} & \dots & h_{2n} \\ h_{31} & h_{32} & h_{33} & h_{34} & \dots & h_{3n} \\ h_{41} & h_{42} & h_{43} & h_{44} & \dots & h_{4n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ h_{n1} & h_{n2} & h_{n3} & h_{n4} & \dots & h_{ni} \end{bmatrix}$$

$$e_{padronizado} = \frac{Y_i - \hat{Y}}{\sqrt{\frac{SQ_{Res}}{(n-(m)-1)}}}$$

$$\begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ X_{12} & X_{22} & X_{32} & \dots & X_{n2} \end{bmatrix} \times \begin{bmatrix} 1 & X_{12} \\ 1 & X_{22} \\ 1 & X_{32} \\ \dots & \dots \\ 1 & X_{n2} \end{bmatrix} \Rightarrow X'X = \begin{bmatrix} n & \sum_{i=1}^n X_{i2} \\ \sum_{i=1}^n X_{i2} & \sum_{i=1}^n X_{i2}^2 \end{bmatrix}$$

$$Cook's.D_{(i)} = \left(\frac{1}{m} \right) \times \left(\frac{h_{ii}}{(1-h_{ii})} \right) \times \left(\frac{e_i^2}{S^2(1-h_{ii})} \right)$$

$$R^2 = \left[1 - \left(\frac{SQ_{Residuo}}{SQ_{Total}} \right) \right]$$

$$PRESS_p = \sum_{i=1}^n \left(Y_i - \hat{Y}_{i(i)} \right)^2$$

$$S_{yx} = \sqrt{\frac{SQ_{Res}}{(n-(m)-1)}} \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = (X'X)^{-1} X'Y$$

$$X = \begin{bmatrix} 1 & X_{12} \\ 1 & X_{22} \\ 1 & X_{32} \\ 1 & X_{42} \\ \dots & \dots \\ 1 & X_{n2} \end{bmatrix}$$

República Federativa do Brasil

Luiz Inácio Lula da Silva

Presidente

Ministério da Agricultura, Pecuária e Abastecimento

Roberto Rodrigues

Ministro

Empresa Brasileira de Pesquisa Agropecuária – Embrapa

Conselho de Administração

Luís Carlos Guedes Pinto

Presidente

Silvio Crestana

Vice-Presidente

Alexandre Kalil Pires

Hélio Tollini

Ernesto Paterniani

Marcelo Barbosa Saintive

Membros

Diretoria-Executiva da Embrapa

Silvio Crestana

Diretor-Presidente

Tatiana Deane de Abreu Sá

José Geraldo Eugênio de França

Kepler Euclides Filho

Diretores-Executivos

Embrapa Acre

Marcus Vinício Neves d'Oliveira

Chefe-Geral

Milcíades Heitor de Abreu Pardo

Chefe-Adjunto de Administração

Luís Cláudio de Oliveira

Chefe-Adjunto de Pesquisa e Desenvolvimento

Francisco de Assis Correa Silva

Chefe-Adjunto de Comunicação, Negócios e Apoio

ISSN 0104-9046

Junho, 2005

**Empresa Brasileira de Pesquisa Agropecuária
Centro de Pesquisa Agroflorestal do Acre
Ministério da Agricultura, Pecuária e Abastecimento**

Documentos 95

Método da Variável Fictícia para Ajuste de Modelos Volumétricos Estáveis e Compatíveis em Povoamentos Florestais

Evandro Orfanó Figueiredo

Rio Branco, AC
2005

Exemplares desta publicação podem ser adquiridos na:

Embrapa Acre

Rodovia BR 364, km 14, sentido Rio Branco/Porto Velho

Caixa Postal, 321

Rio Branco, AC, CEP 69908-970

Fone: (68) 212-3200

Fax: (68) 212-3284

<http://www.cpafac.embrapa.br>

sac@cpafac.embrapa.br

Comitê de Publicações da Unidade

Presidente: *Rivaldalve Coelho Gonçalves **

Secretária-Executiva: *Suely Moreira de Melo*

Membros: *Carlos Mauricio Soares de Andrade, Celso Luís Bergo, Claudenor Pinho de Sá, Cleisa Brasil da Cunha Cartaxo, Henrique José Borges de Araujo *, João Alencar de Sousa, Jonny Everson Scherwinski Pereira, José Tadeu de Souza Marinho, Lúcia Helena de Oliveira Wadt, Luís Cláudio de Oliveira, Marcílio José Thomazini, Patrícia Maria Drumond*

* Revisores deste trabalho

Supervisão editorial: *Claudia Carvalho Sena / Suely Moreira de Melo*

Revisão de texto: *Claudia Carvalho Sena / Suely Moreira de Melo*

Normalização bibliográfica: *Luiza de Marillac Pompeu Braga Gonçalves*

Tratamento de ilustrações: *Fernando Farias Sevá / Iuri Rudá Franca Gomes*

Editoração eletrônica: *Fernando Farias Sevá / Iuri Rudá Franca Gomes*

1ª edição

1ª impressão (2005): 300 exemplares

Todos os direitos reservados.

A reprodução não autorizada desta publicação, no todo ou em parte, constitui violação dos direitos autorais (Lei nº 9.610).

Dados Internacionais de Catalogação na Publicação (CIP).

Embrapa Acre.

F475m Figueiredo, Evandro Orfanó

Método da variável fictícia para ajuste de modelos volumétricos estáveis e compatíveis em povoamentos florestais / por Evandro Orfanó Figueiredo. Rio Branco: Embrapa Acre, 2005.

28 p. il. (Embrapa Acre. Documentos, 95)

1. Floresta. 2. População de planta – Estatística. 3. Software.
4. Monitoramento ambiental. I. Título. II. Série.

CDD (19.ed.) 634.9285

Autor

Evandro Orfanó Figueiredo

Eng. agrôn., M.Sc., Embrapa Acre, Rodovia BR 364, km 14,
Caixa Postal 321, CEP 69908-970, Rio Branco-AC,
orfano@cpafac.embrapa.br

Apresentação

Muitas dúvidas têm surgido no momento de selecionar as equações cujas estimativas de volumes representem melhor o estoque madeireiro num povoamento florestal, visto que é comum a ocorrência de super ou subestimativas do estoque presente. A inconsistência dos dados volumétricos é ainda maior quando se trata de volumes de interesse distintos, a exemplo de volume total e comercial ou volume com casca e sem casca.

Com emprego de procedimentos já bastante conhecidos no campo do conhecimento de programação de software, a técnica da variável fictícia ou Dummy permite os mais diversos usos na biometria florestal. Uma destas possibilidades é o ajuste de modelos volumétricos para obter numa única expressão matemática todos os volumes de interesse. Esta estratégia de ajuste permite gerar modelos volumétricos consistentes e estáveis, uma vez que não haverá oscilações numéricas significativas entre os volumes obtidos com uma equação ajustada com a variável Dummy.

Esta publicação é destinada a técnicos florestais, estudantes e profissionais dos órgãos ambientais ligados ao monitoramento ambiental de plantios florestais. Seu conteúdo é a descrição do procedimento prático do emprego da variável fictícia na obtenção de equações de volume por meio de softwares estatísticos, embora os princípios matemáticos não sejam abordados neste documento.

Marcus Vinicio Neves d'Oliveira
Chefe-Geral da Embrapa Acre

Sumário

Introdução	9
Funcionamento das Variáveis Dummy em Regressões Matemáticas.....	11
Estrutura da Base de Dados.....	12
Construção dos Modelos para Estimar Volume de Árvores Individuais, Considerando as Variáveis Dummy	14
Análise Residual para Observações Influentes em Modelos Lineares com a Variável Dummy	16
Seleção da Equação de Regressão	18
Diagnóstico de Normalidade para Equações com a Variável Dummy	21
Efeito da Aplicação das Variáveis Dummy	22
Exemplos de Alguns Modelos com a Variável Dummy	24
Considerações Finais	27
Referências	27

Método da Variável Fictícia para Ajuste de Modelos Volumétricos Estáveis e Compatíveis em Povoamentos Florestais

Evandro Orfanó Figueiredo

Introdução

Nos diagnósticos florestais em florestas plantadas, um dos principais objetivos é a obtenção de estimativas de volume total e comercial (com e sem casca), utilizando-se as equações de volumes comerciais ou totais, ajustadas de forma independente.

Essas equações servem para estimar volumes individuais de árvores e, com base num procedimento de amostragem de campo, pode-se prever o volume de um povoamento florestal (Veiga, 1981).

No entanto, o ajuste de forma independente não garante que, para todos os indivíduos amostrados, haverá compatibilidade entre os volumes totais e comerciais com e sem casca, ou seja, que o volume comercial em determinado grupo de árvores inventariadas não será igual ou superior ao volume total, situação anormal mas bastante freqüente quando se ajusta o modelo matemático independentemente. A garantia de que os volumes de interesse (comercial, total, com casca, sem casca, etc.) serão compatíveis em toda a amplitude dos dados do diagnóstico florestal somente poderia ser dada se as equações de volume fossem ajustadas de maneira simultânea, em que numa única equação se pudessem obter todas as formas de volume de interesse e, principalmente, uma única

estatística de precisão, a qual permitirá a seleção do modelo matemático que represente todos os volumes.

Uma das possibilidades de maior simplicidade de aplicação é promover a compatibilidade das estimativas, obtendo uma equação geral que calcule todos os volumes de acordo com o interesse. Para isso, se faz necessário a introdução de uma variável fictícia associada a cada variável independente ou parâmetro, conforme for o caso.

Na linguagem de programação, a variável fictícia denominada de Dummy permite realizar distintas simulações de falso ou verdadeiro e, ainda, inúmeras possibilidades de composições binárias. Desta maneira, satisfaz a sintaxe de uma linguagem, mas é substituída quando o programa é executado.

Ignorar a possibilidade de ocorrer incompatibilidade entre os modelos ajustados para distintos volumes em determinadas situações é correr o risco de cometer sérias falhas de super ou subestimativas para determinados grupos de árvores. Se estas árvores compõem o principal grupo comercial de um empreendimento, isto se traduz em significativos prejuízos para o suprimento da indústria florestal.

Com a finalidade de esclarecer o emprego de variáveis fictícias em regressões lineares múltiplas e não-lineares, será demonstrado a seguir como estruturar uma base de dados para ajustes simultâneos e o procedimento de adição da variável Dummy em vários modelos volumétricos lineares e não-lineares a serem ajustados em diversos softwares estatísticos.

Este procedimento de obtenção de distintos volumes é usualmente empregado para povoamentos artificiais (como teca, pinus, eucalipto, acácias, etc.) e florestas nativas destinadas à energia, visto que em florestas nativas na Amazônia o foco principal é o volume do fuste.

Funcionamento das Variáveis Dummy em Regressões Matemáticas

A variável Dummy possibilita a realização de agrupamentos de regressões ajustadas para diferentes conjuntos de dados (Scolforo, 1997). Ao proceder ao agrupamento de regressões, a variável fictícia funciona como um delimitador dos valores preditos, permitindo uma compatibilidade entre as variáveis de interesse, tais como: altura total e comercial, volume com casca e sem casca, diferentes clones, entre outras possibilidades.

Sua aplicação pode ser feita de maneira a introduzir a variável em cada grupo de interesse, no caso: volume total com casca (V_1), volume total sem casca (V_2), volume comercial com casca (V_3) e volume comercial sem casca (V_4). Assim, a variável fictícia assume valor de 1 quando a observação pertence ao grupo de interesse (qualquer dos 4 volumes definidos) e valor de 0 (zero) quando a observação pertence a um grupo distinto. Na Tabela 1 consta o arranjo da variável Dummy para quatro grupos de interesse para ajuste dos modelos volumétricos.

Tabela 1. Arranjo matemático das variáveis Dummy para quatro grupos volumétricos de interesse.

Grupos de interesse	Var. Dummy	Grupos de interesse			
		V_1	V_2	V_3	V_4
Volume total com casca (V_1)	D_1	1	0	0	0
Volume total sem casca (V_2)	D_2	0	1	0	0
Volume comercial com casca (V_3)	D_3	0	0	1	0
Volume comercial sem casca (V_4)	D_4	0	0	0	1

A associação entre as variáveis Dummy com o modelo de regressão volumétrico a ser ajustado permite construir o modelo geral em que cada grupo volumétrico é representado pela variável de valor 1. Desta forma, a variável fictícia D_i é um fator específico dentro de um modelo matemático, o qual será multiplicado pela variável (dependente e independente) e/

ou parâmetro (conforme o caso). Assim, quando D_i é 1, a variável e/ou parâmetro será multiplicado por 1 e, conseqüentemente, parte do modelo é ativado; quando D_i for 0 (zero) os valores da variável e/ou parâmetro são multiplicados por 0 (zero) e parte do modelo será desativado.

Estrutura da Base de Dados

Quando se emprega a alternativa da variável fictícia para gerar a compatibilidade de equações volumétricas, a estruturação da base de dados para ajuste de modelos de regressão em softwares estatísticos consiste numa etapa que requer atenção, pois é neste arranjo de dados que será definido o “ligamento/desligamento” da variável Dummy. Na Tabela 2 consta uma das alternativas de estruturação da base de dados consolidados da cubagem rigorosa.

Tabela 2. Estrutura da base de dados da cubagem rigorosa para ajuste de modelo de regressão com a variável Dummy, considerando quatro grupos volumétricos de interesse. Povoamentos artificiais de *Tectona grandis* L.f., Rio Branco, Acre.

Árv.	Volumes	DAP	Ht	Vol. real	D ₁	D ₂	D ₃	D ₄
1	V ₁	15,5	14,6	0,130	1	0	0	0
1	V ₂	15,5	14,6	0,105	0	1	0	0
1	V ₃	15,5	14,6	0,129	0	0	1	0
1	V ₄	15,5	14,6	0,104	0	0	0	1
2	V ₁	16,4	14,8	0,113	1	0	0	0
2	V ₂	16,4	14,8	0,100	0	1	0	0
2	V ₃	16,4	14,8	0,112	0	0	1	0
2	V ₄	16,4	14,8	0,099	0	0	0	1
3	V ₁	16,6	16,6	0,137	1	0	0	0
3	V ₂	16,6	16,6	0,119	0	1	0	0
3	V ₃	16,6	16,6	0,136	0	0	1	0
3	V ₄	16,6	16,6	0,118	0	0	0	1
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
n _i	V ₁	16,7	11,5	0,115	1	0	0	0
n _i	V ₂	16,7	11,5	0,091	0	1	0	0
n _i	V ₃	16,7	11,5	0,115	0	0	1	0
n _i	V ₄	16,7	11,5	0,090	0	0	0	1

Onde: Árv. = árvore cubada rigorosamente; V₁ = volume total com casca; V₂ = volume total sem casca; V₃ = volume comercial com casca; V₄ = volume comercial sem casca; DAP = diâmetro à altura do peito; Ht = altura total da árvore cubada; Vol. real = volume obtido da cubagem rigorosa; D₁ = variável Dummy para V₁; D₂ = variável Dummy para V₂; D₃ = variável Dummy para V₃; e D₄ = variável Dummy para V₄.

Observa-se na estrutura da Tabela 2 que a variável D_i somente assume valor 1 quando o índice "i" de D_i e V_i é igual, definindo assim quando será utilizada a referida informação.

Construção dos Modelos para Estimar Volume de Árvores Individuais, Considerando as Variáveis Dummy

Para fins de exemplificação será construído o modelo geral com a variável fictícia empregando o modelo de simples entrada de Hohenald-Krenm, expresso por:

$$V = \hat{a}_0 + \hat{a}_1DAP + \hat{a}_2DAP^2 + \hat{a}_i$$

Onde:

V = volume

$\hat{a}_{is}$ = parâmetros estimados

DAP = diâmetro à altura do peito

$\hat{a}_i$ = erro de estimativa

Para construção do modelo geral serão adotados os mesmos volumes de interesse da Tabela 2: V_1 = volume total com casca; V_2 = volume total sem casca; V_3 = volume comercial com casca; e V_4 = volume comercial sem casca. Para isso, será vinculado a cada V_i um D_i específico, conforme os passos a seguir:

Passo 1 – Como são quatro equações matemáticas numa única expressão, também serão obtidos quatro resultados distintos para a variável dependente na equação geral. Assim, se faz necessário construir uma equação com quatro variáveis dependentes, conforme estrutura a seguir:

Variáveis dependentes de interesse	Variáveis Dummy			
Volume total com casca	V_1	X	D_1	Dummy 1
Volume total sem casca	V_2	X	D_2	Dummy 2
Volume comercial com casca	V_3	X	D_3	Dummy 3
Volume comercial sem casca	V_4	X	D_4	Dummy 4

Assim, o conjunto de variáveis dependentes da equação geral é expresso por:

$$(V_1.D_1 + V_2.D_2 + V_3.D_3 + V_4.D_4)$$

Passo 2 – Para construir o modelo geral foi considerada a seguinte estrutura:

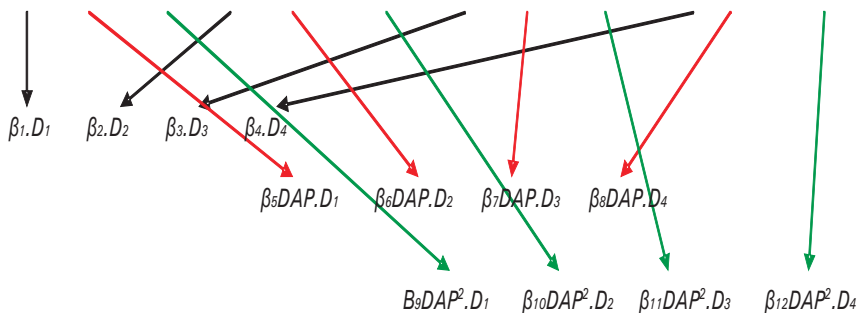
$$\begin{array}{llll} \mathbf{V_1} & \mathbf{V_2} & \mathbf{V_3} & \mathbf{V_4} \\ \beta_0 + \beta_1 DAP + & \beta_0 + \beta_1 DAP + & \beta_0 + \beta_1 DAP + & \beta_0 + \beta_1 DAP + \\ \beta_2 DAP^2 & \beta_2 DAP^2 & \beta_2 DAP^2 & \beta_2 DAP^2 \end{array}$$

Fragmentando os modelos e adicionando a variável Dummy, tem-se:

V ₁			V ₂			V ₃			V ₄		
β_0	$\beta_1 DAP$	$\beta_2 DAP^2$	β_0	$\beta_1 DAP$	$\beta_2 DAP^2$	β_0	$\beta_1 DAP$	$\beta_2 DAP^2$	β_0	$\beta_1 DAP$	$\beta_2 DAP^2$
+			+			+			+		
D ₁			D ₂			D ₃			D ₄		
$\beta_0.D_1$	$\beta_1 DAP.D_1$	$\beta_2 DAP^2.D_1$	$\beta_0.D_2$	$\beta_1 DAP.D_2$	$\beta_2 DAP^2.D_2$	$\beta_0.D_3$	$\beta_1 DAP.D_3$	$\beta_2 DAP^2.D_3$	$\beta_0.D_4$	$\beta_1 DAP.D_4$	$\beta_2 DAP^2.D_4$

Agora basta realizar o arranjo do modelo geral remunerando os parâmetros, em que:

V ₁			V ₂			V ₃			V ₄		
$\beta_0.D_1$	$\beta_1 DAP.D_1$	$\beta_2 DAP^2.D_1$	$\beta_0.D_2$	$\beta_1 DAP.D_2$	$\beta_2 DAP^2.D_2$	$\beta_0.D_3$	$\beta_1 DAP.D_3$	$\beta_2 DAP^2.D_3$	$\beta_0.D_4$	$\beta_1 DAP.D_4$	$\beta_2 DAP^2.D_4$



Passo 3 – Juntando as variáveis dependentes e as independentes tem-se o modelo Hohenald-Krenm geral com a variável Dummy a ser ajustado para obtenção dos quatro volumes de interesse, em que:

$$(V_1 D_1 + V_2 D_2 + V_3 D_3 + V_4 D_4) = \hat{a}_1 D_1 + \hat{a}_2 D_2 + \hat{a}_3 D_3 + \hat{a}_4 D_4 + \hat{a}_5 DAP D_1 + \hat{a}_6 DAP D_2 + \hat{a}_7 DAP D_3 + \hat{a}_8 DAP D_4 + \hat{a}_9 DAP^2 D_1 + \hat{a}_{10} DAP^2 D_2 + \hat{a}_{11} DAP^2 D_3 + \hat{a}_{12} DAP^2 D_4 + \hat{a}_{13}$$

Observa-se que o modelo geral será ajustado sem o intercepto ($\hat{\alpha}_0$), pois o intercepto de cada modelo encontra-se expresso nos parâmetros $\hat{\alpha}_1$, $\hat{\alpha}_2$, $\hat{\alpha}_3$ e $\hat{\alpha}_4$. Após o ajuste matemático em software estatístico ou por meio de planilhas empregando o método dos mínimos quadrados, a equação geral pode ser dividida em equações de volume correspondente a cada volume de interesse, já com os respectivos parâmetros estimados com a variável Dummy.

Análise Residual para Observações Influentes em Modelos Lineares com a Variável Dummy

Nas equações lineares com a variável Dummy sugere-se adotar os mesmos critérios empregados em regressões lineares para identificação de pontos influentes (*outlier*), conforme Figueiredo (2005). Entre os diversos medidores de atipicidade, destacam-se: elementos da diagonal principal da matriz H, DFFITS e Cook's D.

Os elementos da diagonal da matriz H, h_{ii} indicam a influência de y_i sobre o valor estimado $\hat{y}_i$. Além disso, como a soma dos valores da diagonal principal de H é igual a "p", o número de parâmetros do modelo, espera-se uma influência média da ordem de p/n , em que n é o número de observações. Dessa maneira, a análise é realizada da seguinte forma: quanto maior h_{ii} , maior será a distância do valor observado em relação à média de todas as variáveis explicativas. Para Souza (1998), observações que tenham h_{ii} maior que duas vezes a sua média, $2p/n$, devem ser alvo de exame, visto que podem indicar a existência de pontos de "alavancagem" (*leverage*).

A estatística DFFITS considera os efeitos de alavancagem para a avaliação dos pontos extremos (*outliers*) (Ferreira, 1996) que se afastam da mediana e afetam o valor da média de uma ou mais variáveis explicativas. São considerados influentes os pontos que afetam a linha de regressão de tal forma que sua retirada altere significativamente as estimativas. Assim, os

pontos extremos devem ser identificados para avaliar seu grau de influência sobre as estimativas, por meio de uma análise de sensibilidade que consiste em repetir o procedimento de regressão eliminando-se os pontos extremos para verificar qual a sua influência sobre os parâmetros estimados (SAS, 1990). Não havendo modificações significativas, aceita-se que os pontos extremos não são pontos influentes, ou seja, não afetam o modelo. A equação é dada por:

Onde:

h_{ii} = i-ésima diagonal da matriz de projeção no espaço preditor

e_i = i-ésimo resíduo

$S^2_{(i)}$ = i-ésima variância após a deleção da observação

analisada

DFFITS = $\left(\frac{1}{1 - h_{ii}} \right) \times \left(\frac{e_i}{\sqrt{S^2_{(i)}}} \right)$

Conforme sugerido por Chatterjee & Hadi (1986), adotou-se o critério a seguir para considerar os pontos influentes na estatística DFFITS.

$$DFFITS_i = \left(2 \sqrt{\frac{p+1}{n}} \right)$$

Onde:

p = número de parâmetros

n = número de observações

A estatística Cook's D mede alterações em todos os parâmetros da regressão, decorrentes da exclusão da i-ésima observação. Na análise em equações com a variável Dummy sugere-se que os pontos influentes são aqueles com valores maiores que $(2 \times h_{ii(\text{médio})})$.

A estatística Cook's D é definida por:

$$\text{Cook's } D_{(i)} = \left(\frac{1}{p} \right) \times \left(\frac{h_{ii}}{(1-h_{ii})} \right) \times \left(\frac{e_i^2}{S^2(1-h_{ii})} \right)$$

Onde:

h_{ii} = i-ésima da diagonal da matriz de H no espaço preditor

p = número de parâmetros

e_i = i-ésimo resíduo studentizado

$S^2(1-h_{ii})$ = a variância do resíduo da i-ésima observação

Para o ajuste dos modelos lineares com a variável Dummy sugere-se que somente seja considerado ponto influente quando a observação for significativa nos três critérios de identificação de atipicidade.

Seleção da Equação de Regressão

Na análise de regressão com e sem a variável fictícia é importante que o coeficiente de determinação (R^2) não seja a única estatística de análise, pois outras estatísticas devem ser consideradas tais como: gráficos de resíduos, análise de variância, erro padrão da estimativa, critério PRESSp (para modelos lineares). O R^2 não reflete a dispersão dos valores estimados, assim os melhores resultados do coeficiente de determinação podem esconder super ou subestimativas para um determinado grupo de dados que compõem a base de dados do ajuste.

A avaliação da análise da variância (soma de quadrados dos resíduos (SQR), quadrado médio dos resíduos (QMR) e teste F para o modelo completo) consiste no primeiro passo no processo de seleção, juntamente com a observação do coeficiente de determinação para modelos não centrados na média (R^2) e do coeficiente de determinação ajustado (R^2_{Ajustado}), em que:

$$R^2 = \left[1 - \left(\frac{SQ_{\text{Resíduo}}}{SQ_{\text{Total}}} \right) \right]$$

Onde:

R^2 = coeficiente de determinação

$SQ_{\text{Resíduo}}$ = soma de quadrados de resíduos

SQ_{Total} = soma de quadrados total

O R^2_{Ajustado} é preferível como estatística de análise, visto que o R^2 cresce com o simples incremento de variáveis no modelo, mesmo que o benefício provocado pelas variáveis seja pequeno. Isto será mais facilmente observado nas regressões múltiplas com a variável Dummy, sendo:

$$R^2_{\text{Ajustado}} = \left[1 - \left(\frac{(1 - R^2) \times (n - 1)}{n - p - 1} \right) \right]$$

Onde:

R^2_{Ajustado} = coeficiente de determinação ajustado

R^2 = coeficiente de determinação

n = número de dados utilizados no ajuste

p = número de parâmetros, com exceção do intercepto. Vale ressaltar que no caso do ajuste com a variável fictícia o intercepto não será calculado, pois o modelo não é centrado na média

O erro padrão da estimativa ou erro padrão residual (S_{yx}) mede a dispersão média entre os valores observados e estimados ao longo da linha da regressão. Menores valores no valor desta estatística indicam melhores ajustes. No caso de modelos com a variável dependente transformada, os valores do erro padrão residual devem ser corrigidos para a mesma unidade dos dados observados. O erro padrão pode ser obtido por:

$$S_{yx} = \sqrt{\frac{SQ_{\text{Res}}}{(n - p - 1)}}$$

Onde:

S_{yx} = erro padrão residual

SQ_{Res} = soma de quadrados de resíduos

n = número de dados utilizados no ajuste

p = número de parâmetros, com exceção do intercepto

O critério PRESSp (soma de quadrados de predição) é uma medida que avalia a qualidade dos valores ajustados por meio do modelo em questão, ou seja, avalia se a equação ajustada pode prever os valores observados Y_i . É uma importante ferramenta para a seleção de modelos lineares e adota o seguinte procedimento: cada valor ajustado $\hat{Y}$ para o critério PRESSp é obtido eliminando-se o i -ésimo caso do conjunto de dados, estimando a função de regressão para o modelo em questão com as $(n-1)$ observações restantes e, então, usando a equação de regressão ajustada, para obter o valor predito ($Y_{i(i)}$) para a i -ésima observação. O erro de predição PRESSp para o i -ésimo caso é dado por $Y_i - \hat{Y}_{i(i)}$. O critério PRESSp ($Y_i - \hat{Y}_{i(i)}$) é a soma de todos esses erros de predição quadráticos, sendo:

$$PRESS_p = \sum_{i=1}^n (Y_i - \hat{Y}_{i(i)})^2$$

Onde:

$PRESS_p$ = critério PRESSp (soma de quadrados da predição)

Y_i = valor da primeira estimativa

$\hat{Y}_{i(i)}$ = valor predito para a i -ésima observação

Essa estatística é analisada da seguinte forma: as melhores equações ajustadas são aquelas que apresentam valores pequenos de PRESSp (têm menor erro de predição).

A análise da distribuição gráfica dos valores residuais é um critério importante. Esta análise é decisiva na avaliação da qualidade das estimativas, pois permite detectar se há ou não tendenciosidade na estimativa da variável dependente ao longo de toda a linha de regressão.

Diagnóstico de Normalidade para Equações com a Variável Dummy

Uma das etapas finais do processo de diagnóstico em regressão envolve a avaliação da distribuição residual quanto à hipótese de normalidade. Para equações com a variável Dummy se faz necessário fragmentar a equação geral nas distintas equações com os volumes de interesse. No caso exemplificado serão testadas quatro equações retiradas da equação geral, sendo uma para cada variável dependente de interesse (V_1 , V_2 , V_3 e V_4).

O principal elemento para avaliar a distribuição residual é o resíduo padronizado. A padronização do resíduo é obtida por:

$$e_{\text{padronizado}} = \frac{Y_i - \hat{Y}}{\sqrt{\frac{SQ_{\text{Res}}}{(n - p - 1)}}}$$

Onde:

$e_{\text{padronizado}}$ = erro padronizado

SQ_{res} = soma do quadrado do resíduo

Y_i = valor observado

$\hat{Y}_i$ = valor estimado

n = número de dados utilizados no ajuste

p = número de parâmetros, com exceção do intercepto

A normalidade da distribuição residual caracteriza-se por 68% dos valores do erro padronizado encontrarem-se entre -1 e +1 e 95% entre os valores de -1,98 e +1,98. Pequenos desvios nesses percentuais não são considerados problemáticos para o diagnóstico, portanto, para avaliar a significância destes

desvios, os valores residuais devem ser submetidos a um procedimento formal de diagnóstico.

A técnica formal mais adequada para base de dados com $n > 50$ é o teste de Kolmogorov-Smirnov. Este teste de aderência avalia a concordância entre a distribuição observada dos resíduos e uma determinada distribuição teórica. Avalia-se se os dados residuais se aproximam razoavelmente de uma determinada distribuição, no caso normal. Para isso, utiliza-se a função distribuição acumulada observada e compara-se com a teórica, determina-se o ponto em que estas distribuições mais divergem e avalia-se a sua significância (Ayres, 2000).

Sendo $F_0(X)$ uma distribuição teórica acumulada e $S_n(X)$ uma distribuição observada em uma amostra de “n” observações, encontra-se a seguir o maior valor das diferenças (D) entre $F_0(X)$ e $S_n(X)$, ou seja:

Efeito da Aplicação das Variáveis Dummy

Os modelos gerais com as variáveis Dummy apresentam como vantagem a obtenção de estimativas compatíveis de volumes, no entanto, demandam um certo grau de atenção na sua aplicação, devido ao maior número de variáveis. No entanto, após o ajustamento e seleção das melhores equações gerais, estas podem ser fragmentadas de acordo com a variável dependente de interesse (volume total com casca, V_1 ; volume total sem casca, V_2 ; volume comercial com casca, V_3 ; e volume comercial sem casca, V_4), sem que ocorram prejuízos na compatibilidade dos modelos.

A principal característica do método é a obtenção de uma única estatística de precisão para os volumes de interesse, por meio da equação geral. Caso fosse realizado o ajuste independente, haveria o risco de selecionar um modelo

diferente para cada volume, podendo gerar estimativas instáveis e incompatíveis.

Esta compatibilidade se traduz na ausência de riscos de superposição de resultados, tais como volume total igual ou menor que o volume comercial, isto para qualquer situação dentro da amplitude dos valores observados. Contrapondo-se a esta situação desejável, pode ocorrer superposição de resultados ou uma exagerada aproximação em determinados grupos de observações, quando se ajusta a equação de forma independente e ocorre a seleção de modelos diferentes para cada volume. Estudos realizados por Figueiredo (2005) demonstram a incompatibilidade de equações ajustadas independentemente para volume total e comercial, exemplificadas graficamente com resultados preditos pelo modelo Scolforo (1997) e Schumacher & Hall (1933) (Fig. 1).

Na Fig. 1a observa-se o comportamento dos valores estimados para volume total com casca e volume total sem casca, no qual foi utilizada a metodologia de ajuste com a variável Dummy para o modelo Scolforo. É possível verificar que existe uma excelente combinação entre os valores de volumes, pois na medida que o volume total com casca cresce, o volume sem casca acompanha. Já na Fig. 1b observa-se uma completa desarmonia entre os volumes, visto que, neste caso, foi empregado um ajuste independente para volume total com casca, com a seleção do modelo Scolforo, e para volume total sem casca, foi escolhido o modelo Schumacher & Hall (1933) (não-linear). O critério para seleção do modelo Schumacher e Hall foi o mesmo empregado para seleção de modelos lineares, exceto a estatística PRESSp.

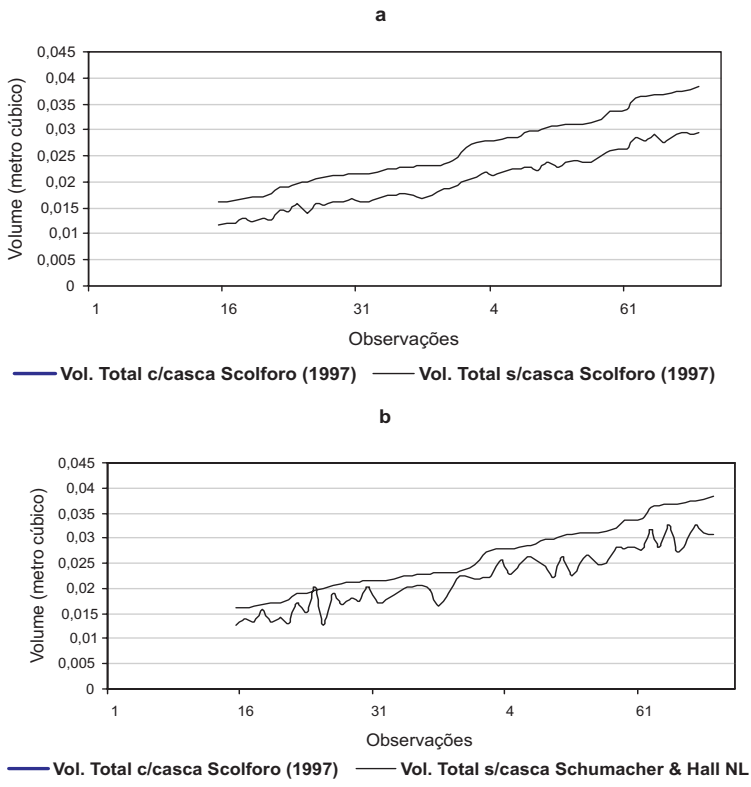


Fig. 1. Predição dos volumes totais e comerciais: a) resultados de predição do modelo Scolforo (1997), ajustado pela variável Dummy; b) resultados de predição do modelo Scolforo (1997) para volume com casca e do modelo Schumacher e Hall não-linear para volume sem casca, ajustados de forma independente, considerando o mesmo conjunto de dados de quatro povoamentos de *Tectona grandis* L.f.

Exemplos de Alguns Modelos com a Variável Dummy

Conforme adaptado por Figueiredo (2005), as Tabelas 3 e 4 apresentam alguns modelos com a variável Dummy já inserida, cujo objetivo é a obtenção de quatro formas de volume de interesse distintas e compatíveis.

Tabela 3. Modelos volumétricos de simples entrada para o volume de madeira comercial e total, com e sem casca, empregando-se as variáveis Dummy.

Modelos gerais com as variáveis Dummy	Autores ¹
$(V_1D_1 + V_2D_2 + V_3D_3 + V_4D_4) = \beta_1D_1 + \beta_2D_2 + \beta_3D_3 + \beta_4D_4 + \beta_5DAP^2D_1 + \beta_6DAP^2D_2 + \beta_7DAP^2D_3 + \beta_8DAP^2D_4 + \varepsilon_i$	Kopecky-Gehrhardt
$(V_1D_1 + V_2D_2 + V_3D_3 + V_4D_4) = \beta_1DAP D_1 + \beta_2DAP D_2 + \beta_3DAP D_3 + \beta_4DAP D_4 + \beta_5DAP^2D_1 + \beta_6DAP^2D_2 + \beta_7DAP^2D_3 + \beta_8DAP^2D_4 + \varepsilon_i$	Dissescu-Meyer
$(V_1D_1 + V_2D_2 + V_3D_3 + V_4D_4) = \beta_1D_1 + \beta_2D_2 + \beta_3D_3 + \beta_4D_4 + \beta_5DAP D_1 + \beta_6DAP D_2 + \beta_7DAP D_3 + \beta_8DAP D_4 + \beta_9DAP^2D_1 + \beta_{10}DAP^2D_2 + \beta_{11}DAP^2D_3 + \beta_{12}DAP^2D_4 + \varepsilon_i$	Hohenald-Krenm
$(V_1D_1 + V_2D_2 + V_3D_3 + V_4D_4) = [\beta_1(DAP)^{\beta_2}D_1 + [\beta_3DAP]^{\beta_4}D_2 + [B_5(DAP)^{\beta_6}D_3 + [\beta_7(DAP)^{\beta_8}D_4 + \varepsilon_i$	Berkhout
$[ln(V_1)]D_1 + [ln(V_2)]D_2 + [ln(V_3)]D_3 + [ln(V_4)]D_4 = \beta_1D_1 + \beta_2D_2 + \beta_3D_3 + \beta_4D_4 + \beta_5(ln DAP) D_1 + \beta_6(ln DAP) D_2 + \beta_7(ln DAP) D_3 + \beta_8(ln DAP) D_4 + \beta_9(DAP^{-1}) D_1 + \beta_{10}(DAP^{-1}) D_2 + \beta_{11}(DAP^{-1}) D_3 + \beta_{12}(DAP^{-1}) D_4 + \varepsilon_i$	Brenac
$[ln(V_1)]D_1 + [ln(V_2)]D_2 + [ln(V_3)]D_3 + [ln(V_4)]D_4 = \beta_1D_1 + \beta_2D_2 + \beta_3D_3 + \beta_4D_4 + \beta_5(ln DAP) D_1 + \beta_6(ln DAP) D_2 + \beta_7(ln DAP) D_3 + \beta_8(ln DAP) D_4 + \varepsilon_i$	Husch

¹Os modelos originais foram citados por Scolforo (1997).

Tabela 4. Modelos volumétricos de dupla entrada para o volume de madeira comercial e total, com e sem casca, empregando-se as variáveis Dummy.

Modelos gerais com as variáveis Dummy	Autores
$(V_1D_1 + V_2D_2 + V_3D_3 + V_4D_4) = [DAP^2/(\beta_1 + \beta_2H_i^{-1})]D_1 + [DAP^2/(\beta_3 + \beta_4H_i^{-1})]D_2 + [DAP^2/(\beta_7 + \beta_8H_i^{-1})]D_3 + [DAP^2/(\beta_9 + \beta_{10}H_i^{-1})]D_4 + \varepsilon_i$	Honner
$(V_1D_1 + V_2D_2 + V_3D_3 + V_4D_4) = \beta_1D_1 + \beta_2D_2 + \beta_3D_3 + \beta_4D_4 + \beta_5DAP D_1 + \beta_6DAP D_2 + \beta_7DAP D_3 + \beta_8DAP D_4 + \beta_9DAP^2D_1 + \beta_{10}DAP^2D_2 + \beta_{11}DAP^2D_3 + \beta_{12}DAP^2D_4 + \beta_{13}DAP H_i D_1 + \beta_{14}DAP H_i D_2 + \beta_{15}DAP H_i D_3 + \beta_{16}DAP H_i D_4 + \beta_{17}DAP^2H_i D_1 + \beta_{18}DAP^2H_i D_2 + \beta_{19}DAP^2H_i D_3 + \beta_{20}DAP^2H_i D_4 + \beta_{21}H_i D_1 + \beta_{22}H_i D_2 + \beta_{23}H_i D_3 + \beta_{24}H_i D_4 + \varepsilon_i$	Meyer ¹
$(V_1D_1 + V_2D_2 + V_3D_3 + V_4D_4) = \beta_1DAP^2 D_1 + \beta_2DAP^2 D_2 + \beta_3DAP^2 D_3 + \beta_4DAP^2 D_4 + \beta_5DAP^2H_i D_1 + \beta_6DAP^2H_i D_2 + \beta_7DAP^2H_i D_3 + \beta_8DAP^2H_i D_4 + \beta_9DAP H_i^2 D_1 + \beta_{10}DAP H_i^2 D_2 + \beta_{11}DAP H_i^2 D_3 + \beta_{12}DAP H_i^2 D_4 + \beta_{13}H_i^2 D_1 + \beta_{14}H_i^2 D_2 + \beta_{15}H_i^2 D_3 + \beta_{16}H_i^2 D_4 + \varepsilon_i$	Naslund ¹
$(V_1D_1 + V_2D_2 + V_3D_3 + V_4D_4) = [DAP^2(\beta_1 + \beta_2H_i)]D_1 + [DAP^2(\beta_3 + \beta_4H_i)]D_2 + [DAP^2(\beta_5 + \beta_6H_i)]D_3 + [DAP^2(\beta_7 + \beta_8H_i)]D_4 + \varepsilon_i$	Ogaya ¹
$(V_1D_1 + V_2D_2 + V_3D_3 + V_4D_4) = [\beta_1DAP^{\beta_2}H_i^{\beta_3}] D_1 + [\beta_4DAP^{\beta_5}H_i^{\beta_6}] D_2 + [\beta_7DAP^{\beta_8}H_i^{\beta_9}] D_3 + [\beta_{10}DAP^{\beta_{11}}H_i^{\beta_{12}}] D_4 + \varepsilon_i$	Schumacher e Hall ¹ não-linear
$[ln(V_1)]D_1 + [ln(V_2)]D_2 + [ln(V_3)]D_3 + [ln(V_4)]D_4 = \beta_1D_1 + \beta_2D_2 + \beta_3D_3 + \beta_4D_4 + \beta_5(ln DAP) D_1 + \beta_6(ln DAP) D_2 + \beta_7(ln DAP) D_3 + \beta_8(ln DAP) D_4 + \beta_9(ln H_i) D_1 + \beta_{10}(ln H_i) D_2 + \beta_{11}(ln H_i) D_3 + \beta_{12}(ln H_i) D_4 + \varepsilon_i$	Schumacher e Hall (1933)

¹Os modelos originais foram citados por Scolforo (1997).

Continua...

Tabela 4. Continuação.

Modelos gerais com as variáveis Dummy	Autores
$(V_1D_1 + V_2D_2 + V_3D_3 + V_4D_4) = \beta_1D_1 + \beta_2D_2 + \beta_3D_3 + \beta_4D_4 + \beta_5DAP/H_t$ $D_1 + \beta_6DAP/H_t D_2 + \beta_7DAP/H_t D_3 + \beta_8DAP/H_t D_4 + \beta_9DAP^2H_t D_1 +$ $\beta_{10}DAP^2H_t D_2 + \beta_{11}DAP^2H_t D_3 + \beta_{12}DAP^2H_t D_4 + \beta_{13}DAP^3H_t D_1 + \beta_{14}DAP^3H_t$ $D_2 + \beta_{15}DAP^3H_t D_3 + \beta_{16}DAP^3H_t D_4 + \epsilon_i$	Scolforo (1997)
$(V_1D_1 + V_2D_2 + V_3D_3 + V_4D_4) = \beta_1D_1 + \beta_2D_2 + \beta_3D_3 + \beta_4D_4 + \beta_5DAP H_t$ $D_1 + \beta_6DAP H_t D_2 + \beta_7DAP H_t D_3 + \beta_8DAP H_t D_4 + \beta_9DAP^3H_t D_1 +$ $\beta_{10}DAP^3H_t D_2 + \beta_{11}DAP^3H_t D_3 + \beta_{12}DAP^3H_t D_4 + \beta_{13}DAP H_t^2 D_1 + \beta_{14}DAP$ $H_t^2 D_2 + \beta_{15}DAP H_t^2 D_3 + \beta_{16}DAP H_t^2 D_4 + \epsilon_i$	Scolforo (Modelo 2) ¹
$(V_1D_1 + V_2D_2 + V_3D_3 + V_4D_4) = \beta_1D_1 + \beta_2D_2 + \beta_3D_3 + \beta_4D_4 + \beta_5H_t D_1 +$ $\beta_6H_t D_2 + \beta_7H_t D_3 + \beta_8H_t D_4 + \beta_9DAP^2H_t D_1 + \beta_{10}DAP^2H_t D_2 + \beta_{11}DAP^2H_t$ $D_3 + \beta_{12}DAP^2H_t D_4 + \beta_{13}DAP H_t^2 D_1 + \beta_{14}DAP H_t^2 D_2 + \beta_{15}DAP H_t^2 D_3 +$ $\beta_{16}DAP H_t^2 D_4 + \epsilon_i$	Scolforo (Modelo 3) ¹
$(V_1D_1 + V_2D_2 + V_3D_3 + V_4D_4) = \beta_1D_1 + \beta_2D_2 + \beta_3D_3 + \beta_4D_4 + \beta_5DAP^2H_t$ $D_1 + \beta_6DAP^2H_t D_2 + \beta_7DAP^2H_t D_3 + \beta_8DAP^2H_t D_4 + \epsilon_i$	Spurr (1952) (Modelo 1)
$[\ln(V_1)]D_1 + [\ln(V_2)]D_2 + [\ln(V_3)]D_3 + [\ln(V_4)]D_4 = \beta_1D_1 + \beta_2D_2 + \beta_3D_3 +$ $\beta_4D_4 + \beta_5\ln(DAP^2 H_t) D_1 + \beta_6\ln(DAP^2 H_t) D_2 + \beta_7\ln(DAP^2 H_t) D_3 +$ $\beta_8\ln(DAP^2 H_t) D_4 + \epsilon_i$	Spurr (logarítmica) ¹
$(V_1D_1 + V_2D_2 + V_3D_3 + V_4D_4) = \beta_1D_1 + \beta_2D_2 + \beta_3D_3 + \beta_4D_4 + \beta_5DAP^2D_1$ $+ \beta_6DAP^2D_2 + \beta_7DAP^2D_3 + \beta_8DAP^2D_4 + \beta_9DAP^2H_t D_1 + \beta_{10}DAP^2H_t D_2 +$ $\beta_{11}DAP^2H_t D_3 + \beta_{12}DAP^2H_t D_4 + \beta_{13}H_t D_1 + \beta_{14}H_t D_2 + \beta_{15}H_t D_3 + \beta_{16}H_t$ $D_4 + \epsilon_i$	Stoate (1945) (australiana)
$(V_1D_1 + V_2D_2 + V_3D_3 + V_4D_4) = [(DAP^2 H_t).(\beta_1 + \beta_2DAP)^{-1}]D_1 + [(DAP^2$ $H_t).(\beta_3 + \beta_4DAP)^{-1}]D_2 + [(DAP^2 H_t).(\beta_5 + \beta_6DAP)^{-1}]D_3 + [(DAP^2 H_t).(\beta_7 +$ $\beta_8DAP)^{-1}]D_4 + \epsilon_i$	Takata ¹

Onde:

- V_1 = volume total com casca
- V_2 = volume total sem casca
- V_3 = volume comercial com casca
- V_4 = volume comercial sem casca
- D_1 = variável Dummy para V_1
- D_2 = variável Dummy para V_2
- D_3 = variável Dummy para V_3
- D_4 = variável Dummy para V_4
- DAP = diâmetro à altura do peito
- H_t = altura total
- ϵ_i = erro da estimativa
- β_{is} = parâmetros a serem estimados
- ln = logaritmo neperiano

Considerações Finais

- O emprego do método da variável fictícia ou Dummy para a obtenção de diversas especificações de volume possibilita compatibilidade entre os volumes total e comercial (com e sem casca).
- A maior estabilidade das estimativas é observada na pequena oscilação gráfica dos volumes com e sem casca, quando se emprega o modelo completo com a variável fictícia.
- O procedimento de construção de modelos completos pode envolver duas ou mais variáveis de interesse.
- As estatísticas de análise de resíduos e normalidade seguem os mesmos procedimentos matemáticos que os modelos múltiplos.

Referências

- AYRES, M. **BioEstat 2.0**: aplicações estatísticas nas áreas das ciências biológicas e médicas. Belém: Sociedade Civil Mamirauá; Brasília: CNPq, 2000. 272 p.
- CHATTERJEE, S.; HADI, A. S. Influential observations, high leverage points, and outliers in linear regression. **Statistical Science**, v. 3, p. 379-416, 1986.
- FERREIRA, D. E. **Recursos computacionais na genética e melhoramento de plantas**. Lavras: UFLA, 1996. 80 p. Apostila.
- FIGUEIREDO, E. O. **Avaliação de povoamentos de teca (*Tectona grandis* L.f.) na microrregião do Baixo Rio Acre**. 2005. 301 p. Dissertação (Mestrado em Engenharia Florestal) - Universidade Federal de Lavras, Lavras.

SAS INSTITUTE. **SAS User's guide**: statistics. Release 6.03. Cary, 1990. 584 p.

SCHUMACHER, F. X.; HALL, F. dos S. Logarithmic expression of timber-tree volume. **Journal of Agricultural Research**, Washington, v. 47, n. 9, p. 719-734, 1933.

SCOLFORO, J. R. S. **Biometria florestal 2**: técnicas de regressão aplicada para estimar: volume, biomassa, relação hipsométrica e múltiplos produtos de madeira. Lavras: UFLA/FAEPE, 1997. 292 p. (Textos Acadêmicos).

SOUZA, G. S. **Introdução aos modelos de regressão linear e não-linear**. Brasília, DF: Embrapa-SPI; Embrapa-SEA, 1998. 505 p.

SPURR, S. H. **Forest inventory**. New York: Ronald, 1952. 476 p.

STOATE, I. N. The use of a volume equation in pine stands. **Australian Forestry**, Camberra, v. 9, p. 48-52, 1945.

VEIGA, R. A. de A. Uso de equações de volume em levantamentos florestais. In: SIMPÓSIO SOBRE INVENTÁRIO FLORESTAL, 1., 1981, Piracicaba. **Anais...** Piracicaba: IPEF, 1981. p. 93-102.

$$[\ln(V_1)]D_1 + [\ln(V_2)]D_2 + [\ln(V_3)]D_3 + [\ln(V_4)]D_4 = 8,827$$
$$1,80842.(\ln(DAP)).D_1 + 1,80842.(\ln(DAP)).D_2 + 1,8762$$
$$0,800847.(\ln(H_0)).D_2 + 0,674498.(\ln(H_0)).D_3 + 0,808362.(\ln$$



Acre

$$DFFITS_i = \left(\sqrt{\frac{h_{ii}}{1-h_{ii}}} \right) \times \left(\frac{e_i}{\sqrt{S^2_{(i)}(1-h_{ii})}} \right)$$

$$H = \begin{bmatrix} h_{ii} & h_{12} & h_{13} & h_{14} & \dots & h_{1n} \\ h_{21} & h_{ii} & h_{23} & h_{24} & \dots & h_{2n} \\ h_{31} & h_{32} & h_{ii} & h_{34} & \dots & h_{3n} \\ h_{41} & h_{42} & h_{43} & h_{ii} & \dots & h_{4n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ h_{n1} & h_{n2} & h_{n3} & h_{n4} & \dots & h_{ii} \end{bmatrix}$$

Cook'

$$R^2 = \left[1 - \left(\frac{SQ}{S} \right) \right]$$

$$S_{yx} = \sqrt{\frac{1}{n}}$$

$$e_{padronizado} = \frac{Y_i - \hat{Y}}{\sqrt{\frac{SQ_{Res}}{(n-(m)-1)}}}$$

$$\begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ X_{12} & X_{22} & X_{32} & \dots & X_{n2} \end{bmatrix} \times \begin{bmatrix} 1 & X_{12} \\ 1 & X_{22} \\ 1 & X_{32} \\ \dots & \dots \\ 1 & X_{n2} \end{bmatrix} \Rightarrow X'X =$$

Ministério da
Agricultura, Pecuária
e Abastecimento

