

**ALTERNATIVA DE MODELO LINEAR PARA ESTIMAÇÃO DA BIOMASSA VERDE DE
Bambusa vulgaris SCHRAD. EX J.C. WENDL NA EXISTÊNCIA DE MULTICOLINEARIDADE**

LINEAR MODEL ALTERNATIVE TO ESTIMATE THE GREEN BIOMASS OF THE *Bambusa vulgaris*
SCHRAD. EX J.C. WENDL. WITHIN THE APPEARANCE OF MULTICOLLINEARITY

Adriano Victor Lopes da Silva¹ Rinaldo Luiz Caraciolo Ferreira²
José Antonio Aleixo da Silva³ German Hugo Gutierrez Cespedes⁴

RESUMO

O objetivo deste trabalho foi utilizar um método estatístico multivariado como opção para estimar a biomassa verde da haste principal do bambu, *Bambusa vulgaris* Schrad. ex J. C. Wendl., na existência de multicolinearidade. Os dados foram provenientes de um experimento conduzido pela empresa Agroindustrial Excelsior S. A. (Agrimex) localizada no Engenho Itapirema, município de Goiana, PE. Foram utilizadas 450 hastes de bambu, que tiveram sua biomassa verde quantificada por meio do peso e mensuração de quatro variáveis independentes. Inicialmente, comprovou-se a existência da multicolinearidade por meio da matriz de correlação das variáveis independentes e pelo fator de inflação da variância e a opção utilizada foi a regressão linear com os componentes principais que tem como base a matriz de covariância. O resultado indicou que ao existir uma interpretação para os componentes principais o modelo apresenta um ajuste satisfatório aos dados, podendo ser utilizado para estimar a biomassa verde da haste principal do bambu.

Palavras-chaves: bambu; componentes principais; regressão por componentes principais.

ABSTRACT

The objective of this work was to use a multivariate statistical method as an alternative to estimate the green biomass of the main bamboo rod, *Bambusa vulgaris* Schrad. i.e.: J.C. Wendl., in the presence of multicollinearity. The data came from an experiment carried out for the Agroindustrial Excelsior S. A. (Agrimex) company located in the city of Goiana – PE. Quantified by their green biomass weight, 450 bamboo rods were used and 4 independent variables measured in the rod. Initially, the presence of the multicollinearity could be verified through the correlation matrix of the independent variables and the variance inflation factors, the alternative used was the regression of the principal components based on the covariate matrix. The result indicates that, when there is an interpretation to the main components, the model shows a satisfactory data adjust, and it could be used to estimate the green biomass of the main bamboo rod.

Keywords: bamboo; principal components; regression by principal components.

INTRODUÇÃO

O bambu é uma planta de grande utilidade industrial, sendo importante opção para a produção de biomassa, particularmente para o Brasil, país que intensamente utiliza essa espécie para produção de papel e energia (BRITTO *et al.*, 1997).

No Brasil, a espécie que apresenta maiores áreas de plantio é *Bambusa vulgaris* Schrad. ex J. C. Wendl.. A Região Nordeste tem a maior área plantada do mundo, com plantios distribuídos nos estados do Maranhão, da Paraíba e de Pernambuco. *Bambusa vulgaris* é, portanto, uma planta essencial ao desenvolvimento florestal do nordeste brasileiro, usada como matéria prima industrial para a produção e papel de fibra longa, que exibe maior resistência para uso em embalagens. No entanto, apesar do seu alto potencial produtivo, há poucas pesquisas em silvicultura e manejo (BONILLA, 1991).

Dentre os possíveis estudos a serem desenvolvidos, destaca-se a necessidade de métodos estatísticos para estimativa de produção de biomassa, o que implica nas decisões sobre espécies, variedades e clones

1. Estatístico, M.Sc., Professor Substituto, Universidade Federal Rural de Pernambuco, Rua Manoel de Medeiros, s/n, Dois Irmãos, CEP 52171-900, Recife (PE). v_adriano@hotmail.com
2. Engenheiro Florestal, Dr., Professor Associado, Universidade Federal Rural de Pernambuco, Rua Manoel de Medeiros, s/n, Dois Irmãos, CEP 52171-900, Recife (PE). rinaldo@dcfl.ufrpe.br
3. Engenheiro Agrônomo, PhD., Professor Associado, Universidade Federal Rural de Pernambuco, Rua Manoel de Medeiros, s/n, Dois Irmãos, CEP 52171-900, Recife (PE). jaaleixo@uol.com.br
4. Engenheiro Florestal, M.Sc., Pesquisador da Empresa Agroindustrial Excelsior S.A. (Agrimex), Engenho Itapicirica, CEP 55900-000, Goiana (PE). germanguitierrez@joaosantos.com.br

Recebido para publicação em 11/08/2008 e aceito em 31/03/2009.

cultivados, tipo de solo, espaçamento e idade de corte (SILVA e BONILLA, 1998). No entanto, em muitos casos, aparecem problemas a serem resolvidos para que os métodos utilizados sejam válidos. Dentre esses problemas, destaca-se a questão da multicolinearidade entre variáveis.

A grande motivação de aplicar regressão com componentes principais foi à utilização de todas as variáveis mensuradas, mesmo sendo variáveis altamente correlacionadas. Porém é muito utilizada, equivocadamente, a regressão linear diretamente nessas variáveis explicativas. Podendo surgir problemas quando o modelo tiver mais de uma variável ou quando as variáveis não tenham uma relação linear com a variável resposta, logo as estimativas não serão confiáveis.

O objetivo deste estudo visa a estabelecer uma metodologia que permita quantificar a biomassa verde da unidade amostral do bambu por meio de modelos de regressão com componentes principais na existência da multicolinearidade.

REVISÃO DE LITERATURA

Multicolinearidade

Segundo Montgomery *et al.* (2006), multicolinearidade é o nome dado ao problema geral que ocorre quando duas ou mais variáveis explicativas são muito correlacionadas entre si, o que torna difícil, utilizando-se apenas o modelo de regressão, distinguir suas influências separadamente, já que uma das suposições é que nenhuma relação linear exata pode existir entre quaisquer covariáveis ou combinações lineares destas.

Preocupações na área florestal quanto ao problema de multicolinearidade entre variáveis explicativas podem ser encontradas em vários trabalhos (CAIXETA *et al.*, 2003; VACCARO *et al.*, 2003; NUTTO e SPATHELF (2003); BATISTA *et al.*, 2004; ANTONANGELO e FENNER, 2005; OLIVEIRA *et al.*, 2005; ROLIM *et al.*, 2006; MARCHIORETTO *et al.*, 2007; TOMÉ *et al.*, 2007)

Segundo Montgomery *et al.* (2006), a consequência mais grave da presença de multicolinearidade é que o teste de significância do coeficiente de uma variável independente, sendo esta correlacionada com alguma outra, pode indicar sua exclusão do modelo, mesmo que exista uma forte relação linear desta com a variável resposta. Os autores afirmaram ainda que os efeitos de multicolinearidade podem ser facilmente demonstrados como segue:

Considerando que os elementos da diagonal da matriz de correlação $C = (X'X)^{-1}$ podem ser escritos como:

$$C_{jj} = \frac{1}{(1 - R_j^2)} \quad j = 1, 2, \dots, p$$

Sendo R_j^2 = coeficiente de determinação múltipla, resultante da regressão de x_j nos outros regressores. Claramente, quanto mais forte for a dependência linear de x_j nos regressores restantes, e por conseguinte mais forte a colinearidade, maior será o valor de R_j^2 . Considerando ainda que $V(\hat{\beta}_j) = \sigma^2 C_{jj}$, diz-se que a variância de $\hat{\beta}_j$ é “inflacionada” pela quantidade $(1 - R_j^2)^{-1}$. Dessa maneira, é definido o fator de inflação da variância para $\hat{\beta}_j$ como:

$$FIV(\hat{\beta}_j) = \frac{1}{(1 - R_j^2)} \quad j = 1, 2, \dots, p$$

Análise de Componente Principal (ACP)

Análise de Componente Principal (ACP) é uma técnica estatística que transforma um conjunto de “p” variáveis em um conjunto com um número menor “k” ($k < p$) de variáveis não correlacionadas, que explica uma parcela substancial das informações do conjunto original (KENDALL, 1950).

Além da redução da dimensionalidade dos dados, o objetivo principal é o de explicar a estrutura de variância e covariância de um vetor aleatório composto de “p” variáveis aleatórias, mediante a construção de combinações lineares das variáveis originais. Essas combinações lineares são chamadas de componentes principais e são não correlacionadas entre si (JOHNSON e WICHERN, 1998).

A suposição de normalidade do vetor aleatório não é requisito necessário para que a técnica de análise de componente principal possa ser utilizada. Entretanto, quando a distribuição de probabilidade do vetor aleatório for normal p-variado, as componentes principais, além de não correlacionadas, são independentes e têm distribuição normal (MINGOTTI, 2005).

Os componentes principais dependem somente da matriz de covariância ou da matriz de correlação do vetor aleatório de interesse. Geometricamente, os componentes principais são as coordenadas dos pontos amostrais em um sistema de eixos obtidos pela rotação do sistema de eixos originais, na direção da variabilidade máxima dos dados (JOHNSON e WICHERN, 1998).

Seja X o vetor de p variáveis originais $x' = (X_1, \dots, X_p)$, com vetor de médias $\mu = (\mu_1, \dots, \mu_p)'$, e matriz de covariância Σ_{pp} . Sejam $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ os autovalores da matriz Σ_{pp} , com respectivos autovetores normalizados e_1, e_2, \dots, e_p , em que $e_i = (e_{i1}, e_{i2}, \dots, e_{ip})'$, isto é, os autovetores e_i satisfazem as seguintes condições: $e_i' e_j = 0; \forall i \neq j; e_i' e_i = 1; \forall i = 1, 2, \dots, p; \sum_{k=1}^p e_i e_{ik} = \lambda_i e_i; \forall i = 1, 2, \dots, p$.

Então, a j-ésima componente principal da matriz Σ_{pp} , $j = 1, 2, \dots, p$ é definida como:

$$CP_j = e_j' X = e_{j1} X_1 + e_{j2} X_2 + \dots + e_{jk} X_p$$

Modelo de Regressão para as Componentes Principais

Se as componentes principais têm um significado intuitivo, é melhor expressar a equação da regressão em termos das componentes. Nos demais casos, é mais conveniente recolocar o modelo nas variáveis originais (MONTGOMERY *et al.*, 2006).

$$Y_i = \beta_0 + \beta_1 CP_{1i} + \beta_2 CP_{2i} + \dots + \beta_p CP_{pi} + \varepsilon_i$$

Em que todas as suposições, estimação e análise do modelo de regressão múltipla devem ser atendidas para o caso do modelo de regressão para as componentes principais, no entanto, as componentes principais, agora variáveis explicativas, são não correlacionados, o que elimina o problema da multicolinearidade (DRAPER e SMITH, 1981).

Vale ressaltar que o problema de multicolinearidade pode ser resolvido por outros métodos estatísticos, tais como, utilização de modelo não linear, *ridge regression* (HOERL e KENNARD, 1970a,b) e regressão com variáveis latentes – *latent root regression* (MALHOTRA, 2001).

MATERIAIS E MÉTODOS

Os dados foram provenientes de um experimento conduzido pela empresa Agroindustrial Excelsior S. A. (Agrimex) localizada no Engenho Itapirema na cidade de Goiana, PE, a cerca de 60 km da Cidade do Recife. A precipitação média anual é de 1900 mm. A topografia é ondulada a fortemente ondulada. Os tipos de solos variam do arenoso ao argiloso com profundidade variável. A região onde se encontra o Engenho é caracterizada por plantios de cana-de-açúcar nas áreas mais planas e bambu nas mais onduladas.

Para coleta de dados, na área de plantio de bambu, foram lançadas três parcelas ao acaso, cada uma medindo 15x15 m. Em cada parcela, foi realizada a contagem do número de touceira, e em seguida, em cada touceira, a contagem do número de hastes. Para cada haste da touceira foram mensuradas a altura (H), circunferências a 0,10 (CAB) e a 1,30m (CAP) do solo e a variável combinada (CAB²H). Posteriormente, procedeu-se ao corte e pesagem das hastes principais sem galhos secundários e folhas.

Para identificação de multicolinearidade nas estimativas dos coeficientes de regressão, considerou-se, como regra de decisão, fator de inflação da variância para os estimadores superior a cinco, com base na afirmativa de Montgomery *et al.* (2006) em que os fatores de inflação da variância não devem exceder 4 ou 5. Para estimativa de $FIV(\hat{\beta}_j)$ foi utilizada a seguinte expressão:

$$FIV(\hat{\beta}_j) = \frac{1}{(1 - R_j^2)} \quad j = 1, 2, \dots, p$$

Em que: $FIV(\hat{\beta}_j)$ = fator de inflação da variância; $\hat{\beta}_j$ = estimativa do j-ésimo coeficiente de regressão; R_j^2 = coeficiente de determinação múltipla j-ésimo coeficiente de regressão.

Depois de realizado o teste de multicolinearidade e verificada a existência desta, analisaram-se os diagramas de dispersão entre a variável resposta (Peso da haste do bambu, P) contra as variáveis explicativas (H, CAB, CAP, CAB²H).

A regressão com componentes principais foi obtida por meio do método de seleção de variável de Stepwise, em que os componentes foram as variáveis explicativas, pois a variabilidade dos componentes não é a maior importância, mas, sim, o seu grau de contribuição para o modelo.

Os parâmetros da equação de regressão foram estimados a partir das interrelações da variável dependente (biomassa) e das variáveis independentes (H, CAB, CAP e CAB²H). As estimativas foram obtidas por meio do método de mínimos quadrados que consistem no procedimento matemático para minimizar os erros quadráticos.

A obtenção das componentes principais foi baseada na matriz de covariância das variáveis originais conforme Johnson e Wichern (1998), excluindo-se a variável dependente (P).

As análises foram realizadas por meio do programa estatístico R, sendo utilizado o método mínimos quadrados para calcular as estimativas dos parâmetros da regressão.

RESULTADOS E DISCUSSÃO

A matriz de correlação entre as variáveis explicativas estão na Tabela 1. Segundo Silva e Silva (1982), pode-se considerar como correlação nula entre duas variáveis quando os valores dos índices de correlação estiverem entre - 0,40 e 0,40. Assim sendo, observa-se que as correlações entre todas as variáveis são não nulas, ou seja, existe de um alto grau de correlação entre as variáveis explicativas podendo resultar no problema de multicolinearidade. Tal situação justifica o uso da regressão por componentes principais, pois as componentes são não correlacionadas e, portanto, não apresentam o problema de multicolinearidade.

TABELA 1: Matriz de correlação entre as variáveis independentes: altura, circunferência na base da haste (CAB), circunferência a 1,30m do solo (CAP), combinada (CAB²H).

TABLE 1: Matrix of correlation between the independent variables height, circumference at the base of the stem (CAB), circumference at 1.30 m above the ground (CAP), combined (CAB²H).

Variável	H	CAB	CAP	CAB ² H
H	1,0000	0,7662	0,8564	0,8801
CAB	0,7662	1,0000	0,9247	0,8949
CAP	0,8564	0,9247	1,0000	0,9008
CAB ² H	0,8801	0,8949	0,9008	1,0000

Por meio do teste de multicolinearidade identificou-se que poderão existir problemas nas estimativas dos coeficientes de regressão, pois os fatores de inflação da variância para os estimadores foram todos acima de cinco, como observado na Tabela 2.

TABELA 2: Fatores de inflação da variância para os estimadores das variáveis independentes: altura, circunferência na base da haste (CAB), circunferência a 1,30m do solo (CAP), combinada (CAB²H).

TABLE 2: The variance inflation factors for the estimators of the independent variables height, circumference at the base of the stem (CAB), circumference at 1.30 m above the ground (CAP), combined (CAB²H).

Variável Retirada	FIV ($\hat{\beta}_j$)
H	7,8308
CAB	7,5988
CAP	7,1022
CAB ² H	5,2576

De acordo com Dias *et al.* (2003) esse procedimento deve ser realizado porque, quando ocorre multicolinearidade, os testes estatísticos podem falhar em detectar diferenças significativas entre os fatores.

Nota-se, na Figura 1, que só a variável explicativa CAB²H tem uma relação linear com a variável resposta, e a importância dessa variável combinada é a normalização dos resíduos no modelo de estimativa da biomassa verde, pois as outras variáveis não têm uma relação linear com o peso da haste, P.

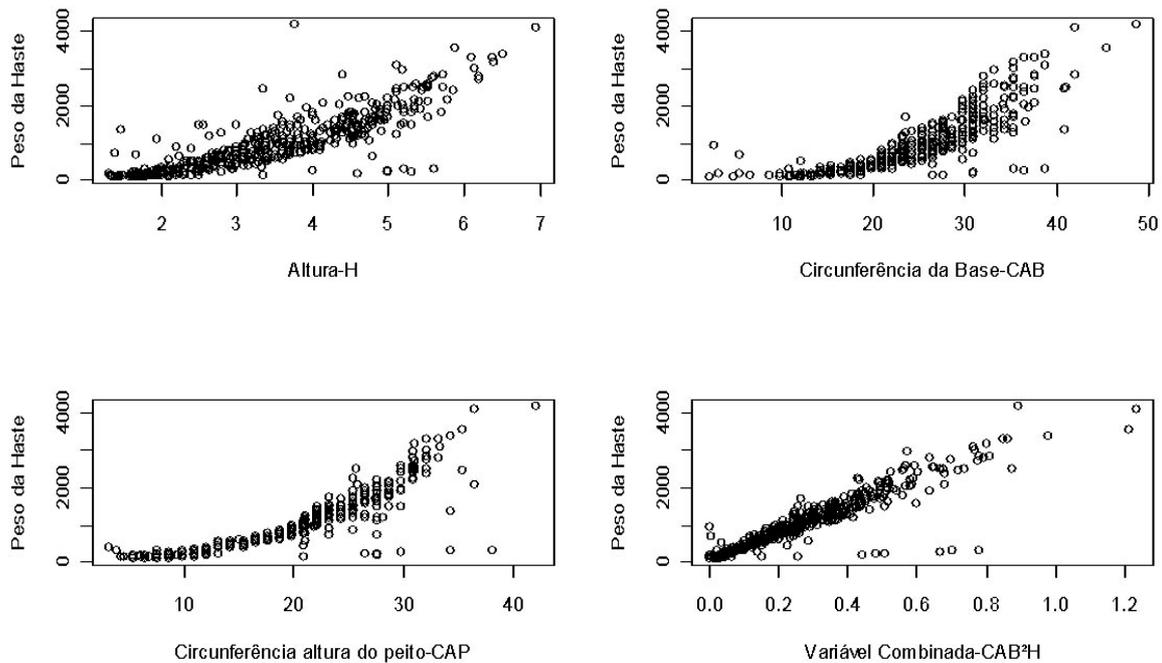


FIGURA 1: Diagramas de dispersão das variáveis: altura, circunferência na base da haste (CAB), circunferência a 1,30m do solo (CAP), combinada (CAB²H) versus variável resposta (P).

FIGURE 1: Diagrams of dispersion of the varying height, circumference at the base of the stem (CAB), circumference at 1.30 m above ground (CAP), combined (CAB²H) vs. variable response (P).

Na Tabela 3, está informada a importância dos componentes (contribuição dos autovalores) por meio da variabilidade partindo da matriz de covariância e a correlação entre as variáveis explicativas e os componentes.

Pela matriz de covariância, a primeira componente (CP1) que é, basicamente, um índice do peso da haste, explica quase 80% da variabilidade total e em que o maior coeficiente de grandeza se refere à variável CAP. Já a segunda componente (CP2) é uma comparação entre as variáveis H, CAB²H e CAB com a variável CAP. A terceira componente (CP3) é representada pelas variáveis H e CAB²H, sendo dominada pela variável H, que tem o maior coeficiente numérico em valor absoluto. A quarta componente (CP4) representa a variável CAB²H, pois mostra uma correlação maior da componente, em valor absoluto, com a variável e

correlação próxima de zero com as outras variáveis é a componente de menor variância amostral.

TABELA 3: Componentes principais pela matriz de covariância das variáveis independentes, altura, circunferência na base da haste (CAB), circunferência a 1,30m do solo (CAP), combinada (CAB²H).

TABLE 3: Main components by the covariance matrix of independent variables, height, circumference at the base of the stem (CAB), circumference at 1.30 m above the ground (CAP), combined (CAB²H).

Variáveis	CP1	CP2	CP3	CP4
H	0,8299	-0,2520	-0,4976	0,0044
CAB	0,9808	0,1950	-0,0027	0,0001
CAP	0,9812	-0,1926	0,0131	7,9708e ⁻⁶
CAB ² H	0,9165	-0,0236	-0,2308	-0,3258
Estatística	-	-	-	-
Desvio -padrão	10,9863	2,1833	0,6206	0,0687
Proporção da variância (%)	79,27%	15,75%	4,49%	0,49%
Proporção acumulada (%)	79,27%	95,02%	99,51%	100%

Na Tabela 4, são apresentadas às estimativas dos parâmetros, e nota-se que apenas o β_2 não é significativo para o modelo ao nível de 1% de probabilidade. No entanto, caso existisse uma correlação entre as variáveis explicativas (multicolinearidade), todos os parâmetros poderiam estar viesados, apesar de grande parte ser significativo ao nível de 1%, o que não é o caso, pois o modelo foi obtido com base nas componentes principais que são não correlacionadas.

TABELA 4: Estimativas dos parâmetros do modelo proposto pelo método de Stepwise para a seleção dos componentes.

TABLE 4: Estimates of the parameters of the model proposed by Stepwise method for the selection of components.

Parâmetros	Estimativas	Erros-padrão	p-valor
β_0	975,202	13,019	$< 2e^{-16}$ **
β_1	370,920	6,857	$< 2e^{-16}$ **
β_2	-65,711	26,453	0,0134 *
β_3	-270,211	42,217	$3,93e^{-10}$ **
β_4	-486,493	58,986	$1,84e^{-15}$ **

Em que: ** significativo a 1% ; * significativo a 10%.

As contribuições das variáveis são vistas na Tabela 5, e verifica-se que são significativas para o modelo. A estimativa da variância é dada por $S^2 = 76278$, enquanto que o coeficiente de determinação é dado por $R^2 = 0,8724$.

TABELA 5: Análise de variância para o modelo proposto pelo método de Stepwise para a seleção dos componentes.

TABLE 5: Analysis of variance for the model proposed by Stepwise method for the components selection.

Fontes de Variação	Graus de liberdade	Somas de Quadrados	Quadrados Médios	p-valor?
CP1	1	223195419	223195419	$< 2,2e^{-16}$ **
CP2	1	470659	470659	0,01336*
CP3	1	3124928	3124928	$3,932e^{-10}$ **
CP4	1	5188719	5188719	$1,839e^{-15}$ **
Resíduo	445	33943797	76278	
Total	449	265923522		

Em que: ** significativo a 1% ; * significativo a 10%.

Na Figura 2, observa-se que todas as suposições de regressão em relação ao erro são atendidas. As variáveis explicativas ou componentes, não são correlacionadas, ou seja, não existe problema de multicolinearidade. Aplicou-se o método de seleção de variáveis de Stepwise nos componentes porque não é a variabilidade do componente que vai obter o melhor modelo e, sim, a contribuição do componente. O

quarto componente (CP4) é o de menor variabilidade, porém é ele que contribui para a suposição do resíduo ser válida, pois, quando comparamos a Figura 2 com a Figura 3 que está sem o quarto componente (CP4), observa-se que uma das suposições residuais não é atendida. Logo, necessita-se de um termo quadrático no modelo. Como visto na Tabela 3, o CP4 representa a variável combinada, CAB²H, pois as correlações com as outras variáveis explicativas são próximas de zero.

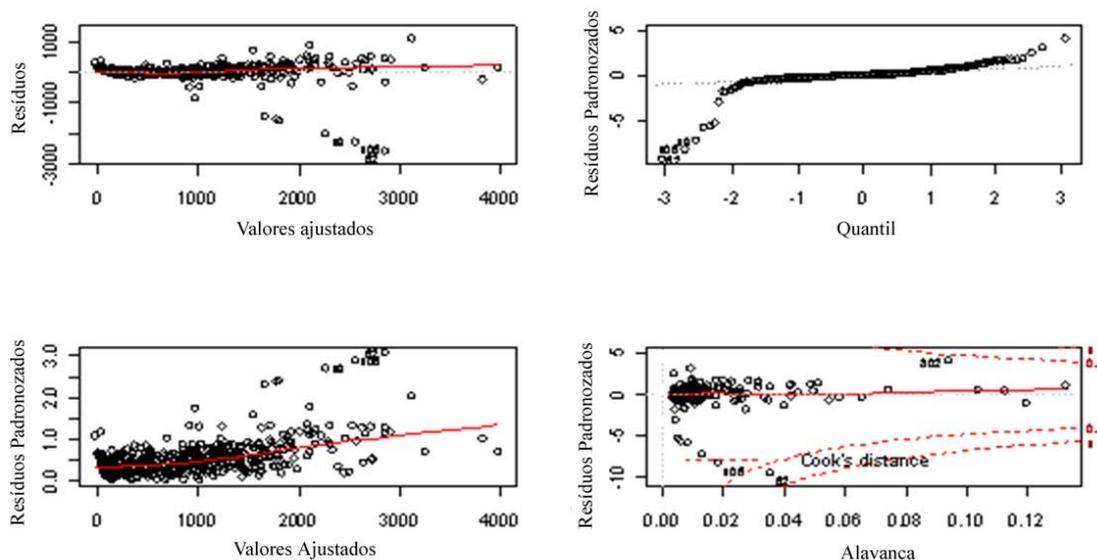


FIGURA 2: Gráficos dos resíduos para a equação obtida pelo método de Stepwise para a seleção dos componentes.

FIGURE 2: Graphs of residues to the equation obtained by the method of Stepwise for selection of components.

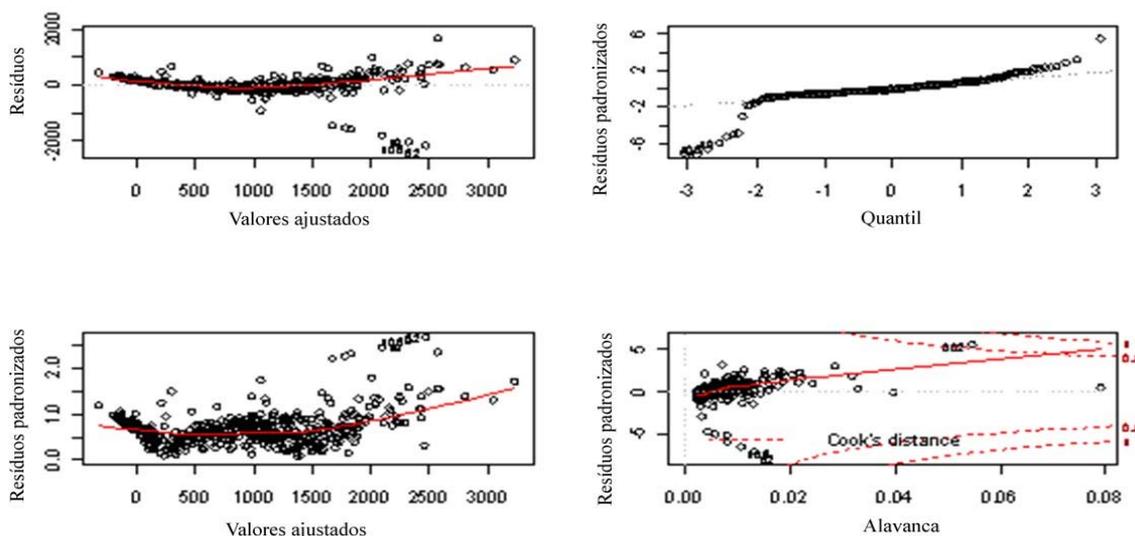


FIGURA 3: Gráficos dos resíduos para a equação obtida pelo método de Stepwise para a seleção dos componentes sem o quarto componente (CP4).

FIGURE 3: Graphs of waste to the equation obtained by the method of Stepwise for selection of components without the fourth component (CP4).

Vale salientar ainda que, na Figura 1, observou-se um comportamento típico de uma relação alométrica característica que pode ser representada por um modelo exponencial envolvendo a variável

diâmetro e altura, conforme o modelo de Schumacher e Hall (1933), e poderia ser uma opção ao método apresentado neste trabalho.

CONCLUSÕES

O problema de multicolinearidade entre as variáveis originais foi solucionado pela metodologia proposta.

O modelo de regressão por componentes principais obtido pode ser utilizado para estimar a biomassa verde da haste principal do bambu.

REFERÊNCIAS BIBLIOGRÁFICAS

- ANTONANGELO, A.; FENNER, P. T. Identificação de riscos de erosão em estradas de uso florestal através do critério do fator topográfico LS. **Energia Agrícola**, Botucatu, v.20, n.3, p.1-20, 2005.
- BATISTA, J.L.F.; MARCHESINI, M.; VIANA, V.M. Equações de volume para árvores de caixeta (*Tabebuia cassinoides*) no estado de São Paulo e sul do estado do Rio de Janeiro. **Scientia Forestalis**, Piracicaba, n. 65, p.162-175, jan./jun. 2004.
- BONILLA, O. H. **Análises quantitativas da produção de bambusa vulgares Scharder ex Wendland for vulgares no estado da Paraíba**. 1991. 89 f. Dissertação (Mestrado em Botânica) – Universidade Federal Rural de Pernambuco, Recife.
- BRITTO, J. O.; TOMAZELLO FILHO, M.; SALGADO, A. L. B. Produção e caracterização do carvão vegetal de espécies e variedades de bambu. **Instituto de Pesquisas Florestais – IPEF**, Piracicaba, n. 36, p. 13-17, 1997.
- CAIXETA, R. P. *et al.* Propriedades e classificação da madeira aplicadas à seleção de genótipos de *Eucalyptus*. **Revista Árvore**, Viçosa, v. 27, n. 1, p. 43-51, jan./fev. 2003.
- DIAS, L. T.; FARO, L.; ALBUQUERQUE, L. G. Estimativas de herdabilidade para perímetro escrotal de animais da raça nelore. **Revista Brasileira de Zootecnia**, Viçosa, v. 32, n. 6, p. 1878-1883, 2003. (Suplemento 2).
- DRAPER, N.; SMITH, H. **Applied regression analysis**. 2nd ed. New York: John Wiley & Sons, 1981. 709 p.
- HOERL, A. E.; KENNARD, R. W. Ridge regression: Applications to nonorthogonal problems. **Technometrics**, v.12, n.1, p.69-82. 1970a.
- HOERL, A.E., KENNARD, R.W. Ridge regression: Biased estimation for nonorthogonal problems. **Technometrics**, v.12, n.1, p.55-68. 1970b.
- JOHNSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis**. 4. ed. Englewood Cliffs: Prentice Hall, 1998. 816p.
- KENDALL. M. G. Factor analysis. **Journal of the Royal Statistical Society**, London, v. 12, p. 60-94, 1950.
- MALHOTRA, N.K. **Pesquisa de marketing: uma orientação aplicada**. 3ª Ed. Porto Alegre: Bookman, 2001. 719p.
- MARCHIOMETTO, M.S.; MAUHS, J.; BUDKE, J.C. Fenologia de espécies arbóreas zoocóricas em uma floresta psamófila no sul do Brasil. **Acta Botânica Brasílica**, v.21, n.1, p.193-201, 2007.
- MINGOTTI, S.A. **Análise de dados através de métodos de estatística multivariada – uma abordagem prática**. Belo Horizonte: Editora UFMG, 2005. 295p.
- MONTGOMERY, D. C.; PECK, E. A.; VINING, G. G. **Introduction to linear regression analysis**. 4. ed. New York: John Wiley & Sons, 2006. 612p.
- NUTTO, L.; SPATHELF, P. Modelagem da desrama natural de *Araucaria angustifolia* (Bert.) O. Ktze. **Floresta**, v. 33, n.3, p.295-309, 2003.
- OLIVEIRA, M.L.R.; SOARES, C.P.B.; SOUZA, A.L.; LEITE, H.G. Equações de volume de povoamento para fragmentos florestais naturais do município de Viçosa, Minas Gerais. **Revista Árvore**, v.29, n.2, Viçosa, 2005.
- PAULA, G. A. **Modelos de regressão com apoio computacional**. São Paulo: IME/USP, 2004. 253p.
- ROLIM, S.G.; COUTO, H.T.Z.; JESUS, R.M. *et al.* Modelos volumétricos para a Floresta Nacional do Tapirapé-Aquirí, Serra dos Carajás (PA). **Acta Amazônica**, v.36, n.1, p. 107-114, 2006.
- SCHUMACHER, F., HALL, F. Logarithmic expression of timber-tree. **Journal of Agricultural Research**, v.47, n.9, p.719-734, 1933.
- SILVA, J. A. A. ; BONILLA, O. H. . An exercise in the determination of optimal plot shape and size for estimating stem biomass in plantations of *Bambusa vulgaris* Scharader ex Wendland in the State of Paraíba, Brazil. **The International Tree Crops Journal**, Oxford, v. 9, p. 261-266, 1998.
- TOMÉ, M.; TOMÉ, J.; RIBEIRO, F. *et al.* Equação de Volume Total, Volume Percentual e de Perfil do Tronco para *Eucalyptus globulus* Labill. em Portugal. **Silva Lusitana**, v.15, n.1, p.25 - 39, 2007.
- VACCARO, S.; FINGER, C.A.G.; SCHNEIDER, P.R. *et al.* Incremento em área basal de árvores de uma floresta estacional decidual, em três fases sucessionais, no município de Santa Tereza, RS. **Ciência Florestal**, Santa Maria, v. 13, n. 2, p. 131-142, 2003.