Determinação e Modelagem do Volume Total de Pinus carabea var bahamensis: Uma Abordagem Multivariada

Mário Diego Rocha Valente
Núcleo de Estatística de Trânsito, DETRAN-PA,
66640-000, Belém, PA
mario.valente@detran.pa.gov.br

João Guimarães Pinheiro.
Depto. de Controle Estatístico da Qualidade, Vitória Régia Madeireira Exportadora.
66070-500, Belém, PA
joaogpinheiro@yahoo.com.br

Waldinei Travassos de Queiroz
Instituto de Ciências Agrárias, Programa de Pós-graduação em Engenharia Florestal, UFRA.
6607-7530, Belém, PA
waldenei.queiroz@ufra.edu.br

Resumo: *Neste trabalho propou-se um método para a construção de um modelo de regressão para determinar o Volume de Madeira Total da espécie florestal Pinus carabea var. bahamensis, em função de suas características (Diâmetro à Altura do Peito (DAP), Índice de Área Foliar, Área Basal, Idade e Altura). O modelo foi determinado utilizando-se a técnica multivariada de Análise de Fatores, através do Método de Máxima Verossimilhança via Rotação Ortogonal do tipo Varimax para extração dos fatores, procurando eliminar o problema da Multicolinearidade. Por fim desenvolveu-se um modelo de Regressão Linear Múltipla com base nos Escores Fatoriais. O modelo determinado apresentou-se de fácil interpretação e utilização, usando-se um fator e proporcionando um bom ajuste (R^2=0,91) aos dados e uma boa capacidade preditiva. Ele atendeu a todas as suposições teóricas para sua existência e utilização*.

Palavras-chave: Volume de Pinus, Multicolinearidade, Análise Multivariada.

Introdução

O conhecimento do volume total da árvore individual sobre casca é de enorme importância para o produtor florestal. Ele permite avaliar, com razoável rigor, o rendimento esperado quando um desbaste, do corte raso numa mata ou apenas de uma avaliação instantânea de existência. Uma vez que é uma variável de difícil medição direta, é necessário proceder-se à sua estimação. Para o efeito, procede-se geralmente à seleção de equações que estimem o volume da árvore em função de variáveis de mais fácil medição; é o caso do diâmetro à altura do peito e da altura total (**TOMÉ, 1990; ALEGRIA, 1993**).

Assim neste contexto que a Estatística Multivariada aparece como uma ferramenta fundamental e que na maioria das vezes não é utilizada. Assim, o objetivo deste trabalho foi explorar os dados dendrométricos disponíveis e provenientes de plantios de Quaruba, de modo a ilustrar a aplicabilidade e a importância da análise multivariada e tentar ajustar um modelo explicativo dos valores de Volume de *Pinus carabea var bahamensis*, a partir de um conjunto de variáveis biofísicas.

Material e Métodos

A área de estudo consiste em plantiu de *Pinus caribea* var. *bahamensis*, de uma Fazenda da Duratex Florestal S/A, cujas idades variam de 9,5 a 13,2 anos, caracterizando talhões maduros. A variável dependente foi o Volume da Madeira (Pinus). As variáveis independentes consideradas inicialmente e, em princípio, explicativas do volume, foram IAF (Índice de Área Foliar), DAF (Diâmetro Angular da Folha), DAP (Diâmetro à Altura do Peito), GAP (medida da Clareira), Idade, Altura e Área Basal da arvores.

Inicialmente, foi ajustado um modelo de regressão linear múltipla para o Volume de madeira de *Pinus*. Contudo, observou-se a presença de multicolinearidade nos dados, ou seja, verificou-se que, os coeficientes tanto de Correlação de Pearson (r = 0,91) quanto de Determinação (R^2=0,82) foram substancialmente altos e apenas a variável Altura foi estatisticamente

significativa com p-valor menor que o nível de significância de 0,05% e utilizando o método stepwise de regressão, também se confirmou que à medida que se introduzia uma variável por vez no modelo, a natureza das mesmas também mudava, por exemplo, quando se colocou a variável DAP no modelo esta mudou o sinal das outras variáveis, indicando assim uma forte multicolinearidade nos dados. Com isso, foi realizado primeiramente a Análise de Fatores com as variáveis originais, para obter fatores não correlacionados das variáveis, e assim contornar o problema da multicolinearidade.

Resultados e Discussão

Primeiramente observou-se a ordem de grandeza das variáveis, visando detectar discrepâncias que pudessem causar problemas na análise. Posteriormente, analisou-se a viabilidade da análise fatorial a partir da matriz de correlações. O primeiro passo é um exame visual das correlações, identificando as que são estatisticamente significantes, com isso, verificou-se que, existe um número substancial de correlações maiores que 0,30 (**GORSUCH, 1983**), supondo-se possíveis inter-relações entre as variáveis.

O próximo passo é avaliar a significância geral da matriz de correlação com o Teste de Esfericidade de Bartlett, neste caso, as correlações em geral são significantes ao nível de 0,000. Entretanto, isso testa apenas a presença de correlações não nulas, e não o padrão dessas correlações, assim, para testar se todas as características oriundas de diversos apartamentos possuem uma possível relação em comum, utilizou-se o Teste de Kaiser-Meyer-Olkin (KMO), ou seja, que vai testar as correlações parciais entre os pares de variáveis sem o efeito das demais, conforme Tabela 1, verificou-se que, neste caso ocupa um intervalo aceitável (acima de 0,50) conforme **KAISER e RICE (1974)**, com um valor de 0,75 classificados como Bom. Todas essas medidas indicam que o conjunto de variáveis é adequado à análise de fatores, e a análise pode prosseguir para os próximos estágios.

Tabela l: Medidas de Adequação da Análise Fatorial para os dados dendrométricos da Empresa Duratex Florestal em 2005.

Estatísticas	**Coeficientes**	**P-valor**
Teste de Bartlett	268,86	0,000
Teste de KMO	0,756	0,000

Assim, aplicou-se a análise de fatores aos dados dendrométricos, utilizando-se o método de máxima verossimilhança para extração dos fatores via Rotação Ortogonal do tipo Varimax para obter melhores combinações e usando como critérios para a escolha do número de fatores a extrair, o Critério da Raiz Latente e critério da percentagem da variância explicada. Assim, após a aplicação da análise de fatores, obtêm-se os seguintes resultados a seguir.

Tabela 2: Variância Total Explicada pelos Fatores dos dados dendrométricos da Duratex Florestal em 2005.

	Autovalores e Variâncias pos Rotação		
Variáveis	Variância	% Variância	% Variância Acumulada
1	3,79	47,38	47,38
2	2,33	29,23	76,61

A Tabela 2 mostra também, os autovalores, as percentagens das variâncias explicadas e acumuladas, calculadas após rotação ortogonal do tipo varimax nos dois primeiro fator extraído. Percebeu-se que um modelo com apenas dois fatores seriam suficientes para representar a estrutura de covariância inicial, com **23,39%** de perda de informações e que expliquem **76,61%** da variabilidade total dos dados originais. Sendo que o primeiro fator explica **47,38%** e o segundo fator explica **29,23%** Com isso, buscou-se identificar as variáveis que mais influenciam em cada fator, ou seja, as que possuem maiores cargas fatoriais.

Com os resultados obtidos na Tabela 3, pode-se observar que o primeiro fator possui pesos mais altos nas variáveis: DAP, Altura, Idade e Área Basal, sendo identificado como **Estrutura Volumar**. O segundo fator possui cargas fatoriais mais altas nas variáveis: IAF, GAP e DAF, sendo identificado como **Localização Volumar.**

Tabela 3: Matriz de Cargas Fatoriais dos dados dendrométricos da Empresa Duratex Florestal em 2005.

Características	Fator 1	Fator 2	h^2
Diâmetro a Altura do Peito	**0,93**	0,02	0,95
Idade	**0,81**	0,15	0,87
Altura	**0,95**	0,04	0,90
Área Basal	**0,72**	-0,47	0,72
Índice de Área Foliar (IAF)	0,1	**-0,93**	0,87
Distribuição Angular da Folha (DAF)	0,17	**0,71**	0,79
Clareira (GAP)	-0,04	**0,97**	0,79

De acordo com a Tabela 3, verificou-se que, a soma dos quadrados das cargas fatoriais para cada variável resulta num valor estimado da Comunalidade (h^2), que é à parte da variância total explicada pelo primeiro fator, assim, o tamanho da comunalidade é um índice útil para avaliar o quanto de variância em uma dada variável é explicada pela solução fatorial, com isso, pode-se observar nesta tabela, que todas as variáveis

são importantes na estrutura de covariância, pois as comunalidades são altas, acima de 0,70 conforme (**RENCHER, 2002).**

Em seguida, com a obtenção dos escores fatoriais, ou seja, medidas compostas de cada fator computado para cada árvore, substituindo-se as variáveis originais pelos escores fatoriais e realizou-se a **Regressão Linear Múltipla**. Considerou-se como variável resposta o Volume de Pinus e como variáveis explicativas os escores fatoriais referentes aos dois fatores extraídos.

Ao ajustar-se um modelo com os escores fornecidos pelo primeiro fator, verificou-se que os mesmos foram significativamente importante devido ao valor de p ser menor que o nível de significância de 0,05, como mostra a Tabela 4.

Tabela 4: Modelo de Regressão Linear para o Volume de Pinus carabea var bahamensis da Empresa Duratex Florestal em 2005.

Variáveis	Coeficientes	Teste t	P-valor
Constante	59,59	4,31	0,000
Fator 1	242,61	11,85	0,000
Fator 2	-11,39	-0,572	0,003

Após analisar a significância individual de cada fator de acordo com o teste t e por meio do valor descritivo (P-valor), partiu-se para analisar a significância geral do modelo de regressão, aplicando-se a Análise de variância (ANOVA) aos fatores.

A análise de variância mostra que se rejeita a hipótese de não haver regressão, isto é, o modelo é significativo a um nível de significância de 0, 05, e conclui-se que pelo menos uma variável explanatória esta relacionada com o valor de Volume de madeira de Pinus, conforme a Tabela 5.

Tabela 5: Análise de Variância para o Volume de Pinus carabea var bahamensis da Empresa Duratex Florestal em 2005.

F.V	G.L	S.Q	Q.M	F	P-valor
Reg.	2	115315,69	57657,850	70,49	0,000
Res.	33	26991,61	817,920		
Total	35				

Analisando-se as medidas de ajuste, observa-se que, os valores do coeficiente de correlação múltiplo (R múltiplo), o coeficiente de determinação (R^2) e o coeficiente de determinação ajustado (R^2 ajustado), são consideravelmente altos, indicando alta correlação da variável dependente com as variáveis independentes, e alta explicação da variável dependente pelas variáveis independentes, indicando que o modelo ajustado explica bem a variabilidade do valor do Volume de Madeira de Pinus.

Tabela 6: Análise de Variância para o Volume de Pinus carabea var bahamensis da Empresa Duratex Florestal em 2005.

Medidas	Valores
R múltiplo	0,90
R^2	0,93
R^2 ajustado	0,91

Em suma, o coeficiente de determinação múltiplo, que representa a proporção da variação em *Y(Valor)* que é explicada através do conjunto de variáveis explanatórias selecionadas, apresentou um valor igual a 91% da variação no volume pode ser explicado através da variação nas variáveis e 9% do volume são explicados por outras variáveis que não constam no modelo, conforme Tabela 6.

Portanto a equação de regressão linear múltipla para os dados do volume de pinus que descreve o relacionamento entre o valor de mercado e os dois fatores independentes é:

VOLUME $_{Pinus}$ = 59,5978 +242,6155***Escore do Fator 1** -10,3907***Escore do Fator 2**.

Sendo que o escore do fator 1 seria uma combinação linear entre as variáveis DAP, Idade, Altura e Área Basal e o escore do fator 2 por uma combinação das variáveis IAF, DAF e GAP obtidas por meio de um método de regressão linear múltipla, sendo desmembrado da seguinte forma.

VOLUME $_{Pinus}$ = 59,5978 + 242,61* [(0,93***DAP**) + (0,81***Idade**) + (0,95***Altura**) + (0,72***Área Basal**)] - 10,3907[(- 0,93***IAF**) + (0,71***DAF**) + (0,97***GAP**)]

Assim, a equação estimada para o volume de Pinus em função de suas características pode ser usada para predizer o volume de outras árvores sem o problema da multicolinearidade.

Análise de Resíduos

As suposições do modelo com a equação de regressão ajustada podem ser verificadas através da análise gráfica, utilizando-se os resíduos, a fim de procurar evidências sobre violações das suposições de Homocedasticidade, Normalidade e Independência dos resíduos, referentes aos dados dendrométricos de Pinus carabea var bahamensis

da Empresa Duratex Florestal em 2005.

Homocedasticidade dos Resíduos

A suposição de variância constante ou homogeneidade de variâncias é verificada facilmente através da Figura 1, que apresenta os pontos distribuídos aleatoriamente em torno de uma reta horizontal que passa pela origem sem qualquer padrão. Esta disposição dos pontos indica que a suposição de variância constante é razoável.

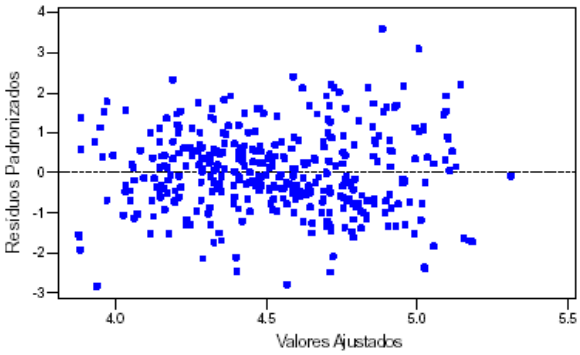

Figura 1: Resíduos Padronizados & Valores Ajustados para o Volume de Pinus carabea var bahamensis da Empresa Duratex Florestal em 2005.

Normalidade dos Resíduos

A Figura 2 construída pelos resíduos padronizados versus respectivos valores teóricos da distribuição normal, mostra a suposição de normalidade dos erros devidos seus pontos se distribuírem aleatoriamente em torno de uma linha reta.

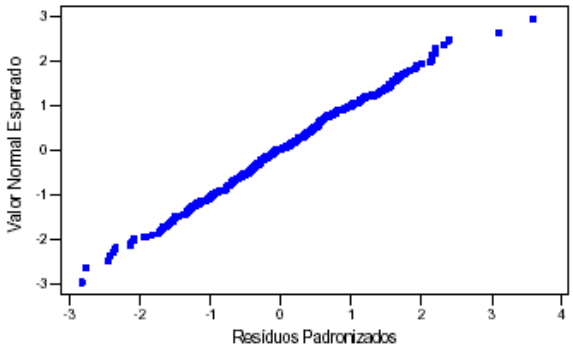

Figura 2: Valor Normal Esperado & Resíduos Padronizados para o Volume de Pinus carabea var bahamensis da Empresa Duratex Florestal em 2005.

Independência dos Resíduos

A existência de autocorrelação dos erros pode ser investigada através do gráfico dos resíduos versus sequência no tempo ou sequência de coleta dos dados, pode-se verificar a independências dos resíduos quando se distribuem aleatoriamente, em torno de zero, conforme a Figura 3.

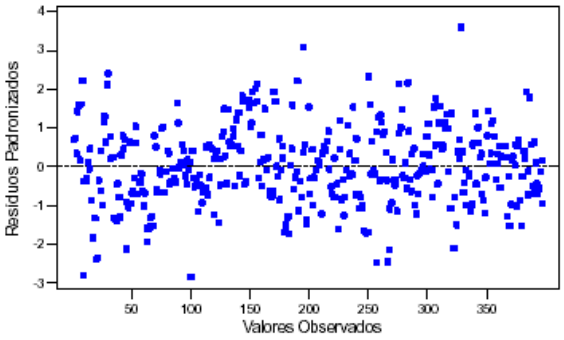

Figura 3: Resíduos Padronizados & Valores Observados para o Volume de Pinus carabea var bahamensis da Empresa Duratex Florestal em 2005.

Em suma, o modelo proposto para modelar o volume total de pinus carabea var bahamensis, ficou bem ajustado de acordo com a análise de resíduo.

Considerações Finais

Este trabalho teve como objetivo a aplicação conjunta das Técnicas de Análise Multivariada, para determinação do Volume de Madeira de Pinus da Empresa Duratex Florestal em 2005. A transformação e redução do número de variáveis podem ser obtidas pelo método da Análise de Fatores, sem perda significativa de informações, e obtendo-se variáveis não correlacionadas, por meio de dois fatores. Com os resultados obtidos da Análise Fatorial partiu-se para a construção do modelo de regressão linear múltipla para o volume de pinus, evitando-se o problema de multicolinearidade. No modelo ajustado obteve-se regressão com R^2 maior que 91%, o que garante a consistência da regressão para estimativa do volume de madeira de plantiu de pinus.

Referências

[l] ALEGRIA, C. M. Predição do volume total, volumes mercantis, perfil do tronco e sistemas de equações compatíveis para a Pinus pinaster Aiton. No distrito de Castelo Branco. Tese de Mestrado. ISA/UTL, 1993.

[2] GORSUCH, R. L. Factor Analysis. New Jersey: Lawrence Erlbaum, 1983.

[3] KAISER, H. F.; RICE, J. Little Jiffy, mark IV. Education and Psychological Measurement, 1974.

[4] NETER, J.; WASSERMAN, W. Applied linear statistical models. Illinois: Richard D. Irwin, 1974.

[5] RENCHER, A. Methods of Multivariate Analysis. 2 ed. New York: John Wiley & Son, 2002.

[6] TOME, J. A. L. Estimação do volume total, de volumes mercantis e modelação do perfil do tronco em Eucalyptus globulus Labill. Tese de Mestrado. ISA/UTL, 1990.