

MODELO DE PREDIÇÃO PARA O VOLUME TOTAL DE QUARUBA (*Vochysia inundata ducke*) VIA ANÁLISE DE FATORES E REGRESSÃO¹

Mário Diego Rocha Valente², Waldenei Travassos de Queiroz³, João Guimarães Pinheiro³ e Luis Augusto da Silva Monteiro⁴

RESUMO – Neste trabalho propôs-se um método para a construção de um modelo de regressão para determinar o Volume de Madeira Total da espécie florestal (*Vochysia inundata ducke*) Quaruba, em função de suas características (Diâmetro à Altura do Peito (DAP), Idade e Altura). O modelo foi determinado utilizando-se a técnica estatística multivariada de Análise de Fatores com do Método das Componentes Principais via Rotação Ortogonal do Tipo Varimax para Extração dos Fatores, procurando contornar o problema da Multicolinearidade. Por fim desenvolveu-se um modelo de Regressão Linear Simples com base nos Escores Fatoriais. O modelo determinado apresentou-se de fácil interpretação e utilização, usando-se um fator e proporcionando um bom ajuste ($R^2 = 96\%$) aos dados e uma boa capacidade preditiva. Ele atendeu a todas as suposições teóricas para sua existência e utilização.

Palavras-chave: Volume Total de Quaruba, Análise de Fatores e Regressão Linear.

PREDICTION MODEL FOR QUARUBA (*Vochysia inundata ducke*) TOTAL VOLUME VIA FACTOR ANALYSIS AND REGRESSION

ABSTRACT – This study proposed a method for building a regression model to determine the volume of total wood forest belonging to the *Vochysia inundata* Ducke (quaruba) species, according to their characteristics (diameter at breast height - DBH, Age and Height). The model was determined using the multivariate statistical technique of Factor Analysis with the Principal Components Method via Orthogonal Rotation from the Varimax Type for the Factors Extraction, seeking to circumvent the Multicollinearity problem. Finally, we developed a Simple Linear Regression model based on the Factorial Scores. The given model presented easy interpretation and usage, using a factor and providing a good fit ($R^2 = 96\%$) to the data and with good predictive ability. It met all the theoretical assumptions for its existence and use.

Keywords: Quaruba total volume, Factors analysis and Linear regression..

1. INTRODUÇÃO

A Flora Amazônica ainda é Praticamente desconhecida, com fantástica diversidade de plantas. O vastíssimo mar verde amazônico tem número incalculável de espécies. Sabe-se que a Amazônia possui gigantesco potencial madeireiro, despertando grande comércio, nessa área, na região, mas é necessário que tais recursos sejam mantidos de forma renovável, pois o uso indiscriminado dos vegetais pode causar, entre

outros males, a Desertificação da região. Entre as muitas espécies amazônicas, encontra-se a quaruba, árvore das matas de Terra Firme e várzea alta, não inundada, principalmente nos Estados do Pará e Amapá e, ainda, nas Guianas, porém pode ser encontrada desde o Sul do México até o Norte do Brasil (MELLO et al., 2009). Sua variabilidade é enorme, mas é comum encontrar árvores de 30 m de altura e 60 cm de diâmetro, entretanto já foram observadas árvores com 57 m de altura e 180 cm de diâmetro (REIS et al., 2009).

¹ Recebido em 12.09.2008 e aceito para publicação em 16.12.2010.

² Departamento de Trânsito do Estado do Pará, DETRAN-PA, Brasil. E-mail: <mario.valente@detran.pa.gov.br>.

³ Universidade Federal Rural da Amazônia, UFRA, Brasil. E-mail: <waldenei.queiroz@ufra.edu.br> e <joaogpinheiro@yahoo.com.br>.

⁴ Secretaria de Estado de Meio Ambiente do Pará, SEMA-PA, Brasil. E-mail: <augustofloresta@yahoo.com.br>.



Vochysia inundata Ducke (quaruba) é espécie florestal amazônica altamente promissora não somente pela qualidade de sua madeira, como também pelos excelentes resultados que vêm sendo obtidos na sua silvicultura (YARED et al., 1983). Com características semelhantes àquelas do cedro (*Cedrella odorata*), sua madeira tem sido utilizada principalmente em construção civil e naval, marcenaria, caixotaria, tábuas, ripas, forros, parte interior de móveis, brinquedos etc. (INSTITUTO BRASILEIRO DE DESENVOLVIMENTO FLORESTAL, 1981). Devido à sua condição econômica e utilitária, o monitoramento das características (variáveis) da quaruba é de extrema importância para o equilíbrio ambiental da região amazônica.

No setor florestal, o conhecimento sobre os recursos se dá por meio da medição ou estimação de atributos das árvores e das florestas, além de muitas características das áreas sobre as quais as árvores estão crescendo, por meio de instrumentos e métodos apropriados (HUSH et al., 1993).

Existem três fatores que governam a escolha das medidas a serem realizadas nas árvores: a facilidade e velocidade que as medidas podem ser realizadas; a exatidão com que podem ser feitas; e a correlação entre as medidas e as características às quais se desejam numa estimativa (SPURR, 1952).

De acordo com Paula Neto (1977), as medidas podem ser classificadas como diretas e indiretas. Quando são realizadas medidas diretas, faz-se, na realidade, uma determinação, enquanto as indiretas se referem à estimação.

Como medidas diretas, têm-se os diâmetros e as circunferências dos fustes e dos galhos das árvores, os comprimentos das toras, a espessura da casca e as alturas das árvores abatidas, entre outras. Como medidas indiretas, citam-se aquelas que estão fora do alcance direto do homem, sendo necessária, muitas vezes, a utilização de métodos óticos. Como exemplos dessas medidas, podem ser citados a altura e o volume das árvores em pé (SOARES et al., 2007).

O conhecimento do volume total da árvore individual sobre a casca é de enorme importância para o produtor florestal. Permite avaliar, com razoável rigor, o rendimento esperado quando do desbaste, do corte raso numa mata ou apenas de avaliação instantânea de existência. Uma vez que é variável de difícil medição direta, é

necessário proceder-se à sua estimação. Para o efeito de medição, 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).

Diferentes condições de clima e solo influenciam o crescimento e relação entre as variáveis da árvore. Dessa forma, a criação de modelos baseados em dados de um povoamento ou de uma região sujeita às condições uniformes de clima e de solo permite o fornecimento aos produtores florestais de ferramentas que possibilitam a estimação mais exata das variáveis nesse povoamento ou região.

Estudos envolvendo relações entre dados dendrométricos de dosséis florestais têm-se tornado cada vez mais frequentes. Nesses estudos, é comum a aplicação de testes estatísticos com o objetivo de avaliar o grau de associação entre esses parâmetros, por meio de abordagens que incluem a análise individual da ação de um parâmetro específico sobre outro e variável independente *versus* variável dependente (SOARES et al., 2007).

Normalmente, são ajustados modelos entre parâmetros biofísicos, de forma que seja possível inferir sobre as características de um dossel florestal a partir de dados dendrométricos. Entretanto, na utilização de vários parâmetros simultaneamente nem sempre os modelos são obtidos considerando-se todos os aspectos estatísticos necessários. Quando se dispõe de grande número de variáveis, é importante entender como elas estão se comportando e como estão se relacionando.

Nesse Contexto, as técnicas estatísticas surgem nesse cenário como importantes fontes de produção de conhecimento, principalmente para estimação do volume comercial, em que o uso de equações de volume e de relações hipsométricas em inventário florestal vem-se constituindo em operação rotineira para cálculo de volume de madeira em pé e estimativa da altura das árvores através da relação DAP e altura. Assim, o objetivo deste trabalho foi explorar os dados dendrométricos disponíveis e provenientes de plantios de quaruba na comunidade de Benjamin Constant, no Município de Bragança, de modo a ilustrar a aplicabilidade e a importância do Modelo de Regressão em conjunto com a Análise Fatorial e tentar ajustar um modelo explicativo dos valores de volume de quaruba dessa região, a partir de um conjunto de variáveis biofísicas.

2. MATERIAL E MÉTODO

2.1. Localização e Área de Estudo

A área de estudo localiza-se na comunidade Benjamin Constant, no Município de Bragança (01° 11' 22" S e 46° 40' 41" W), Nordeste paraense, em propriedade de pequeno agricultor.

A cobertura vegetal original do município (terra-firme) foi substituída por floresta secundária em vários estágios de regeneração, devido à ação dos desmatamentos. A topografia é suavemente ondulada, não havendo valores altimétricos expressivos, e as cotas mais elevadas giram em torno de 30 m, posicionadas na sua porção meridional, até atingir o nível do mar, na região litorânea. O clima do município de Bragança é equatorial superúmido, com temperatura máxima de 33 °C e mínima de 18 °C, apresentando média de 27 °C e elevada pluviosidade de 2.501 mm/ano, com período chuvoso nos primeiros seis meses do ano.

2.2. Levantamento dos Dados

A espécie de árvore utilizada no presente estudo originou-se da supressão da vegetação secundária de um pequeno agricultor para a fabricação de carvão vegetal, em que 32 indivíduos de *Vachysia inudata ducke* (quaruba) em 1 ha de floresta de aproximadamente 40 anos foram mensurados. A variável dependente foi o volume da madeira (quaruba). As variáveis independentes consideradas inicialmente e, em princípio, explicativas do volume foram Diâmetro à Altura do Peito (DAP), Altura e Idade.

Para o cálculo do volume real de madeira, empregou-se a fórmula de Smalian (MACHADO e FILHO, 2006), transformando os correspondentes diâmetros em áreas seccionais (A_1) e o respectivo comprimento, ou altura (h) do fuste ou galho considerado, de acordo com a fórmula:

$$V_i = \frac{A_1 + A_2}{2} \times h$$

2.3. Fundamentação Teórica

Em modelos de equação para estimar o volume de árvores em pé, utiliza-se constantemente técnica estatística chamada de Análise de Regressão Linear Múltipla, em que existe frequente interesse com relação

à natureza e significância das relações entre as variáveis independentes (Altura, DAP, Área Basal, Idade etc.) e a variável dependente (Volume).

Em muitas aplicações de administração, economia e principalmente nas ciências agrárias e florestais, frequentemente encontram-se variáveis independentes que estão correlacionadas entre elas mesmas e, também, com outras variáveis que não estão incluídas no modelo, mas estão relacionadas às variáveis dependentes (NETER; WASSERMAN, 1974). O fato de muitas funções de regressões diferentes proporcionarem bons ajustes para um mesmo conjunto de dados é porque os coeficientes de regressão atendem a várias amostras em que as variáveis independentes são altamente correlacionadas. Assim, os coeficientes de regressão estimados variam de uma amostra para outra quando as variáveis independentes estão altamente correlacionadas. Isso leva à informação imprecisa a respeito dos coeficientes verdadeiros, sendo esse fenômeno chamado de multicolinearidade (NETER e WASSERMAN, 1974).

2.4. Multicolinearidade

Multicolinearidade consiste em um problema comum em regressões, em que as variáveis independentes possuem relações lineares exatas ou aproximadamente exatas. O indício mais claro da existência da multicolinearidade é quando o R^2 é bastante alto, mas nenhum dos coeficientes da regressão é estatisticamente significativo segundo o teste t convencional. As consequências da multicolinearidade em regressão são a de erros-padrão elevados no caso de multicolinearidade moderada ou severa e, até mesmo, a impossibilidade de qualquer estimação se a multicolinearidade for perfeita (GUJARATI, 2000).

Quando as variáveis independentes são correlacionadas, o coeficiente de regressão de uma variável independente depende de qual outra variável independente é incluída no modelo, pois adicionando ou deletando uma das variáveis independentes mudam-se os coeficientes de regressão. Assim, um coeficiente de regressão deixa de refletir os efeitos inerentes de particular variável independente sobre a variável dependente, mas reflete apenas efeito parcial (GUJARATI, 2000).

Existem muitas sugestões, ou métodos propostos, para detectar a multicolinearidade. Os mais comumente usados são: Coeficiente de Correlação Simples (R),

Coefficiente de Explicação (R^2), Regressões Auxiliares, Raízes Características, Mudança de Sinal após a Introdução de uma Variável Independente, Gráficos dos Resíduos etc. (HAIR et al., 2005).

A existência de multicolinearidade tendo sido detectada e considerada prejudicial indica que o pesquisador deve procurar soluções para suavizar seus efeitos ruins. Várias medidas corretivas têm sido propostas, desde simples às mais complexas, para suavizar os efeitos provocados pela multicolinearidade. As mais usadas e citadas na literatura são: Remoção de Variáveis, Ampliação do Tamanho da Amostra, Ridge Regression, Método das Componentes Principais, Análise Fatorial, Redes Neurais etc. (ELIAN, 1988; HAIR et al., 2005).

Valente et al. (2008) estimaram um modelo de regressão linear múltipla para o volume de *Pinus carabea* var *Bahamensis*, com base nos dados da Empresa Duratex Florestal S/A na cidade de São Paulo, em 2007, quando verificaram a presença de multicolinearidade nas características das árvores, considerando problema para estimação de equação de volume e aplicando a técnica Análise Fatorial para contorná-lo e, em seguida, o método de mínimos quadrados para estimação do volume da árvore.

Neste trabalho, pretendeu-se adotar um método de maior rigor estatístico e que permita estimar o volume total da espécie *Vochysia inundata* Ducke (quaruba), contornando o problema da multicolinearidade que em dados da área florestal ocorre constantemente, empregando a técnica da Análise Fatorial e seguindo as noções de Valente et al. (2008).

2.5. Análise Fatorial

A análise de fatores ou análise fatorial (AF), de modo geral, é uma técnica estatística multivariada que tem como princípio analisar a estrutura das relações ou inter-relações (correlações) de grande número de variáveis, ou seja, descrever a estrutura de dependência de um conjunto de variáveis através da criação de fatores que são variáveis que, supostamente, medem aspectos comuns (HAIR et al., 2005). Com o emprego dessa técnica, inicialmente podem-se identificar as dimensões isoladas da estrutura dos dados e, então, determinar o grau em que cada variável é explicada por cada dimensão ou fator.

Segundo Míngoti (2005), a AF tem como objetivo principal descrever a variabilidade original do vetor aleatório X , em termos de um número menor M de variáveis

aleatórias, chamadas de fatores comuns e que estão relacionadas com o vetor original X através de modelo linear. Nesse modelo, parte da variabilidade de X é atribuída aos fatores comuns, sendo o restante da variabilidade de X creditado às variáveis que não foram incluídas no modelo, ou seja, o erro aleatório.

Para Barroso e Artes (2003), a AF é técnica que descreve a estrutura de dependência de um conjunto de variáveis, através da criação de fatores que são variáveis que, supostamente, medem aspectos comuns. Reis (2001) descreveu a AF como conjunto de técnicas cuja finalidade é representar ou descrever um número de variáveis iniciais a partir do menor número de variáveis hipotéticas. Trata-se de técnica estatística multivariada que, a partir da estrutura de dependência entre as variáveis de interesse (em geral representada pelas correlações ou covariâncias entre essas variáveis), permite a criação de um conjunto menor de variáveis (variáveis latentes ou fatores), obtidas a partir das originais.

Modelo Estatístico da Análise de Fatores

O modelo estatístico usado na análise fatorial explica a estrutura de correlação entre as características e é diretamente observado por meio de combinação linear de variáveis, as quais não são diretamente verificadas, denominadas fatores comuns, acrescidas de componente residual. Um modelo de análise fatorial pode ser apresentado na forma matricial, conforme citado por Dillon e Goldstein (1984).

$$X = \alpha F + \varepsilon$$

em que:

X = o p -dimensional vetor transposto das variáveis observadas (características), denotado por

$$Y' = (y_1, y_2, \dots, y_k) ;$$

α = uma matriz (p, k) tal que cada elemento expressa α_{ij} a correlação entre o volume y_{ij} e o fator f_j , sendo Λ denominado matriz de cargas fatoriais com o número k de fatores menor que o número p de variáveis;

F = o q -dimensional vetor transposto de variáveis não observáveis ou variáveis latentes chamadas de fatores comuns, denotado por $F' = (f_1, f_2, \dots, f_k)$, sendo $k < p$; e

ε = o p -dimensional vetor transposto de variáveis aleatórias ou fatores únicos, ou seja, vetor de componentes residuais, denotado por .

$$\varepsilon' = (e_1, e_2, \dots, e_k)$$

No modelo de análise fatorial, pressupõe-se que os fatores específicos são ortogonais entre si e com todos os fatores comuns. Normalmente, $E(E) = E(F) = 0$ e $Cov(E, F) = 0$.

2.5.1. Métodos para Extração e Escolha do Número de Fatores

Há na literatura vários métodos para a extração ou obtenção de fatores. Aborda-se o mais popularmente usado: análise fatorial via Método das Componentes Principais. Uma análise das componentes principais se ocupa em explicar a estrutura da variância e da covariância através de poucas combinações lineares das variáveis originais. Seu objetivo geral consiste tanto em reduzir os dados quanto interpretá-los adequadamente. Uma análise das componentes principais frequentemente revela relações que não eram previamente consideradas e, assim, permitem interpretações que não iriam, de outro modo, aparecer (BARROSO e ARTES, 2003).

A escolha dos fatores foi realizada por meio do Critério de Raiz Latente. Essa técnica parte do princípio de que qualquer fator individual deve explicar a variância de pelo menos uma variável para que seja mantida para interpretação. Cada variável contribui com um valor 1 do autovalor total. Com efeito, apenas os fatores que têm raízes latentes ou autovalores maiores que 1 são considerados significantes, e os demais fatores com autovalores menores do que 1 são insignificantes e descartados (HAIR et al., 2005; MINGOTI, 2005; SANTANA, 2006). A matriz de cargas fatoriais, que mede a correlação entre os fatores comuns e as variáveis observáveis, é determinada por meio da matriz de correlação, conforme Dillon e Goldstein (1984).

2.5.2. Rotação dos Fatores e Escores Fatoriais

Existe variedade de métodos de rotação dos fatores, baseados em concepções científicas pertinentes, que geram fatores correlacionados (Rotação Oblíqua) e não correlacionados (Rotação Ortogonal). Neste tópico, aborda-se o método de rotação dos fatores mais utilizados em análise fatorial, o método de Rotação Ortogonal, principalmente quando a solução inicial de uma análise fatorial não for facilmente interpretável; nesse caso, podem-se procurar rotações que levem a melhores soluções (JOHNSON e WICHERN, 1998).

A estrutura inicial utilizada para determinar a matriz de Cargas Fatoriais, em geral, pode não fornecer padrão significativo de cargas das variáveis, por isso não é

definitiva. A confirmação ou não dessa estrutura inicial pode ser feita por meio de vários métodos de rotação dos fatores (DILLON e GOLDSTEIN, 1984; JOHNSON e WICHERN, 1992). Nesta pesquisa, utilizou-se o método Varimax de rotação ortogonal dos fatores, que é um processo em que os eixos e referência dos fatores são rotacionados em torno da origem até que alguma outra posição seja alcançada. O objetivo é redistribuir a variância dos primeiros fatores para os demais e atingir um padrão fatorial mais simples e teoricamente mais significativo (REIS, 2001; HAIR et al., 2005; SANTANA, 2005).

Para a construção do modelo de equação de volume, estimou-se a matriz de escores fatoriais após a rotação ortogonal da estrutura fatorial inicial. O escore fatorial, por definição, situa cada observação no espaço dos fatores comuns. Para cada fator f_i , o i -ésimo escore fatorial extraído é definido por F_{ij} , expresso da seguinte forma (DILLON e GOLDSTEIN, 1984):

$$F_{ij} = b_1x_{i1} + b_2x_{i2} + \dots + b_px_{ip}$$

em que:

b_i são os coeficientes de regressão estimados para os n escores fatoriais comuns; e

x_{ij} são as n observações das p variáveis observadas.

Assim, os escores fatoriais podem ser utilizados posteriormente para construção de gráficos, mapas de aptidão agrícola, indicadores setoriais (MALHOTRA, 2001), como variáveis-resposta ou explicativas para algum outro procedimento estatístico. Neste trabalho, utilizaram-se os escores fatoriais para estimar um modelo de regressão linear para o volume de madeira de quaruba com base na combinação linear das variáveis originais.

2.6. Análise da Confiabilidade e Viabilidade da Análise de Fatores

Uma das premissas de uma análise fatorial é que exista estrutura de dependência clara entre as variáveis da matriz de covariância ou de correlação. A existência de tal estrutura significa que uma variável pode, dentro de certos limites, ser prevista pelas demais. Para verificar esse fato, podem-se calcular os coeficientes de correlação parcial entre os pares de variáveis, eliminando o efeito das demais variáveis. Espera-se que os valores obtidos sejam baixos. Diante desse problema, foram feitos testes para aferir a adequação do método à amostra de dados. Os dois principais testes aplicados foram: o teste de Esfericidade de Bartlett e o teste de Kaiser-Meyer-Olkin (KMO).

2.6.1. Teste de Esfericidade de Bartlett

Este teste avalia a significância geral da matriz de correlação, ou seja, testa se todas as variáveis oriundas de diversos setores possuem possível relação em comum (DILLON e GOLDSTEIN, 1984; REIS, 2001).

O teste de Bartlett testa as seguintes hipóteses nulas:

$$H_0: R=I \quad \text{ou} \quad H_0: \lambda_1 = \lambda_2 = \dots = \lambda_p$$

$$\chi^2 = -[n-1 - \frac{1}{6}(2p+5)] \ln |R|$$

ou

$$\chi^2 = -[n-1 - \frac{1}{6}(2p+5)] \sum_{i=1}^p \ln \lambda_i$$

em que $|R|$ é o determinante da matriz de correlação amostral, λ_i é a variância explicada por cada fator, n é o número de observações e p é o número de variáveis. A estatística tem distribuição assintótica de com $(0,5p(p-1))$ graus de liberdade.

2.6.2. Teste de Kaiser-Meyer-Olkin (KMO)

Este teste é usado para comparar as correlações parciais entre os pares de variáveis sem o efeito das demais, ou seja, testa variáveis duas a duas que possuem algum tipo de relação entre si (DILLON e GOLDSTEIN, 1984; KAISER, 1958; REIS, 2001).

$$KMO = \frac{\sum_i \sum_j r_{ij}^2}{\sum_i \sum_j r_{ij}^2 + \sum_i \sum_j a_{ij}^2}$$

em que r_{ij} é o coeficiente de correlação observado entre as variáveis i e j , e a_{ij} é o coeficiente de correlação parcial entre as mesmas variáveis que são, simultaneamente, estimativa das correlações entre os fatores, eliminado o efeito das demais variáveis. Os r_{ij} deverão assumir valores próximos de zero, uma vez que se pressupõe que os fatores são ortogonais entre si.

Conforme Kaiser e Rice, (1974), os valores do teste são classificados da seguinte forma:

Valores do teste abaixo de 0,50 não são aceitáveis.

Test values below 0.50 are not acceptable.

Tabela 1 – Esquema Geral de Classificação do Teste de KMO.
Table 1 – General classification scheme of the KMO test.

KMO	Classificação
0,90 - 1,00	Excelente
0,80 - 0,90	Ótimo
0,70 - 0,80	Bom
0,60 - 0,70	Regular
0,50 - 0,60	Ruim
0,00 - 0,50	Inadequado

Valores do teste abaixo de 0,50 não são aceitáveis.

Test values below 0.50 are not acceptable.

3. RESULTADOS E DISCUSSÃO

Primeiramente, observou-se a ordem de grandeza das variáveis, visando detectar discrepâncias que pudessem causar problemas na análise. Na Tabela 2, mostram-se algumas estatísticas descritivas dos dados e é possível perceber que as informações acerca das 32 árvores de quaruba, no que diz respeito à variável volume, apresentam valores mínimos (66,29 m³) e máximos (311,09 m³) bem extremos, indicando possíveis valores discrepantes, enquanto as demais variáveis exibem comportamento dentro dos padrões normais.

Posteriormente, analisou-se a viabilidade da análise fatorial a partir da matriz de correlações (Tabela 3). O primeiro passo foi um exame visual das correlações, verificando-se que existe número substancial de correlações maiores que 0,70 (GORSUCH, 1983). Foram supostas possíveis inter-relações entre as variáveis, o que foi confirmado por meio da aplicação do teste de Correlação de Pearson, com valor de p menor que

Tabela 2 – Estatísticas Descritivas dos dados dendrométricos referente ao Volume, Idade DAP² e Altura de Quaruba (*VACHYSIA INUDATA DUCKE*) na comunidade de Benjamin Constant no município de Bragança em 2001.

Table 2 – Descriptive Statistics of the dendrometric data relating to volume, age, breast height (DBH2), and the quaruba (*Vachysia inudata Ducke*) height in Benjamin Constant community, Bragança municipality, in 2001.

Variáveis	Estatísticas			
	Média	Desvio-Padrão	Mínimo	Máximo
Volume	158,0	63,76	66,29	311,09
Idade	11,26	1,34	9,50	14,8
DAP	19,9	2,49	16,0	25,9
Altura	18,31	3,97	10,6	27,1

Tabela 3 – Matriz de correlação referente aos dados dendrométricos referente à Quaruba (*Vachysia unudata ducke*) na comunidade de Benjamin Constant no município de Bragança em 2001.

Table 3 – Correlation matrix for the quaruba (*Vachysia unudata Ducke*) dendrometric data in Benjamin Constant community, Bragança municipality, in 2001

Dados Dendrométricos	Volume	Idade	Dap	Altura
Volume	1	0,708	0,743	0,894
Idade	0,708	1	0,873	0,801
DAP	0,743	0,873	1	0,847
Altura	0,894	0,801	0,847	1

o nível de significância de 0,01, identificando-se que todas as correlações entre as variáveis são estatisticamente significantes.

De acordo com a Tabela 3, verificou-se que as correlações das variáveis independentes (Idade, DAP e Altura) com a variável dependente (volume) são moderadas acima de 0,70. A variável mais correlacionada com o volume de madeira é a altura ($r = 0,89$).

Foi ajustado um modelo de Regressão Linear Múltiplo para o Volume de madeira de quaruba, seguindo as noções de Campos et al. (1990) e Soares et al. (1996), que utilizaram uma equação de regressão para estimar o volume das árvores em estudo, de acordo com a Tabela 4.

Observou-se, contudo, a presença de multicolinearidade nos dados conforme Neter e Wasserman (1974) e Draper e Smith (1981), ou seja, verificou-se que os coeficientes tanto de Correlação de Pearson ($r = 0,99$) quanto de Determinação ou Explicação ($R^2 = 0,98$) foram substancialmente altos e apenas a variável Altura foi estatisticamente significativa

Tabela 4 – Ajuste Inicial do Modelo de Regressão Linear Múltiplo para Determinação da Equação de Volume de Quaruba (*Vachysia unudata ducke*) na comunidade de Benjamin Constant no município de Bragança em 2001.

Table 4 – Initial adjustment of the Multiple Linear Regression Model for determining the quaruba (*Vachysia unudata Ducke*) volume equation in Benjamin Constant community, Bragança municipality, in 2001.

Variáveis	Coefficientes	Valor de t	Valor de p
Constante	-86,29	-1,71	0,96
Idade	1,47	0,19	0,84
DAP	-2,77	-0,57	0,56
Altura	15,44	6,11	0,01

com valor de p menor que o nível de significância de 0,05; utilizando o Método de Regressão Passo a Passo, também se confirmou que, à medida que se introduzia uma variável por vez no modelo, a natureza de cada uma delas também mudava. Por exemplo, quando se colocou a variável DAP no modelo, esta mudou o sinal das outras variáveis, indicando, assim, forte multicolinearidade nos dados. Com isso, foi realizada a Análise Fatorial com as variáveis originais para obter fatores não correlacionados e, assim, contornar o problema da multicolinearidade, conforme as noções de Elian (1988) e Hair et al. (2005).

O próximo passo é avaliar a significância geral da matriz de correlação com o teste de Esfericidade de Bartlett. Nesse caso, as correlações em geral são significantes a 1% (ver Tabela 4). Entretanto, isso testa apenas a presença de correlações não nulas e não o padrão delas. Assim, para testar se todas as características oriundas de diversas árvores possuem 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. Na Tabela 5, verifica-se que, nesse caso, ocupa um intervalo aceitável (acima de 0,50), conforme Kaiser e Rice (1974), com o valor de 0,74 indicando que as variáveis estão correlacionadas e o modelo fatorial apresentou nível bom de adequação aos dados.

Aplicou-se, assim, a análise de fatores aos dados dendrométricos, utilizando o método das Componentes Principais para extração dos fatores via rotação ortogonal do tipo Varimax para obter melhores combinações e usando como critério para a escolha do número de fatores a extrair o Critério da Raiz Latente (HAIR et al., 2005; MINGOTI, 2005; SANTANA, 2005).

Tabela 5 – Medidas de Avaliação da Adequação da Análise de Fatores para os dados de Quaruba (*Vachysia unudata ducke*) na comunidade de Benjamin Constant no município de Bragança em 200.

Table 5 – Measures for evaluating the appropriateness of the quaruba (*Vachysia unudata Ducke*) data factor analysis in Benjamin Constant community, Bragança municipality, in 2001.

Estatísticas	Coefficientes	Valor de p
Teste de Esfericidade de Bartlett	162,27	0,0012
Teste de Kaiser-Meyer-Olkin (KMO)	0,742	0,0015

Na Tabela 6, mostram-se, também, os autovalores e as percentagens das variâncias explicadas e acumuladas calculadas pelo método Não Rotacionada, contendo as informações sobre os três fatores possíveis e seu poder explanatório relativo. Utilizaram-se os valores da análise do primeiro fator extraído. Percebeu-se que um modelo com apenas um fator seria suficiente para representar a estrutura de covariância inicial, com 3,48% de perda de informações, e que expliquem 96,52% da variabilidade total dos dados originais. Com isso, buscou-se identificar as variáveis que mais influenciavam cada fator, ou seja, as que possuam maiores cargas fatoriais.

Com os resultados obtidos na Tabela 7, pode-se observar que o primeiro fator possui pesos mais altos nas variáveis: DAP, Idade e Altura, sendo identificado como Estrutura Volumar da Árvore. De acordo com a tabela acima, verificou-se que, a soma dos quadrados das cargas fatoriais para cada variável resulta num valor estimado chamado de Comunalidade, 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 na Tabela 7, 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

Tabela 6 – Resultados dos Autovalores para Extração de Fatores, Componentes e Variância Total Explicada pelos Fatores para os dados de Quaruba (*Vachysia unudata ducke*) na comunidade de Benjamin Constant no município de Bragança em 2001.

Table 6 – Eigenvalues results for factors extraction, components, and total variance explained by the quaruba (*Vachysia unudata Ducke*) data factors in Benjamin Constant community, Bragança municipality, in 2001.

Componentes	Autovalores e Variâncias		
	Autovalores	% Variância	Variância Acumulada
1	2,896	96,52	96,52
2	0.088	2,96	99,48
3	0,015	0,51	100,00

Tabela 7 – Matriz de Cargas Fatoriais (Pesos) para os Dados de Quaruba (*Vachysia unudata ducke*) na comunidade de Benjamin Constant no município de Bragança em 2001.

Table 7 – Factors load matrix (weights) of the quaruba (*Vachysia unudata Ducke*) data in Benjamin Constant community, Bragança municipality, in 2001.

Características	Fator 1	Comunalidade
Diâmetro à Altura do Peito (DAP)	0,991	0,982
Idade	0,987	0,974
Altura	0,970	0,940

pelos escores fatoriais e realizou-se a Regressão Linear Simples. Considerou-se como variável resposta o Volume de Quaruba e como variáveis explicativas os escores fatoriais referentes ao fator extraído.

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 8.

Após estudar a significância individual dos coeficientes de acordo com o teste t e por meio do valor de p, analisou-se a significância geral do Modelo de Regressão aplicando o teste F aos coeficientes (GUJARATI, 2000).

De acordo com o teste F, rejeita-se a hipótese de não haver regressão, isto é, o modelo é significativo a 0,05 de significância, e conclui-se que pelo menos uma variável explanatória esteja relacionada com o valor de volume de madeira (Tabela 9).

Analisando as medidas de ajuste, observou-se que os valores do coeficiente de correlação múltiplo (R múltiplo), o coeficiente de determinação (R²) e o

Tabela 8 – Ajuste do Modelo de Regressão Linear simples para Determinação da Equação de Volume de Quaruba (*Vachysia unudata ducke*) na comunidade de Benjamin Constant no município de Bragança em 2001.

Table 8 – Setting of the Simple Linear Regression Model for determining the quaruba (*Vachysia unudata Ducke*) volume equation in Benjamin Constant community, Bragança municipality, in 2001.

Variáveis	Coefficientes	Valor de t	Valor de p
Constante	0,506	48,977	0,000
Fator 1	0,289	27,521	0,000

Tabela 9 – Análise de Variância para a Significância da Equação de Volume de Quaruba de Quaruba (*Vachysia unudata ducke*) na comunidade de Benjamin Constant no Município de Bragança em 2001.

Table 9 – Variance analysis of the quaruba (*Vachysia unudata Ducke*) volume equation's significance in Benjamin Constant community, Bragança municipality, in 2001.

EV	GL	S.Q	Q.M	F	Valor de p
Regressão	1	2,583	2,583	757,401	0,000
Resíduo	30	0,102	0,003		
Total	31	2,685			

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, o que evidencia que o modelo ajustado explica bem a variabilidade do valor do volume de madeira de quaruba.

Em suma, o coeficiente de determinação, que representa a proporção da variação em Y (*Volume*) que é explicada através do conjunto de variáveis explanatórias selecionadas, apresentou valor igual a 96% da variação no volume que pode ser explicado através da variação das variáveis, e 4% do volume é explicado por outras variáveis que não constam no modelo (Tabela 10).

A equação, portanto, de regressão linear simples dos dados de volume em função do DAP, Idade e Altura e que descreve o relacionamento entre o volume e o fator independente é:

$$\text{VOLUME} = 0,506 + 0,289 \cdot \text{Escore do Fator 1.}$$

sendo o escore do fator 1 a combinação linear entre as variáveis DAP, Idade e Altura, obtidas por meio de um método de regressão linear simples e desmembrado da seguinte forma:

Tabela 10 – Medidas Referentes Ao Ajuste da Equação de Volume de Madeira de Quaruba de dados de Quaruba (*Vachysia unudata ducke*) na comunidade de Benjamin Constant no Município de Bragança em 2001.

Table 10 – Measures relating to adjustment of the quaruba (*Vachysia unudata Ducke*) volume equation in Benjamin Constant community, Bragança municipality, in 2001.

Medidas de Ajuste	Valores
Coeficiente de Correlação Múltiplo (R)	0,981
Coeficiente de Determinação (R^2)	0,962
Coeficiente de Determinação Ajustado (R^2)	0,961

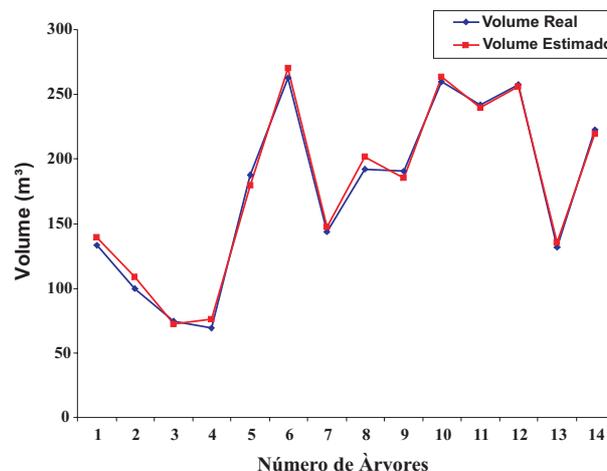


Figura 1 – Volume Real e Estimado de Quaruba (*Vachysia unudata ducke*) na comunidade de Benjamin Constant no município de Bragança em 2001.

Figure 1 – Real and estimated volume of the quaruba (*vachysia unudata ducke*) in Benjamin Constant community, Bragança municipality, in 2001.

$$\text{VOLUME} = 0,506 + 0,289 \cdot [(0,991 \cdot \text{DAP}) + (0,987 \cdot \text{Idade}) + (0,97 \cdot \text{Altura})]$$

Assim, a equação estimada do volume de quaruba em função de suas características pode ser usada para prever o volume de outras árvores da espécie *Vochysia* na mesma região da comunidade de Benjamin Constant, no Município de Bragança, e dosséis com características idênticas ao usado no experimento, sem o problema da multicolinearidade.

3.1. Validação do modelo ajustado

Utilizando os dados observados em 14 pontos de amostragem, selecionados aleatoriamente em fragmentos na comunidade de Benjamin Constant, no Município de Bragança, em 2001, e que não foram utilizados nos ajustes do modelo avaliado, procedeu-se à validação do modelo selecionado para estimar o volume total de quaruba.

De acordo com a Figura 1, verificou-se que o modelo de regressão aplicado aos dados após a utilização da análise fatorial obteve grande poder de explicação e predição, pois os volumes reais observados na amostra ficaram próximos do volume estimado, sendo o modelo proposto o mais adequado para estimar os volumes totais das árvores da espécie do tipo *Vachysia unudata* Ducke.

4. CONCLUSÃO

Este trabalho teve como objetivo a aplicação conjunta das técnicas de análise multivariada para determinação do volume de madeira de quaruba (*Vachysia unudata* Ducke) na Comunidade de Benjamin Constant, no Município de Bragança. A transformação e redução do número de variáveis podem ser obtidas pelo método da análise de fatores, sem perda significativa de informações, obtendo-se variáveis não correlacionadas, por meio de um fator. Com os resultados da análise fatorial, partiu-se para a construção do modelo de regressão linear simples do volume de quaruba, evitando o problema da multicolinearidade. No modelo ajustado, obteve-se regressão com R^2 maior que 96%, o que garante a consistência da regressão para estimativa do volume de madeira somente daquelas árvores com características regionais.

5. REFERÊNCIAS

- 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 Brando.** 1993. FOLHAS. Dissertação (Mestrado em Ciência Florestais) - ISA/UTL, LOCAL, 1993.
- BARROSO, L. P.; ARTES, R. **Análise multivariada.** São Paulo: IME-USP, 2003.
- CAMPOS, J. C. C. et al. Relações entre espaçamento, volume e peso de madeira em plantações de eucaliptos. **Revista Árvore**, v.14, n.1, p.119-133, 1990.
- DRAPER, N. R.; SMITH, H. **Applied regression analysis.** New York: John Wiley & Sons, 1981.
- DILON, W.R.; GOLDSTEIN, M. **Multivariate Analysis: methods and applications.** New York: John Wiley & Son, 1984.
- ELIAN, S. N. **Análise de regressão.** São Paulo: IME, 1998.
- GUIMARÃES, D. P.; LEITE, H. G. Influência do número de árvores na determinação de equação volumétrica para *Eucalyptus grandis*. **Scientia Forestalis**, n.50, p.37-42, 1996.
- GUJARATI, D. N. **Econometria básica.** São Paulo: Makron Book, 2000.
- GORSUCH, R. L. **Factor analysis.** New Jersey: Lawrence Erlbaum, 1983.
- HUSCH, B.; MILLER, C. I.; BERRS, T. W. **Forest mensuration.** 2.ed. New York: The Ronald Press Company, 1972.
- HAIR JR, J. F. et al. **Análise multivariada de dados.** 5.ed. Porto Alegre: Bookman, 2005.
- INSTITUTO BRASILEIRO DE DESENVOLVIMENTO FLORESTAL – IBDF. **Madeiras da Amazônia: características e utilização.** Brasília: 1981. v.1.
- JOHNSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis.** [S.l.]: Prentice-Hall, 1992.
- KAISER, H. F. The varimax criterion for analytic rotation in factor analysis. **Psychometrika**, v.23, p.187-200, 1958.
- KAISER, H. F.; RICE, J. Little Jiffy, mark IV. **Education and Psychological Measurement**, v.34, p.111–117, 1974.
- MALHOTRA, N. K. **Marketing Research: an applied orientation.** New Jersey: Prentice Hall, 2001.
- MACHADO, S.A.; FIGUEREDO FILHO, A. **Dendrometria.** 2.ed. Guarapuava: Unicentro, 2006.
- MELLO, M. N.; RAMOS, E. M. I. S.; ALMEIDA, S. S. . Modelagem de árvores quaruba via gráfico de controle de regressão funcional. In: ESCOLA DE MODELOS DE REGRESSÃO, 11., 2009, Recife. São Paulo: Associação Brasileira de Estatística, 2009.
- MINGOTI, S. A. **Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada.** Belo Horizonte: UFMG, 2005.
- NETER, J.; WASSERMAN, W. **Applied linear statistical models.** Illinois: Richard D. Irwin, 1974.
- PAULA NETO, F. Tabelas volumétricas com e sem casca para *Eucalyptus saligma*. **Revista Árvore**, v.1, n.1, p.31-54, 1977.

- RIOS, M. et al. **Benefícios das plantas da capoeira para a comunidade de Benjamin Constant, Pará, Amazônia Brasileira**. Belém: CIFOR, 2001.54p.
- REIS, E. **Estatística multivariada aplicada**. 2.ed. Lisboa: Silabo, 2001.
- RENCHER, A. **Methods of multivariate analysis**. 2.ed. New York: John Wiley & Son, 2002.
- REIS, C. P.; TORRES, R. V.; ALMEIDA, S. S. Comparação dos estimadores de regressão clássica e funcional aplicado ao estudo das árvores quaruba. In: ESCOLA DE MODELOS DE REGRESSÃO, 11., 2009, Recife. São Paulo: Associação Brasileira de Estatística, 2009.
- SANTANA, A. C. Análise do desempenho competitivo das agroindústrias de frutas do Estado do Pará. In: CONGRESSO BRASILEIRO DE ECONOMIA E SOCIOLOGIA RURAL, 44., 2006, Fortaleza. **Anais...** Brasília, DF: Sober, 2006.
- SANTANA, A. C. Elementos de economia, agronegócio e desenvolvimento local. Belém: GTZ; TUD; UFRA, 2005. (Série Acadêmica, 1).
- SPURR, S. H. **Forest inventory**. New York: The Ronald Press, 1952.
- SOARES, C. P. B. et al. **Modelos para estimar a biomassa da parte aérea em um povoamento de Eucalyptus gradis na região de Viçosa, Minas Gerais**. *Revista Árvore*, v.20, n.2, p.179-189, 1996.
- SOARES, C. P. B.; PAULA NETO, F.; SOUZA, A. L. **Dendrometria e inventário florestal**. Viçosa, MG: Universidade Federal de Viçosa, 2007.
- TOME, J. A. L. **Estimação do volume total, de volumes mercantis e modelação do perfil do tronco em Eucalyptus globulus Labill**. 1990. FOLHAS. Dissertação (Mestrado em Ciências Florestais) - ISA/UTL, LOCAL, 1990.
- TORRES, R. V. ; REIS, C. P. ; ALMEIDA, S. S. . Modelagem do volume da quaruba via regressão linear. In: ESCOLA DE MODELOS DE REGRESSÃO, 11., 2009, Recife. São Paulo: Associação Brasileira de Estatística, 2009.
- VALENTE, M. D. R. ; PNHEIRO, J. G.; QUEIROZ, W. T. . Determinação e modelagem do volume total de Pinus carabea var Bahamensis; uma abordagem multivariada. In: ENCONTRO REGIONAL DE MATEMÁTICA APLICADA E COMPUTACIONAL, 8., 2008, Natal. **Anais...** Natal: Sociedade Brasileira de Matemática Aplicada e Computacional, 2008.
- YARED, J. A. G. et al. **Crescimento de quaruba-verdadeira em diferentes métodos de regeneração artificial**. Belém: Embrapa-CPATU, 1983. 2p. (Pesquisa em Andamento, 105).

