

LEONARDO MENDES DE LIMA JUNIOR

**UTILIZAÇÃO DE TÉCNICAS MULTIVARIADAS NA
CLASSIFICAÇÃO DE FASES DE CRESCIMENTO DA
LEUCENA**

RECIFE

2006

UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
Programa de Pós-Graduação em Biometria

LEONARDO MENDES DE LIMA JUNIOR

**UTILIZAÇÃO DE TÉCNICAS MULTIVARIADAS NA
CLASSIFICAÇÃO DE FASES DE CRESCIMENTO DA
LEUCENA**

Dissertação apresentada à
Universidade Federal Rural de
Pernambuco para obtenção do título
de Mestre em Biometria.

Orientador: Prof. DS. Rinaldo Luiz
Caraciolo Ferreira.

Co-orientador: Prof. PhD. José
Antônio Aleixo da Silva.

RECIFE

2006

Ficha catalográfica
Setor de Processos Técnicos da Biblioteca Central – UFRPE

L732u Lima Junior, Leonardo Mendes
Utilização de técnicas multivariadas na classificação de fases de crescimento da leucena / Leonardo Mendes de Lima Junior - 2006.
78f.

Orientador: Rinaldo Luiz Caraciolo Ferreira
Dissertação (Mestrado em Biometria) -- Universidade Federal Rural de Pernambuco. Departamento de Estatística e Informática.
Inclui bibliografia.

CDD 574.018 2

1. Análise multivariada
 2. Sítios florestais
 3. *Leucaena leucocephala*
 4. Análise de agrupamento
 5. Análise fatorial
- I. Ferreira, Rinaldo Luiz Caraciolo
 - II. Título

UTILIZAÇÃO DE TÉCNICAS MULTIVARIADAS NA CLASSIFICAÇÃO DE FASES DE CRESCIMENTO DA LEUCENA

LEONARDO MENDES DE LIMA JUNIOR

Dissertação defendida e aprovada em 30/05/2006 pela Banca Examinadora

Orientador

Prof. DS. Rinaldo Luiz Caraciolo Ferreira – UFRPE

Examinadores

Prof. PhD. José Antônio Aleixo da Silva – UFRPE

Prof. PhD. Borko Stosic – UFRPE

Prof. PhD. Mário de Andrade Lira Júnior – UFRPE

Recife-PE

AGRADECIMENTOS

A Deus, por permitir minha existência durante esses anos de vida.

Aos meus pais, pela minha formação pessoal, amor, coragem e renúncia. Sempre presentes nos momentos mais difíceis da minha vida.

A minha esposa Pryscilla Lima, pelo companheirismo, compreensão, apoio e incentivo.

Ao Departamento de Estatística e Informática, da Universidade Federal Rural de Pernambuco, pela oportunidade de ser aluno deste curso.

Ao coordenador do curso de Pós-Graduação em Biometria, prof. Dr. Eufrásio de Souza Santos pela dedicação e esforço incansável pelo curso.

Ao prof. Dr. Rinaldo Luiz Caraciolo Ferreira, orientador. Pela clareza e objetividade das informações dadas, pela paciência e sinceridade.

Ao prof. Dr. José Antônio Aleixo da Silva, co-orientador. Pelos incentivos prestados, apoio e amizade.

Ao corpo docente do mestrado em Biometria pelas informações adquiridas e que foram de grande relevância para minha formação durante o curso.

Aos colegas e amigos que estiveram comigo neste curso, convivendo e aprendendo: Ady Marinho, Ilzes Celi, Franklin Tupinambá, Arundo Nunes, Fábio Cavalcanti, Cícero Carlos, Hérbetes Cordeiro, Heliovânio, Sérgio Leitão, Carlos e Dâmocles Aurélio.

A Empresa NeuroTech Ltda, pela oportunidade de trabalho como estatístico e pela concessão de alguns expedientes para conclusão deste trabalho.

Ao agrônomo Ivan Ferraz, da Estação Experimental de Caruaru, pela exposição das leucenas, do experimento ao qual foi estudado.

A todos os que não foram citados, mas que de contribuíram de forma direta ou indiretamente para a realização deste trabalho. Meus sinceros agradecimentos.

*“Nenhum vento ajuda a quem não
sabe para que porto velejar.”
(Montaigne)*

LIMA JUNIOR, L.M. Utilização de técnicas multivariadas na classificação de fases de crescimento da leucena. 2006. Orientador: Prof. DS. Rinaldo Luiz Caraciolo Ferreira. Co-orientador: Prof. PhD. José Antônio Aleixo da Silva.

RESUMO

O objetivo deste trabalho foi identificar as fases de crescimento da leucena (*Leucaena leucocephala* (Lam.) de Wit.), ao longo do tempo, por meio da utilização das técnicas multivariadas. Os dados foram obtidos de um experimento, localizado na Estação Experimental de Caruaru-PE, que pertence à Empresa Pernambucana de Pesquisa Agropecuária – IPA. Nesse estudo, consideraram-se além das 20 medições em altura (em m) ao longo do tempo, outras mais 17 medições, interpoladas através do Modelo de Weibull e também por interpolação do tipo linear, dividido em 2 grupos distintos de tratamentos: segundo a presença ou não de composto de resíduo orgânico. A análise fatorial permitiu a redução da dimensionalidade das medições, em função de 3 fatores, com autovalores superiores a uma unidade e percentual do total da variação explicada de 94,60% e 94,30%, nos tratamentos sem e com composto orgânico, respectivamente, usando a rotação varimax. Dos fatores retidos, foram obtidos os escores e submetidos à análise de agrupamento do tipo k-médias, sendo o número de grupos escolhidos previamente, para de k de 3 a 10. Os grupos de melhor classificação foram verificados na análise discriminante, que avaliou a eficiência destes agrupamentos, em 95,2% e 95,1% de percentual de classificação correta. O estudo mostrou que as plantas que receberam o composto orgânico tiveram as maiores alturas, em média. A análise gráfica permitiu a comparação entre os dados que tiveram tratamento com composto orgânico e sem composto orgânico, conforme os grupos, em sua fase de crescimento.

LIMA JUNIOR, L.M. The use of multivariate analysis on classification in leucena (*Leucaena leucocephala* (Lam.) of Wit) growth phases. 2006. Adviser: Prof. DS. Rinaldo Luiz Caraciolo Ferreira. Co-adviser: Prof. PhD. José Antônio Aleixo da Silva.

ABSTRACT

The aim of this work was to identify the growth phases of leucena (*Leucaena leucocephala* (Lam.) of Wit.), as a function of time, by using of the multivariate analysis techniques. The data set was obtained through an experiment performed at the Experimental Station of Caruaru-PE of the Institute of Agricultural Research - IPA. Besides 20 actual measures of height along time, additional 17 values were used, obtained by interpolation of best fit curves to the Weibull Model, as well as by linear type interpolation, for two distinct treatments groups: according to presence or absence of organic compound residue. The factor analysis was used to reduce the dimensionality of measured data to three factors with eigenvalues higher than unity, explaining 94.60% and 94.30% of the total variation for the treatments with and without organic compound, respectively, using the varimax rotation. The resulting factor scores were subjected to k-means cluster analysis, using previously selected, number of groups k from 3 to 10. The discriminant analysis was then employed to verify the efficiency of clustering of the best classification groups, which was found to be 95.2% and 95.1% of correct classification for the two treatments. The study it showed that the leucena trees which has received with treatment with organic on average attained greater height. The graphical analysis allowed the comparison among the data treatment with and without organic compound treatment, according to groups, in growth phase.

LISTA DE FIGURAS

Figura		Página
1	Gráfico Scree Plot dos Autovalores do Modelo em Função do Número de Fatores.....	31
2	Crescimento em altura da <i>Leucaena leucocephala</i> (Lam.) de Wit, ao longo do tempo, no tratamento sem composto orgânico, no agrupamento com 3 grupos.....	59
3	Crescimento em altura da <i>Leucaena leucocephala</i> (Lam.) de Wit, ao longo do tempo, no tratamento sem composto orgânico, no agrupamento com 9 grupos.....	59
4	Crescimento em altura da <i>Leucaena leucocephala</i> (Lam.) de Wit, ao longo do tempo, no tratamento com composto orgânico, no agrupamento com 10 grupos.....	67

LISTA DE TABELAS

Tabela		Página
1	Quadrante do ângulo 4ϕ em função dos sinais do numerador e denominador de (12).....	34
2	Classificação do ajuste do modelo de análise fatorial em relação ao valor da estatística de Kaiser-Meyer-Olkin.....	36
3	Tratamentos aplicados no povoamento de <i>Leucaena leucocephala</i> (Lam.) de Wit., em Caruaru, PE.....	46
4	Medições em altura, segundo a data, idade em meses e dias após o plantio.....	47
5	Parâmetros estimados e Índice de Ajuste, segundo o modelo de Weibull aplicado nos tratamentos sem composto e com composto orgânico (BRITO, 2005).....	48
6	Autovalores e Percentual de Explicação da Variância dos Fatores por Componentes Principais e na Rotação Varimax, no Tratamento Sem Composto Orgânico.....	52
7	Matriz de cargas fatoriais, comunalidades e variâncias específicas segundo a idade da <i>Leucena leucocephala</i> (Lam.) de Wit., obtidos na Rotação Varimax, no Tratamento Sem Composto.....	53
8	Número de plantas agrupadas pela técnica k-médias para de k de 3 a 10 no tratamento sem o composto orgânico.....	54
9	Resultado da Análise Discriminante, segundo Percentual de Classificação Correta dos Grupos, para a <i>Leucena leucocephala</i> (Lam.) de Wit., no tratamento sem Composto Orgânico.....	55

Tabela	Página
10 Médias e Desvios-padrão das alturas da <i>Leucaena leucocephala</i> (Lam.) de Wit., agrupados em 3 grupos, de acordo com a idade, para o tratamento sem composto orgânico.....	56
11 Médias e Desvios-padrão das alturas da <i>Leucaena leucocephala</i> (Lam.) de Wit., agrupados em 9 grupos, de acordo com a idade, para o tratamento sem composto orgânico.....	57
12 Autovalores e percentual de explicação da variância na extração dos fatores por componentes principais e na rotação varimax, no tratamento com composto orgânico.....	61
13 Matriz de cargas fatoriais, comunalidades e variâncias específicas segundo a idade da <i>Leucaena leucocephala</i> (Lam.) de Wit., obtidos na Rotação Varimax, no Tratamento Com Composto.....	62
14 Número de plantas agrupadas obtidas no agrupamento do k-médias para de k de 3 a 10 no tratamento com o composto orgânico.....	63
15 Percentual de Classificação dos Grupos Obtidos, usando Análise Discriminante para as leucenas com Composto Orgânico.....	64
16 Médias e Desvios-padrão das alturas da <i>Leucaena leucocephala</i> (Lam.) de Wit., agrupados em 10 grupos, de acordo com a idade, para o tratamento com composto orgânico.....	65

SUMÁRIO

	Página
1. INTRODUÇÃO	13
2. REVISÃO DE LITERATURA	15
2.1 LEUCENA.....	15
2.2 CLASSIFICAÇÃO DE SÍTIOS FLORESTAIS.....	18
2.3 MÉTODOS DE DETERMINAÇÃO DA QUALIDADE DO SITIO....	19
2.4 USO DA ESTATÍSTICA MULTIVARIADA EM CLASSIFICAÇÃO DE SÍTIOS.....	22
2.5 ANÁLISE FATORIAL.....	25
2.5.1 Definição do Modelo Análise Fatorial	25
2.5.2 O Modelo Matemático	25
2.5.3 As Comunalidades e Especificidades	27
2.5.4 Métodos de Estimação dos Fatores	28
2.5.5 Extração dos Fatores por Componentes Principais ...	29
2.5.6 A Escolha do Número de Fatores	30
2.5.7 A Rotação dos Fatores	32
2.5.8 Testes de Adequabilidade do Modelo	35
2.5.8.1 Teste KMO.....	35
2.5.8.2 O Teste de Esfericidade de Bartlett (χ^2).....	36
2.5.9 Os Escores Fatoriais	37
2.5.9.1 Método dos Mínimos Quadrados Ponderados...	37
2.5.9.2 Método de Regressão.....	38
2.6 ANÁLISE DE AGRUPAMENTO.....	40
2.6.1 Análise de Agrupamento por k-médias	41
2.7 ANÁLISE DISCRIMINANTE.....	42
3. MATERIAL E MÉTODOS	45
3.1 DADOS UTILIZADOS.....	45
3.2 O MÉTODO ESTATÍSTICO.....	48
3.2.1 Interpolação de Medidas com o Modelo de Weibull	48
3.2.2 Análise Multivariada	49

	Página
4. RESULTADOS E DISCUSSÃO	52
4.1 ANÁLISE DOS INDIVÍDUOS NA AUSÊNCIA DE COMPOSTO ORGÂNICO.....	52
4.2 ANÁLISE DOS INDIVÍDUOS NA PRESENÇA DE COMPOSTO ORGÂNICO.....	61
5. CONCLUSOES	68
6. REFERÊNCIAS	69

1. INTRODUÇÃO

A cada dia expandem-se as áreas sem vegetação arbórea, seja para utilização agrícola e ou pecuária, e praticamente não existe a preocupação com áreas de florestas, ou com a recuperação de áreas degradadas através de reflorestamento. Entre as espécies mais promissoras para este fim, encontra-se a leucena (*Leucaena leucocephala*); uma leguminosa que pertence a tribo Mimosae, subfamília Mimosideae, de boa adaptação em todas as regiões brasileiras, na produção de florestas energéticas, contribuindo por completo com produtividade e ajudando a estabelecer enriquecer o solo (OLIVEIRA et al., 2004).

A leucena cresce bem nas regiões tropicais, em diferentes tipos de ambientes, apresentando um sistema radicular profundo, alta taxa de crescimento, resistência às pragas e doenças, simbiose com bactérias fixadoras de nitrogênio e tolerância a seca FREITAS et al. (1991). Mesmo havendo grande interesse dos produtores pela leucena na região Nordeste, a sua utilização ainda é bastante restrita, talvez devido a falta tradição no uso de leguminosas (SILVA, 1992).

Na implantação de um povoamento florestal, o manejador precisa levar em consideração, entre outros fatores, as condições gerais de crescimento e produção (sítios), dentro de uma mesma área. Nesse caso, a classificação em áreas homogêneas permite inferir sobre a capacidade produtiva de classe de sítio e fundamentar a elaboração e execução dos planos de manejo.

Dentre os principais métodos usuais de classificação de sítio, podem-se citar: medições de fatores do sítio, como variáveis climáticas, edáficas e da própria vegetação; e medições de características dos povoamentos florestais, como a altura dominante, o estoque de madeira produzida, as características de árvores e a ocorrência de plantas indicadoras (HUSCH et al., 1993).

Na classificação de sítio há necessidade de utilização de várias variáveis e, conseqüentemente, a aplicação de técnicas estatísticas apropriadas para o estudo das inter-relações entre os diversos fatores, evidenciando-se a importância da aplicação de técnicas multivariadas.

As técnicas de análise multivariada possibilitam avaliar um conjunto de características, levando em consideração as correlações existentes, o que permite que interferências sobre o conjunto de variáveis sejam feitas em um nível de significância conhecido.

Segundo Burton et al. (1991), as técnicas estatísticas multivariadas têm sido amplamente empregadas em estudos envolvendo simultaneamente variáveis de clima, solo, relevo, vegetação e geologia na classificação de áreas com cobertura florestal. Essas técnicas são utilizadas com objetivos básicos de ordenamento, visando determinar a influência de fatores do meio na composição e produtividade do local, e de agrupamento, com o propósito de classificação.

Segundo Anderson (1984), existe, basicamente, duas formas de classificar as análises multivariadas: as que permitem extrair informações a respeito da independência entre as variáveis que caracterizam cada elemento, tais como análise fatorial, análise de agrupamento, análise canônica, análise de ordenamento multidimensional e análise de componentes principais; e as que permitem extrair informações a respeito da dependência entre uma ou mais variáveis ou uma com relação à outra, tais como análise de regressão multivariada, análise de contingência múltipla, análise discriminante e análise de variância multivariada.

Dentre as técnicas estatísticas multivariadas, as técnicas de análises de componentes principais, de agrupamento, fatorial e discriminante são as mais utilizadas na classificação de sítios (ASPIAZÚ, 1979; QUEIROZ, 1984; SOUZA et al., 1990; BRAGA, 1997; SOUZA et al., 1997; SALOMÃO, 1998).

Por isso, o objetivo deste trabalho foi identificar as fases de crescimento da leucena ao longo do tempo, empregando-se as técnicas multivariadas de análise fatorial, de agrupamento e discriminante.

2. REVISÃO DE LITERATURA

2.1 LEUCENA

A leucena (*Leucaena leucocephala*) é uma leguminosa arbustiva, perene, que apresenta raízes profundas, característica que lhe confere excelente tolerância à seca. Altamente palatável, produz elevadas quantidades de forragem com altos teores de proteína e minerais, notadamente em solos de alta fertilidade natural, sendo, portanto, uma alternativa de baixo custo para a substituição parcial dos produtos comerciais comumente utilizados na suplementação animal (PEREIRA et al., 2004, citando COSTA, 1987).

A importância econômica dessa espécie se deve ao seu valor como árvore de sombreamento e adubo verde, além do interesse na sua utilização nos trópicos para restauração da fertilidade do solo, forragem e recuperação de áreas degradadas (FREITAS et al., 1991).

Até o momento, foram classificadas doze espécies do gênero *Leucaena*, sendo *Leucaena leucocephala* a que apresenta maior importância em nível internacional (Brewbaker, 1987). Segundo Seiffert e Thiago (1983), as leucenas diferem grandemente em porte, sendo conhecidas mais de 100 variedades que são agrupadas em três tipos: Havaiano, Salvadorenho e Peruano.

As do tipo havaiano, o qual foi estudado no presente trabalho, são variedades arbustivas com até 5 m de altura, que florescem jovens (com 4 a 6 meses). O florescimento ocorre durante todo o ano e apresenta pouca produção de madeira e folhas, e sua produção, abundante de sementes, pode tornar esta planta uma invasora. É comumente encontrada na costa do México, tendo sido largamente dispersada nos trópicos.

Como forrageira, a leucena é altamente palatável e de grande valor nutritivo. A folhagem e os frutos mais novos chegam a apresentar teores protéicos de 35%, enquanto na folhagem mais velha este teor fica em torno de 25%. Alguns autores australianos relatam que a folhagem de leucena é tóxica quando ministrada como alimento único por período prolongado, pela grande quantidade de mimosina existente na composição dessa forragem.

Entretanto, no Brasil, a ocorrência de intoxicações é praticamente inexistente, devido à existência de bactérias que digerem satisfatoriamente a mimosina no rúmen dos animais. (DRUMOND, 2001).

Já Franco e Souto (1986), relatam que a mimosina, presente na leucena causa a queda de pelos, quando ministrada acima de 50% da dieta de forma contínua, por períodos longos, especialmente a não ruminantes ou ruminantes ainda não adaptados a tê-la em sua dieta. Para os autores, o uso da leucena como ração para ruminantes deve ser introduzida aos poucos, devendo atingir um máximo de 25 a 30% da dieta.

Souza (2003) relata que o sistema radicular profundo propicia a reciclagem de nutrientes de subsolo, como também a absorção de água das camadas profundas, fazendo com que a planta cresça e produza massa verde durante todo o ano.

As leucenas crescem nos trópicos e subtropicais em regiões de até 500 m de altitude, suportando grandes diferenças de precipitação, luminosidade, salinidade do solo, inundações periódicas, fogo, geadas leves e seca. O seu melhor desenvolvimento, no entanto, é obtido em áreas onde chove de 600 a 1.700 mm suportando bem épocas curtas de estiagem. É uma planta que prefere insolação direta, perdendo as folhas na sombra e com geadas leves, rebrotando, no entanto, logo após a sua ocorrência. (SEIFFERT e THIAGO, 1983).

Drumond (2001), afirma que o crescimento da leucena é uma espécie de rápido crescimento, chegando a crescer até 3 metros de altura no primeiro ano, e com grande capacidade de regeneração.

Porém, Seiffert e Thiago (1983), relatam que o crescimento da leucena nos três primeiros meses é bastante lento e o seu estabelecimento é por semeadura direta, devendo ser semeada no início do período chuvoso (outubro/novembro). O espaçamento e a densidade de semeadura variam de acordo com o objetivo da utilização.

Como a leucena é uma planta bastante perseguida por formigas, cupins, lagartas e herbívoros silvestres, a fase que decorre entre a semeadura e os primeiros 90 dias é bastante delicada, exigindo, em nosso meio, freqüente replantio (SEIFFERT e THIAGO, 1983).

Para controle de invasoras, torna-se necessária a realização de 3 ou mais capinas, até que as plantas atinjam 1 metro de altura, quando terão rápido crescimento, cobrindo o terreno. Em áreas onde o plantio direto torna-se dificultado devido às pragas, é possível efetuar o plantio, usando-se mudas previamente desenvolvidas em sacos plásticos com 1 a 2 kg de solo (SEIFFERT e THIAGO, 1983).

Com este sistema podem ser dispensados cuidados especiais às mudas na sua fase inicial e, quando apresentarem porte de 15 a 20 cm, serem levadas ao campo já em condições de resistirem ao ataque das pragas, diminuindo a necessidade de replantios. (SEIFFERT e THIAGO, 1983).

Souza (2003) em seu experimento, afirmou que as leucenas atingiram sua altura máxima, em 8 anos de pesquisa. O mesmo autor observou que, em todos os tratamentos aplicados, as árvores cresceram rapidamente nos primeiros anos, atingindo em média 60% do seu crescimento máximo nos primeiros 2 anos e 40% nos outros 6 anos.

A leucena é uma planta de crescimento inicial lento, recomendando-se seu plantio por mudas. Assim, no semi-árido do Nordeste, onde o problema é agravado devido ao pouco desenvolvimento da planta na estação seca, quando do plantio direto por sementes, a implantação de áreas de leucena por meio de mudas é mais recomendável para rápida obtenção de estandes uniformes e vigorosos (TELES, et al., 2000) citando (SOUZA, 1990; NASCIMENTO, 1982; XAVIER, 1989).

Nas condições do semi-árido do Nordeste brasileiro, a leucena não tem apresentado problemas com doenças que limitem o seu cultivo. No semi-árido do Nordeste brasileiro, o comportamento silvicultural da leucena, destaca-se em relação a outras espécies cultivadas, principalmente às nativas, em especial quando se consideram a densidade da madeira, rendimento gravimétrico de carbonização, teores de carbono fixo e cinza apresentado pelas espécies (DRUMOND, 2001).

2.2 CLASSIFICAÇÃO DE SÍTIOS FLORESTAIS

O sítio pode ser considerado sob dois pontos de vista: como uma unidade geográfica uniforme, caracterizada por uma certa combinação estável dos fatores do meio e na qual se pode esperar o desenvolvimento de uma cobertura vegetal mais ou menos homogênea (SPURR e BARNES, 1980); ou como a capacidade produtiva de uma dada área florestal, determinada pela ação e interação de fatores do meio e influenciadas pelas práticas de manejo florestal (BARROS, 1974). No entanto Kilian (1984), citado por Scolforo (1992), apontou que esses conceitos são complementares e que uma classificação de sítio, do ponto de vista ecológico, pode e deve preceder à determinação da capacidade produtiva desses locais.

A princípio pode ser relativamente simples e fácil enumerar fatores do ambiente onde se desenvolve e o crescimento das árvores. Entretanto, pode ser difícil entender e avaliar o somatório de interações entre esses fatores e seus efeitos sobre o crescimento da floresta, no complexo denominado sítio.

Para estudo e compreensão, os fatores do meio podem ser divididos em categorias amplas, envolvendo clima (radiação, temperatura, umidade relativa do ar, precipitação, vento etc.), topografia (fisiografia, exposição, inclinação etc.), solo (textura, estrutura, regime térmico e hídrico, minerais, pH, matéria orgânica etc.), fatores bióticos (plantas, animais, insetos, fungos) e antrópicos.

As inter-relações entre esses fatores resultam em recursos diretamente disponíveis para o crescimento das árvores, como luz, calor, água, fatores químicos (CO_2 , O_2 , fertilidade, compostos alelopáticos etc.) e fatores mecânicos (danos causados pelo fogo, geada, animais, atividades impactantes, manejo florestal etc.), (SPURR e BARNES, 1980).

A classificação de uma floresta em relação à sua produtividade é de grande importância, pois a produtividade do sítio está relacionada com diferentes respostas de certas culturas, que influem diretamente sobre a viabilidade de empreendimentos florestais; com as pesquisas ecológicas, visando a estudos de luminosidade, biomassa e exportação de nutrientes dos produtos em diferentes idades; na adoção de técnicas de manejo e conservação do solo e no planejamento da produção por meio da idade ótima e econômica de corte (TONINI et al., 2002).

Nesse contexto, o cientista florestal dedica-se ao segmento floresta, que é apenas uma parte do ecossistema e tem como objetivo integrar todos os fatores do

sítio para produzir uma estimativa da qualidade do sítio florestal, buscando parâmetros de crescimento na própria vegetação, visto que os ecossistemas funcionam como um complexo ciclo de energia, marcado por perdas e ganhos.

Para Schneider (1993), os principais objetivos da avaliação da qualidade de sítios são: a) as estimativas do rendimento global dos povoamentos; b) o planejamento e execução de trabalhos de pesquisa como, por exemplo, os desbastes, que são cortes intermediários; c) a programação e execução dos trabalhos de manutenção (limpezas) das plantas existentes; d) a extensão da classificação da qualidade de sítio em áreas a serem plantadas para seleção adequada das espécies.

2.3 MÉTODOS DE DETERMINAÇÃO DA QUALIDADE DO SÍTIO

Com relação ao meio físico, é de grande importância para o manejo e planejamento florestal estimar e inventariar a capacidade produtiva do sítio, razão pela qual é recomendável a existência de métodos alternativos para classificar e mapear os sítios. Segundo Lepsch et al. (1983), a classificação de qualquer objeto tem por finalidade ordenar os conhecimentos a seu respeito de maneira simples e precisas. Objetos iguais ou semelhantes em suas características e propriedades são agrupados nas mesmas classes.

Existem várias alternativas para efetuar a classificação da produtividade dos locais podendo-se citar os fatores edáficos, fatores climáticos, fatores fisiográficos, a vegetação rasteira, multifatorial de sítio, registro histórico e elementos dendrométricos (volume, área basal e altura).

A estimativa da qualidade florestal pode ser feita de dois caminhos: primeiro, considerando uma área ou localidade que simplesmente suporta árvores em crescimento sem considerar diretamente as características de sítio (métodos diretos) e, segundo, considerando a capacidade da área ou localidade em sustentar o crescimento, em que características do sítio como clima, solo e vegetação são considerados (métodos indiretos) (TONINI et al., 2000).

Para Clutter et al. (1983), os métodos para predizer o crescimento de florestas podem ser classificados em indiretos e diretos. Os métodos indiretos avaliam a qualidade do sítio a partir de atributos do ambiente, levando-se em consideração características como clima, solo e vegetação. Os métodos diretos utilizam-se

basicamente de indicadores na própria vegetação, que refletem as interações de todos esses fatores de sítio.

Os métodos indiretos podem ser divididos em: quantitativos, também conhecido como análise fator-sítio ou solo, e qualitativos, que envolvem a divisão das terras em unidades com características, potencial de produção e de uso uniformes (SCHÖNAU, 1987).

De acordo com Carmean (1975), os métodos diretos avaliam a capacidade produtiva em termos do crescimento das árvores em altura, área basal, volume, biomassa. Os principais são: o método da curva-guia, o método das tendências correlacionadas e o método do índice de sítio preliminar (“a priori”), sendo o índice de sítio o indicador mais usado para expressar a capacidade produtiva, que é uma expressão da qualidade do sítio baseada na altura das árvores dominantes e codominantes em uma determinada idade-índice.

Na literatura, observa-se o método de índice de sítio (baseia-se na altura média das árvores dominantes na idade de referência) como o mais universalmente utilizado para definir a potencialidade dos sítios florestais e assim o será até que métodos que conjuguem a soma de fatores ambientais possam ser traduzidos de forma numérica e acessível aos usuários do setor. (CUNHA NETO et al, 1996).

Selle et al. (1994) relata que, devido às árvores apresentarem taxas de crescimento diferenciadas ao longo do tempo, variando de acordo com o ambiente e intervenções humanas, é possível construir os fundamentos matemáticos para estimar o crescimento das árvores em determinado sítio. Esta descrição do crescimento pode ser feita, em geral, com uma função apropriada, empregando técnicas de regressão múltipla ou não linear.

No entanto, quando se expressa o crescimento, por meio de funções matemáticas, é de vital importância analisar como é a forma de crescimento de árvores de locais diferentes, ou seja, se elas crescem segundo o mesmo padrão ou não. A ocorrência de diferenciação no crescimento da altura em diferentes sítios, para uma mesma espécie, é denominada polimorfismo entre as formas de crescimento, mostrando a necessidade de ajustar funções individuais para estes locais obtendo, com isso, maior precisão na descrição dos sítios. (SELLE et al., 1994).

A classificação das propriedades florestais quanto aos possíveis níveis de produtividade de grande importância, pois da maior ou menor produtividade do sítio

dependem: a dimensão dos produtos advindos da floresta nas várias idades; a viabilidade ou não de projetos florestais; as diferentes respostas de certas culturas florestais; a identificação do potencial de produtividade da propriedade florestal, além de fornecer referência para diagnóstico e prescrição do manejo e conservação do solo. (SCOLFORO, 1993)

Em síntese a estimativa do índice de sítio assim como o mapeamento dos sítios clarifica as possibilidades e riscos para o manejo de florestas, servindo como base para planejamento de curto e longo prazo, seja local ou regional. (SCOLFORO, 1992).

Quanto ao estudo dos sítios florestais pode-se dizer que, em nosso País, poucos são os estudos pertinentes a esta área e apenas algumas empresas possuem suas florestas mapeadas, em termos de capacidade de produção. A determinação da produtividade florestal é um recurso básico quando o objetivo é dispor e aplicar um sistema integrado de manejo visando à racionalização e otimização do uso dos recursos florestais (SELLE et al., 1994).

Na pesquisa de sítio, o maior problema em classificação é que a sua qualidade não é constante. Ela varia de acordo com a espécie florestal, genótipo, preparo, fertilização do solo e outros fatores que influenciam diretamente o crescimento como precipitação, compactação e erosão do solo, incidências de doenças, danos causados por insetos, fungos etc (SCHÖNAU, 1987).

Também há uma tendência de se discutir os diversos fatores do meio separadamente, o que afeta de forma particular e definida o crescimento da árvore. Os fatores individuais do meio não são forças isoladas que operam sobre as plantas, mas sim influências interdependentes, que devem ser consideradas em última análise (SPURR e BARNES, 1980).

Johnston et al. (1977), afirmaram que o principal problema na classificação de sítios está em encontrar fatores que sejam bons índices de crescimento e que ainda sejam ideais para mais de uma espécie.

Para Gerding e Schlatter (1995), nos estudos de classificação de sítios, deve-se analisar a dependência da produtividade em função de um conjunto de variáveis de sítio de distinta natureza para destacar as mais significativas (ou relevantes) e, para uma melhor interpretação dos resultados é conveniente agrupar tais variáveis em conjuntos que caracterizem determinadas condições de sítio, sendo esses grupos de variáveis denominados de fatores de sítio.

Portanto, na avaliação da qualidade do sítio, os enfoques mais holísticos envolvendo múltiplos fatores simultaneamente como clima, solo, relevo, vegetação, geologia e outros, podem fornecer melhores resultados, do que os enfoques reducionistas. Isto justifica o emprego de técnicas de análise multivariada.

2.4 USO DA ESTATÍSTICA MULTIVARIADA EM CLASSIFICAÇÃO DE SÍTIOS

O uso de técnicas de estatística multivariada, gerando uma classificação de sítios, pode ser adequado, pois avalia simultaneamente um grande número de variáveis, identificando aquelas com real poder de discriminação dos sítios e permitindo a compreensão dos relacionamentos entre as variáveis e as classes de qualidade de sítio (WILLIAMS, 1983; HARDING et al., 1985; LAAR, 1987; BURTON et al., 1991; ROWE, 1992; FINCHER e SMITH, 1994).

Segundo Burton et al. (1991), os resultados dos métodos multivariados são comparáveis aos métodos mais tradicionais de classificação de sítios, com a vantagem de admitirem múltiplas comparações de vários locais, usando uma extensa gama de dados. Eles afirmaram ainda que a análise estatística multivariada tenha sido usada na pesquisa de sítio com três objetivos básicos (BURTON et al., 1991):

- Ordenamento, com o propósito de determinar a influência de vários fatores do meio na composição e na produtividade florestal;
- Agrupamento com a finalidade de classificação, utilizados em prescrições de manejos;
- Estabelecimento de um modelo com espaço dimensional mais reduzido, parcimonioso, mas que efetivamente descreva as diferenças observadas entre os grupos.

Entre as técnicas multivariadas mais utilizadas em classificação de sítio na literatura são: análise de componentes principais, fatorial, agrupamento e discriminante.

McBoyle (1973), mediante a análise de componentes principais, usou 20 variáveis climáticas que forneceram três fatores que explicaram, juntas, 86,7% da variância total, sendo esta classificação considerada útil em termos descritivos e na avaliação de potencialidades, e surgida com o desenvolvimento de softwares apropriados, em classificações na Austrália.

Omi et al. (1979) promoveram uma classificação em áreas homogêneas, tendo como objetivo o uso e manejo planejado da terra, recorrendo a técnicas multivariadas de análise fatorial, agrupamento e discriminante.

Pregitzer e Barnes (1982) usaram a vegetação rasteira como indicadora de condições edáficas (drenagem, textura e nível de fertilidade) de ecossistemas, empregando análise de componentes principais e discriminante em Michigan, no Canadá.

Newnham (1992) apresenta a análise de agrupamento como alternativa usada para agregar povoamentos, quanto ao estoque de madeira, definindo áreas geográficas homogêneas em termos de composição de espécies, volume/ha, idade etc.

Nos últimos anos com a crescente demanda de pacotes estatísticos computacionais mais refinados, tem aumentado de um modo geral, o número de trabalhos publicados no Brasil usando as técnicas multivariadas. Ferreira e Souza (1997) descrevem as potencialidades do uso das técnicas multivariadas para fins de manejo florestal, utilizando a análise fatorial, componentes principais, discriminante, agrupamento, entre outros.

Laven (1982) empregou a análise de agrupamento, análise de componentes principais e análise discriminante para estabelecimento de homogeneidade em estudos de sucessão florestal, num estudo que envolveu dados relacionados com a vegetação, ambiente, bem como a combinações destes no Estado do Paraná.

Queiroz (1984) estudou a operacionalidade da análise fatorial apresentando os fundamentos teóricos do método da máxima verossimilhança, com a finalidade de caracterizar em detalhes a estrutura de um povoamento florestal tropical de mais de 5 mil hectares, localizado na Floresta Nacional do Tapajós, na Região Amazônica, o que tornou o método viável computacionalmente, obtendo resultados satisfatórios.

Aspiazú et al. (1990) utilizaram a análise fatorial do tipo R, com rotação varimax e a análise de agrupamento usando a distância euclidiana como medida de similaridade na delimitação de 10 regiões climáticas homogêneas numa área-teste do Estado de Minas Gerais com dados de temperatura, umidade relativa e precipitação.

Martins (1992) ajustou um modelo fatorial com 17 variáveis, reduzindo para três fatores ou índices, dados referentes ao clima e a altitude. O trabalho visava subsidiar estudos de introdução e manejo de espécies e procedências florestais. O

nível de dissimilaridade foi obtido pela distância euclidiana média e uma verificação final dos grupos (regiões) por meio da análise discriminante, obtendo oito regiões ecológicas no Estado do Espírito Santo.

Braga (1997) utilizou a análise discriminante usando o Método Stepwise e a Distância Generalizada de Mahalanobis para identificar características ambientais que determinam e separam três classes de qualidade de sítio, visando áreas para plantio de florestas comerciais com eucaliptos (*Eucalyptus grandis*), em Minas Gerais.

Scheeren e Finger (1998) estudaram o efeito dos nutrientes do solo e da copa no crescimento de um povoamento plantado de *Araucaria angustifolia* (Bert.) O. Ktze. na Floresta Nacional de Canela, no Estado do Rio Grande do Sul. Por meio da Análise Fatorial, pelo método das Componentes Principais, reduziu-se a dimensão da matriz de 54 variáveis (altura dominante, teores de nutrientes do solo e das acículas) para nove fatores comuns, não observáveis, que explicam 97% da variância total. Observou-se a existência de variação do efeito das variáveis químicas no crescimento em altura em diferentes locais, para a espécie estudada.

Andrade et al. (1999, 2000) promoveram uma classificação ecológica no Estado da Paraíba, definindo 10 sub-regiões a partir de variáveis climáticas e biopedológicas, no uso de técnicas multivariadas de análise fatorial, de agrupamento e discriminante.

Gerhardt *et al.* (2001) estudaram a influência dos fatores físicos e morfológicos do solo e da qualidade de nutrientes da serapilheira, na classificação de sítios em um povoamento de *Araucaria angustifolia* (Bert.) O. Ktze no Rio Grande do Sul. Para isto usou a análise de agrupamento, classificando em três grupos distintos e em seguida uma análise discriminante comprovando a existência dos grupos previamente obtidos, ao nível de significância de 0,01.

2.5 A ANÁLISE FATORIAL

2.5.1 Definição do Modelo Análise Fatorial

Reis (1997) define a análise fatorial como "um conjunto de técnicas estatísticas cujo objetivo é representar ou descrever um número de variáveis iniciais a partir de um menor número de variáveis hipotéticas".

Trata-se de uma técnica estatística multivariada que, a partir da estrutura de dependência existente 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 como função das variáveis originais. Além disso, é possível saber o quanto cada fator está associado a cada variável e o quanto o conjunto de fatores explica da variabilidade geral dos dados originais.

2.5.2 O Modelo Matemático

Sendo o vetor aleatório observável X , com p componentes, media μ e covariância Σ , o Modelo Fatorial postula que X é linearmente dependente de algumas variáveis aleatórias não observáveis F_1, F_2, \dots, F_m , chamadas fatores comuns ($m \leq p$), e p fontes de variação $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p$, chamadas erros ou fatores específicos.

De acordo com Cooley e Lohnes (1971), em particular, o Modelo de Análise Fatorial é dado por:

$$X_{(p \times 1)} = \mu_{(p \times 1)} + L_{(p \times m)} \cdot F_{(m \times 1)} + \varepsilon_{(p \times 1)} \quad (1)$$

ou ainda,

$$\begin{aligned} (X_1 - \mu_1) &= \ell_{11}F_1 + \ell_{12}F_2 + \dots + \ell_{1m}F_m + \varepsilon_1 \\ (X_2 - \mu_2) &= \ell_{21}F_1 + \ell_{22}F_2 + \dots + \ell_{2m}F_m + \varepsilon_2 \\ &\vdots \\ (X_p - \mu_p) &= \ell_{p1}F_1 + \ell_{p2}F_2 + \dots + \ell_{pm}F_m + \varepsilon_p \end{aligned} \quad (2)$$

em que:

$(X - \mu)$ é um vetor de desvios ($p \times 1$) de elementos $(X_i - \mu_i)$, $i = 1, \dots, p$.

F_j é o j -ésimo fator comum ($j = 1, \dots, m$)

ℓ_{ij} o carregamento da i -ésima variável no j -ésimo fator comum, também chamada de carga fatorial (em inglês, loadings).

ε_i é o i -ésimo fator específico (ou a parte aleatória), que não é comum às restantes variáveis.

Sob o modelo (1) é necessário fazer as seguintes suposições sobre os vetores aleatórios F e ε :

- F e ε são independentes;
- $E(F) = 0$ e $Cov(F) = I_m$, em que I_m é a matriz identidade de ordem m ;
- $E(\varepsilon) = 0$ e $Cov(\varepsilon) = \Psi$, sendo Ψ é uma matriz diagonal com variância específica ψ_i ($i = 1, \dots, p$) na diagonal principal.

Os fatores comuns são independentes uns dos outros e tem valor médio nulo e variância um; os fatores específicos são independentes entre si e dos fatores comuns, tendo média zero e variância ψ_i específica de cada variável X_i .

Aplicando as suposições acima, teremos os seguintes resultados:

$$i) Cov(F, \varepsilon) = 0.$$

$$ii) Cov(X_i, X_k) = \sum_{j=1}^m \ell_{ij} \ell_{kj} = \ell_{i1} \ell_{k1} + \ell_{i2} \ell_{k2} + \dots + \ell_{im} \ell_{km}.$$

$$iii) Cov(X_i, F_j) = \ell_{ij}, \text{ logo } Corr(X_i, F_j) = \frac{Cov(X_i, F_j)}{\sqrt{Var(X_i) \cdot Var(F_j)}} = \ell_{ij}.$$

As cargas fatoriais ℓ_{ij} são as covariâncias medidas entre as variáveis observadas e os fatores comuns. No caso do fator comum é expresso pela correlação entre o fator e cada variável observada, sendo padronizado com média zero e variância um.

Uma vez que a análise fatorial tem como objetivo a explicação das covariâncias em termos das variáveis iniciais, convém expressar essas covariâncias em termos das cargas fatoriais e das variâncias específicas, ou seja:

$$\Sigma = Cov(X) = Cov(L \cdot F + \varepsilon) = L \cdot Cov(F) \cdot L' + Cov(\varepsilon) = LL' + \Psi,$$

que é uma alternativa de expressar o modelo de análise fatorial, sendo usado na estimativa de seus parâmetros. Um resultado importante é que, se Ψ satisfaz a

relação $\Sigma = LL' + \Psi$, então há infinitas matrizes que também satisfazem esta relação. Isto pode ser verificado, considerando uma matriz ortogonal T qualquer ($TT' = T'T = I$) e seja uma transformação $L^* = LT$, logo:

$$L^*L^{*'} + \Psi = (LT)(LT)' + \Psi = LTT'L' + \Psi = LL' + \Psi = \Sigma$$

Sobre o modelo de análise fatorial ortogonal, o número de cargas fatoriais estimados é $(p \times m)$, o número de variâncias específicas é p , que é o número de variáveis; ou seja, o total de parâmetros estimados do modelo é $p(m + 1)$, sendo m o número de fatores comuns.

Quando m for bem menor que p ($m < p$), o modelo pode ser usado, desde que exista a matriz fatorada Σ , tal que $\Sigma = LL' + \Psi$.

Quando $m = p$, a matriz pode ser reproduzida por $\Sigma = LL'$, onde Ψ é uma matriz zero, usada na extração dos fatores por componentes principais.

Quando m for muito menor que p , a matriz Σ não poderá ser escrita como $\Sigma = LL' + \Psi$; neste caso, as estimativas obtidas poderão ser inconsistentes para a variância dos fatores comuns e dos fatores específicos.

2.5.3 As Comunalidades e Especificidades

Podemos reescrever o modelo de análise fatorial ortogonal em (2), com X_i padronizados de média zero e variância um dado por:

$$X_i = \sum_{j=1}^m \ell_{ij} F_j + \varepsilon_i, \quad i = 1, \dots, p \quad (3)$$

Obtemos a variância do modelo, que é dada por:

$$Var(X_i) = Var\left[\sum_{j=1}^m \ell_{ij} F_j + \varepsilon_i\right] = \sum_{j=1}^m \ell_{ij}^2 + Var(\varepsilon_i) = \sum_{j=1}^m \ell_{ij}^2 + \psi_i \quad (4)$$

o qual teremos: $Var(X_i) = \sigma_i^2 = h_i^2 + \psi_i$, $i = 1, \dots, p$.

Isto é, a matriz de covariância é composta em 2 partes: a porção da variância da variável aleatória X_i advinda como contribuição dos m fatores comuns, a *comunalidade* i , denotada por h_i^2 e a porção da variância desta variável atribuída ao fator específico ε_i , a *especificidade* ou *variância específica* i , denotada por ψ_i . Significa que as variáveis mais explicadas são as que apresentam valores altos para as cargas fatoriais, e ambos variam de -1 a 1.

Como as variáveis são padronizadas, a expressão (4) resulta em $\psi_i = 1 - \sum_{j=1}^m \ell_{ij}^2$, ou ainda na forma matricial (ANDERSON, 1984):

$$\Psi = 1 - LL' = \text{diag}(\psi_1, \dots, \psi_p), \text{ sendo } L = \{\ell_{ij}\}_{p \times m} = \begin{bmatrix} \ell_{11} & \ell_{12} & \dots & \ell_{1m} \\ \ell_{21} & \ell_{22} & \dots & \ell_{2m} \\ \vdots & & \ddots & \vdots \\ \ell_{p1} & \ell_{p2} & \dots & \ell_{pm} \end{bmatrix}_{p \times m}, \text{ a matriz}$$

das cargas fatoriais.

2.5.4 Métodos de Estimação dos Fatores

O Modelo Fatorial procura representar adequadamente os dados com um pequeno número de fatores não-observáveis. A matriz de covariância S é um estimador da matriz populacional Σ desconhecida, assim como a matriz de correlação R é um estimador de ρ . Se Σ desvia significativamente de uma matriz diagonal, então o modelo fatorial pode ser usado e o problema inicial é a estimação das cargas fatoriais ℓ_{ij} e variâncias específicas ψ_i , conforme Johnson e Wichern (1998).

O método de estimação dos fatores por componentes principais usando a matriz de correlação R é apresentado neste trabalho, devido a sua rápida convergência e de simples operacionalização (FERREIRA, 1996). Meyer e Braga (1999) citam que, quando o problema for a multicolinearidade entre variáveis, ou quando o objetivo for a classificação por intermédio da criação de um índice, este método é o mais indicado.

Outro método bastante conhecido para a obtenção dos fatores, citado anteriormente, é o da máxima verossimilhança, em que, em sua versão usual, supõe que as variáveis envolvidas tenham distribuição normal. Esse método, em geral, não é indicado para os casos em que a suposição de normalidade das variáveis envolvidas não esteja satisfeita.

2.5.5 Extração dos Fatores por Componentes Principais

A rigor, a análise fatorial (AF) e a análise de componentes principais (ACP) são técnicas multivariadas distintas e com objetivos diferentes. A ACP, é uma técnica de transformação de variáveis, no qual a partir de um conjunto de p variáveis ou medidas de interesse, cria-se um novo conjunto p de variáveis (p componentes principais), de tal modo que a primeira componente seja aquela com maior variância possível dentre todas as possíveis combinações lineares das variáveis originais, e os demais componentes são ordenados em termos de sua variabilidade.

A análise fatorial apenas utiliza o método numérico das componentes principais para a obtenção dos fatores, tomando-se $m = p$.

O método de extração de fatores iniciais através do modelo geral da análise por componentes principais é dado por:

$$Y_i = \sum_{j=1}^p a_{ij} X_j, i = 1, \dots, p \quad (5)$$

em que X_1, X_2, \dots, X_p são as variáveis medidas e correlacionadas entre si, Y_1, Y_2, \dots, Y_n são variáveis não-correlacionadas que designam as componentes principais, a_{ij} são os coeficientes ponderados das variáveis com cada componente, de forma a explicar a máxima variabilidade existente nos dados.

Usando uma transformação ortogonal em (5), escrevemos as variáveis X_i como função das componentes normalizadas, conforme Manly (1997), ou seja:

$$\begin{cases} X_1 = a_{11}Z_1 + a_{21}Z_2 + \dots + a_{p1}Z_p \\ X_2 = a_{12}Z_1 + a_{22}Z_2 + \dots + a_{p2}Z_p \\ \vdots \\ X_p = a_{1p}Z_1 + a_{2p}Z_2 + \dots + a_{pp}Z_p \end{cases} \quad (6)$$

sendo Z_i o i -ésimo valor padronizado da variável X_i .

Como no modelo fatorial, apenas m das componentes principais são retidos, o modelo (6) fica:

$$\begin{cases} X_1 = a_{11}Z_1 + a_{21}Z_2 + \dots + a_{m1}Z_p + \varepsilon_1 \\ X_2 = a_{12}Z_1 + a_{22}Z_2 + \dots + a_{m2}Z_p + \varepsilon_2 \\ \vdots \\ X_p = a_{1p}Z_1 + a_{2p}Z_2 + \dots + a_{mp}Z_p + \varepsilon_p \end{cases} \quad (7)$$

no qual ε_i é a combinação linear das componentes principais, tomados de Z_{m+1} a Z_p .

Para obtermos a escala das componentes principais com variância unitária em (7), Z_1, Z_2, \dots, Z_m deve ser dividido pelo desvio padrão, $\sqrt{\lambda_i}$, a raiz quadrada do autovalor correspondente a matriz R. Neste caso:

$$\begin{cases} X_1 = \sqrt{\lambda_1} a_{11} Z_1 + \sqrt{\lambda_2} a_{21} Z_2 + \dots + \sqrt{\lambda_m} a_{m1} Z_p + \varepsilon_1 \\ X_2 = \sqrt{\lambda_1} a_{12} Z_1 + \sqrt{\lambda_2} a_{22} Z_2 + \dots + \sqrt{\lambda_m} a_{m2} Z_p + \varepsilon_2 \\ \vdots \\ X_p = \sqrt{\lambda_1} a_{1p} Z_1 + \sqrt{\lambda_2} a_{2p} Z_2 + \dots + \sqrt{\lambda_m} a_{mp} Z_p + \varepsilon_p \end{cases} \quad (8)$$

Sendo as cargas fatoriais dadas por $\ell_{ij} = \sqrt{\lambda_i} a_{ji}$, e $F_i = \frac{Z_i}{\sqrt{\lambda_i}}$, o modelo

fatorial (8) pode ser escrito como sistema de equações, dado por:

$$\begin{cases} X_1 = \ell_{11} F_1 + \ell_{12} F_2 + \dots + \ell_{1m} F_m + \varepsilon_1 \\ X_2 = \ell_{21} F_1 + \ell_{22} F_2 + \dots + \ell_{2m} F_m + \varepsilon_2 \\ \vdots \\ X_p = \ell_{p1} F_1 + \ell_{p2} F_2 + \dots + \ell_{pm} F_m + \varepsilon_p \end{cases} \quad (9)$$

ou simplesmente, $X_i = \sum_{j=1}^m \ell_{ij} F_j + \varepsilon_i$, $i = 1, \dots, p$, como descrito anteriormente.

2.5.6 A Escolha do Número de Fatores

A escolha do número de fatores é uma das tarefas mais importantes de uma Análise Fatorial. Hair et al. (1995) discute que, se o pesquisador opta por um número muito reduzido, ele pode não identificar estruturas importantes existentes nos dados e, por outro lado, se o número é excessivo, ele pode vir a ter problemas de interpretabilidade dos fatores. Barroso (2003) cita alguns critérios como ponto de partida para uma solução final:

Critério de Kaiser – Esse critério, desenvolvido por Kaiser (1958), também conhecido como critério da raiz latente, determina que o número de fatores m deve ser igual ao número de autovalores maiores ou iguais à média das variâncias das variáveis analisadas. Na situação em que a AF é feito sobre a matriz de correlação,

com variáveis padronizadas, esse critério corresponde à exclusão de fatores com autovalores inferiores a um, ou seja:

$$\frac{\sum_{i=1}^p \lambda_i}{p} = \frac{\sum_{i=1}^p \lambda_i}{\text{traço}(R)} = 1$$

Este critério foi proposto pela primeira vez por Guttman, em 1954. Kaiser apresentou boas razões para se confirmar à confiabilidade do critério, mostrando que funciona muito bem na prática.

Critério da porcentagem da variância explicada – Os fatores são adotados até atingir o nível pré-determinado do percentual da variância explicada. Por exemplo, estipulando um nível de explicação de 70% da variabilidade total dos dados, por esse critério, os fatores que não acumulem um grau de explicação igual ao valor preestabelecido, serão descartados.

Critério "scree test" – O número ótimo de fatores é obtido quando a variação da explicação entre fatores consecutivos passa a ser pequena. Isso porque o primeiro autovalor difere do segundo, que difere do terceiro, e assim por diante. O gráfico scree plot é um gráfico representando os autovalores na ordenada e na abscissa o número da respectiva componente (os fatores), como na Figura 1.

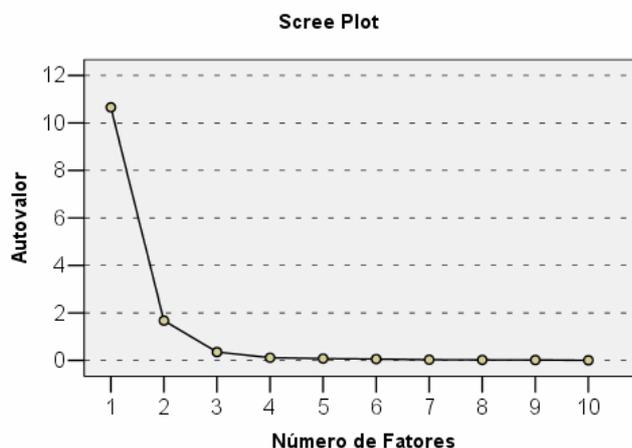


Figura 1: Gráfico “Scree Plot” dos Autovalores do Modelo em Função do Número de Fatores

Métodos inferenciais – Outros métodos foram desenvolvidos para os casos em que as variáveis originais seguem uma distribuição normal. Esses métodos consistem no desenvolvimento de testes estatísticos que se alicerçam na suposição de normalidade.

Dentre esses testes destacamos o de Bartlett (JOHNSON e WICHERN, 1998) que verifica a adequabilidade do modelo de Análise Fatorial estimado pelo método da máxima verossimilhança para representar a estrutura de dependência dos dados.

2.5.7 A Rotação dos Fatores

Os carregamentos fatoriais estão sujeitos a uma rotação de fatores sem alteração das comunalidades e variâncias específicas. O objetivo principal da rotação é uma melhor interpretação e, conseqüentemente, uma visão mais clara dos agrupamentos das variáveis nos fatores. Basicamente, os carregamentos serão multiplicados por uma matriz ortogonal T sem prejudicar a reprodução da matriz de covariância em $\Sigma = LL' + \Psi$ (FURTADO et al., 2003).

Ou seja, para que os m fatores ortogonais situados no espaço p-dimensional sejam mais facilmente entendidos, é necessário que estejam em tal posição de modo que as projeções de cada variável sobre o eixo fatorial situem-se melhor possível, seja junto à extremidade, seja junto à origem. Em outras palavras, haverá necessidade de um critério de maximização da variância dos carregamentos sobre os fatores, o que conseguido pela rotação dos eixos fatoriais.

Conforme Manly (1997), a rotação dos fatores pode ser ortogonal ou oblíqua. Na rotação ortogonal, os novos fatores são não-correlacionados entre si e os eixos dos fatores são perpendiculares, rotados por um ângulo de rotação ϕ .

Na rotação oblíqua, os eixos são não perpendiculares e os novos fatores são correlacionados, ou seja, $Cov(F^*) = Q'IQ = Q'Q \neq I$, sendo Q é uma matriz oblíqua e $F^* = Q'F$. As principais rotações mais usadas são (SPSS, 1999):

Varimax: é um método de rotação ortogonal mais usado que minimiza o número de variáveis que cada agrupamento terá, simplificando a interpretação dos fatores.

Quartimax (ortogonal): é um método que minimiza o número de fatores necessários para explicar cada variável e simplifica a interpretação das variáveis obtidas.

Equamax (ortogonal): é também um método que busca uma combinação dos outros (varimax e quartimax). O número de variáveis obtido terá carga fatorial maior e o número de fatores será minimizado.

Direct Oblimin: este método, diferentemente dos três anteriores, é oblíquo (não ortogonal). Quando certo valor δ (delta) assume o valor zero (*default*), a solução torna-se mais oblíqua. Tomando-se delta mais negativo, os fatores ficam menos oblíquos. Ignorando-se o *default* delta de zero, deve-se usar um número menor ou igual a 0,8.

Promax: é também um método oblíquo de rotação, o qual possibilita os fatores correlatos, podendo ser calculado mais rapidamente que a rotação *direct oblmin*. Assim ele é usado para grandes grupos de dados, tomando-se certo valor κ (kappa), de valor quatro na maioria dos casos.

Seja qual for a rotação a ser usada, uma carga fatorial próximo de zero significa que X_i não é fortemente correlacionada ao fator F_j , enquanto que um valor ℓ_{ij} perto de um (positivo ou negativo) revela que a variável X_i tem forte correlação com o fator j (MANLY, 1997).

O método de rotação freqüentemente usado na literatura é a rotação ortogonal varimax. Baseia-se na suposição de que a interpretabilidade do fator j possa ser medido pela variância das cargas fatoriais ao quadrado, ou seja, a variância de $\ell_{1j}^2, \ell_{2j}^2, \dots, \ell_{pj}^2$. Se esta variância for alta, os valores de ℓ_{ij}^2 tendem a serem próximas de zero ou próximas de um (MANLY, 1997).

Kaiser (1958) definiu o critério de rotação varimax escolhendo uma transformação ortogonal T, que dá um valor V tão grande quanto possível, ou seja:

$$V = \frac{1}{p} \sum_{j=1}^m \left\{ \sum_{i=1}^p \ell_{ij}^{*4} - \frac{1}{p} \left[\sum_{i=1}^p \ell_{ij}^{*2} \right]^2 \right\} \quad (10)$$

sendo ℓ_{ij}^* os coeficientes rotacionados dados por $\ell_{ij}^* = \frac{\ell_{ij}}{h_i}$ com

$$h_i = (1 - \psi_i)^{1/2} = \sqrt{\sum_{j=1}^m \ell_{ij}^2}, \quad i = 1, \dots, p.$$

Após a transformação ortogonal $L^* = LT$, as cargas fatoriais ℓ_{ij}^* são multiplicadas por \hat{h}_i de modo que as comunalidades fiquem inalteradas.

Kaiser modificou mais tarde este procedimento pela normalização das cargas fatoriais antes de maximizar as variâncias de seus quadrados, apresentando

melhores resultados. Todavia, a rotação pode ser usada com ou sem o critério de normalização de Kaiser (MANLY, 1997).

O processo computacional para obter o ângulo de rotação ϕ , na rotação varimax, com cargas normalizadas ℓ_{ir}^* e ℓ_{is}^* nos pares de fatores r e s ($s > r$), e p variáveis do modelo, deve satisfazer a relação (FERREIRA, 1996):

$$\operatorname{tg}(\phi) = \frac{2 \left[2p \sum_{i=1}^p (\ell_{is}^2 - \ell_{ir}^2) \ell_{ir} \ell_{is} - \sum_{i=1}^p (\ell_{is}^2 - \ell_{ir}^2) \cdot 2 \sum_{i=1}^p \ell_{ir} \ell_{is} \right]}{p \sum_{i=1}^p [(\ell_{is}^2 - \ell_{ir}^2)^2 - (2\ell_{ir} \ell_{is})^2] - \left\{ \left[\sum_{i=1}^p (\ell_{is}^2 - \ell_{ir}^2) \right]^2 \left(\sum_{i=1}^p 2\ell_{ir} \ell_{is} \right)^2 \right\}} \quad (11)$$

o que é equivalente à:

$$\operatorname{tg}(4\phi) = \frac{D - \frac{2}{p} AB}{C - \frac{1}{p} (A^2 - B^2)} \quad (12)$$

$$A = \sum_{i=1}^p u_i, \text{ sendo } u_i = \ell_{ir}^2 - \ell_{is}^2,$$

$$B = \sum_{i=1}^p v_i, \text{ sendo } v_i = 2 \cdot \ell_{ir} \ell_{is},$$

$$C = \sum_{i=1}^p (u_i^2 - v_i^2) \text{ e } D = 2 \sum_{i=1}^p u_i v_i,$$

Conforme Harman (1976), a escolha do quadrante é designada pelos sinais do numerador e denominador de (12), o qual especifica na Tabela 1 o quadrante de 4ϕ em função dos sinais.

Tabela 1: Quadrante do ângulo 4ϕ em função dos sinais do numerador e denominador de (12)

Sinal do Denominador	Sinal do Numerador	
	+ (positivo)	- (negativo)
+ (positivo)	$0 \leq 4\phi < 90^0$	$-90^0 \leq 4\phi < 180^0$
- (negativo)	$90^0 \leq 4\phi < 180^0$	$-180^0 \leq 4\phi < -90^0$

Como solução iterativa tem-se os seguintes procedimentos: obtendo o ângulo de rotação ϕ gera o primeiro e segundo fatores rotacionados, sendo $L^* = \{\ell_{ij}^*\}_{p \times m}$ no qual $L^* = LT$, onde T é a matriz ortogonal de ordem (m x m). Em seguida, o novo

primeiro fator é rotado com o terceiro fator original e assim por diante, até que os $m(m-1)/2$ pares de rotações tenham sido executados.

A seqüência de rotações é repetida até que todos os ângulos sejam menores que o critério de convergência especificado dentro de um ciclo. Harman (1976) mostra que a iteração termina quando $|V_{(i)} - V_{(i-1)}| \leq 10^{-5}$, no qual o índice (i) indica a seqüência da iteração.

2.4.8 Testes de Adequabilidade do Modelo

Esses testes indicam qual é o nível de confiança que se pode esperar dos dados quando seu tratamento se pelo método multivariado de análise fatorial seja empregado com sucesso (HAIR et al, 1988).

No presente trabalho são apresentados dois testes, visto que foi utilizado o método das componentes principais: o Teste KMO e o Teste de Esfericidade de Bartlett.

2.4.8.1 Teste KMO

O Teste Kaiser-Meyer-Olkin (KMO) é uma estatística que indica a proporção da variância dos dados que pode ser considerada comum a todas as variáveis, ou seja, que pode ser atribuída a um fator comum, então: quanto mais próximo de um mais adequado é a amostra à aplicação da análise fatorial.

$$KMO = \frac{\sum_{i \neq j} r_{ij}^2}{\sum_{i \neq j} r_{ij}^2 + \sum_{i \neq j} v_{ij}^2}, \text{ sendo } 0 \leq KMO \leq 1 \quad (13)$$

Se $KMO \rightarrow 0$, o método da Análise Fatorial é inadequado (que as correlações simples entre as variáveis são pequenas). Se $KMO \rightarrow 1$, o método de Análise Fatorial é perfeitamente adequado para o tratamento dos dados (as correlações parciais entre as variáveis são pequenas).

Os valores r_{ij} e v_{ij} são os (i, j)-ésimos elementos das matrizes R e V, respectivamente, sendo a matriz V dada por:

$$V = UR^{-1}U, \text{ em que: } U = \left[(\text{diag}R^{-1})^{\frac{1}{2}} \right]^{-1}$$

Kaiser (1958) também sugere outra maneira de classificar o grau de ajuste do modelo em relação ao valor do KMO e concluir se o modelo se ajusta bem aos dados ou não (Tabela 2).

Tabela 2: Classificação do ajuste do modelo de análise fatorial em relação ao valor da estatística de Kaiser-Meyer-Olkin

KMO	Grau de Ajuste ao Modelo
0,9 -1,0	Muito Bom
0,8 - 0,9	Bom
0,7 - 0,8	Média
0,6 - 0,7	Razoável
0,5 - 0,6	Ruim
< 0,5	Inaceitável

2.4.9.2 O Teste de Esfericidade de Bartlett (χ^2)

O Teste de Esfericidade de Bartlett testa se a matriz de correlação é uma matriz identidade ($H_0 : R = I$), o que indicaria que não há correlação entre os dados. Dessa forma, procura-se para um determinado nível de significância, rejeitar a hipótese nula de matriz de correlação identidade. Também usada no caso de testar se Σ é uma matriz diagonal ($H_0 : \Sigma = \sigma^2 I$), com $\Sigma = \text{diag}(\sigma_{11}, \dots, \sigma_{pp})$, $\sigma_{ii} > 0$, $i = 1, \dots, p$. A estatística do teste é dada por (JONHSON e WICHERN, 1998):

$$\Lambda = \frac{|S|^{n/2}}{\prod_{i=1}^p s_{ii}^{n/2}} = |R|^{n/2} < c \quad (14)$$

Como $|R| = \prod_{i=1}^p \lambda_i$, o teste de razão de verossimilhança $-2 \ln \Lambda$, descrito em

(14) pode ser escrito na seguinte forma:

$$-2 \ln \Lambda = -2 \ln |R|^{n/2} = -n \ln \sum_{i=1}^p \lambda_i \quad (15)$$

Em grandes amostras, o teste da razão de verossimilhança, com base na distribuição normal, a estatística $-2 \ln \Lambda$ tende a se aproximar de distribuição Qui-quadrado (χ^2) com $p(p-1)/2$ graus de liberdade.

Bartlett (1954) sugere uma correção da aproximação dessa estatística χ^2 , acrescentando um fator de correção $[1 - (2p + 1)/6n]$ no resultado (15), obtendo:

$$-2 \ln A = -\left[n - \left(\frac{1}{6}\right)(2p + 1)\right] \ln \sum_{i=1}^p \lambda_i \quad (16)$$

Para a hipótese $H_0 : R = I$, rejeita-se H_0 se a estatística acima for menor do que um valor pré-fixado c , onde o valor crítico c para grandes amostras, a um nível α é $c = \chi_{p(p-1)/2}^2(\alpha)$.

2.4.9 Os Escores Fatoriais

Os escores fatoriais são os valores estimados dos fatores, visto que os fatores e os erros são variáveis não observáveis, muito embora seus valores possam ser estimados. Neste trabalho dois métodos de estimação serão propostos: o dos mínimos quadrados ponderados e o da regressão. Ambos tratam as cargas fatoriais e as variâncias específicas estimadas, como se fossem os verdadeiros valores desconhecidos.

Havendo rotação, os escores são obtidos a partir das cargas fatoriais que sofreram a rotação e não a partir das originais. Não obstante, as fórmulas não distinguirão entre as situações em que ocorreu rotação daquelas em que não ocorreu, uma vez que estas fórmulas não são alteradas pelas rotações.

2.4.9.1 Método dos Mínimos Quadrados Ponderados

Suponha que μ , L e Ψ sejam considerados inicialmente como conhecidos para o modelo fatorial $X - \mu = LF + \varepsilon$. Sabendo que $Var(\varepsilon_i) = \psi_i$, não é necessariamente igual para todo i , Bartlett (1937) sugeriu o uso dos quadrados mínimos ponderados, usando como peso o recíproco das variâncias específicas. A soma de quadrados dos resíduos do modelo fatorial ponderado é dada por:

$$\sum_{i=1}^p \frac{\varepsilon_i^2}{\psi_i} = \varepsilon \Psi^{-1} \varepsilon = (X - \mu - LF)' \Psi^{-1} (X - \mu - LF) \quad (17)$$

A solução \hat{F} que minimiza (17) proposta por Bartlett (1937) é:

$$\hat{F} = (L'\Psi^{-1}L)^{-1}L'\Psi^{-1}(X - \mu) \quad (18)$$

Como, de fato, L, Ψ e μ são desconhecidos, os respectivos estimadores devem ser utilizados para a obtenção dos escores fatoriais, ou seja:

$$\hat{F}_j = (\hat{L}'\hat{\Psi}^{-1}\hat{L})^{-1}\hat{L}'\hat{\Psi}^{-1}(X_j - \bar{X}) , \quad j = 1, 2, \dots, n \quad (19)$$

sendo $\bar{X} = \hat{\mu}$, a estimativa de máxima verossimilhança. A condição única para esse escore é que a matriz $(\hat{L}'\hat{\Psi}^{-1}\hat{L})^{-1}$ seja uma matriz diagonal. Se a matriz de correlação for utilizada, temos:

$$\hat{F}_j = (\hat{L}'_Z \hat{\Psi}_Z^{-1} \hat{L}_Z)^{-1} \hat{L}'_Z \hat{\Psi}_Z^{-1} Z_j , \quad j = 1, 2, \dots, n \quad (20)$$

em que:

$$Z_j = D^{-1/2}(X_j - \bar{X})$$

$$D^{-1/2} = \text{diag}(s_{11}^{-1/2}, \dots, s_{pp}^{-1/2}) = \text{diag}(s_{ii}^{-1/2}), \quad i = 1, \dots, p$$

$\hat{L}_Z = \hat{V}^{-1/2}\hat{L}$, o qual $\hat{V}^{-1/2}$ é uma matriz diagonal, com o i -ésimo elemento de $\sigma_{ii}^{-1/2}$ estimada.

Se as cargas fatoriais sofrerem rotação, então $\hat{L}^* = \hat{L}T$, logo os escores serão obtidos por:

$$\hat{F}_j^* = T\hat{F}_j \quad (21)$$

2.4.9.2 Método de Regressão

Considerando que L e Ψ são conhecidas a partir do modelo de fatores originais $X - \mu = LF + \varepsilon$, as variáveis aleatórias F e ε possuem distribuição normal multivariada com vetor de médias 0 e variâncias dadas por Σ , segue-se que a combinação linear do modelo (1) tem distribuição $N_p(0, LL' + \Psi)$. De acordo com Ferreira (1996), a distribuição conjunta de $X - \mu$ e F é, também, normal com parâmetros $N_{m+p}(0, \Sigma^*)$, em que:

$$\Sigma^* = \begin{bmatrix} LL' + \Psi & L \\ L' & I \end{bmatrix} \quad (22)$$

A média 0 é um vetor $[(m + p) \times 1]$ de zeros. A distribuição condicional de $F | x$ é normal com média e variância dada por:

$$E(F | x) = L'\Sigma^{-1}(x - \mu) = L'(LL' + \Psi)^{-1} \quad (23)$$

e

$$\text{Cov}(F | x) = I - L'\Sigma^{-1}L = I - L'(LL' + \Psi)^{-1}L \quad (24)$$

Os coeficientes $L'(LL' + \Psi)^{-1}$ são os coeficientes de uma regressão multivariada dos fatores com as variáveis originais. As estimativas desses coeficientes produzem os escores fatoriais. Dados as observações X_j e tomando se os estimadores de máxima verossimilhança \hat{L} e $\hat{\Psi}$, os escores dos fatores são dados por:

$$\hat{F}_j = \hat{L}'(\hat{L}\hat{L}' + \hat{\Psi})^{-1}(X_j - \bar{X}) \quad (25)$$

O uso da identidade de matrizes $\hat{L}'(\hat{L}\hat{L}' + \hat{\Psi})^{-1} = (I + \hat{L}'\hat{\Psi}\hat{L})^{-1}\hat{L}'\hat{\Psi}^{-1}$ pode simplificar o cálculo dos escores dos fatores, os quais são dados por:

$$\hat{F}_j = (I + \hat{L}'\hat{\Psi}\hat{L})^{-1}\hat{L}'\hat{\Psi}^{-1}(X_j - \bar{X}) \quad (26)$$

Se as cargas fatoriais sofrerem rotação, então $\hat{L}^* = \hat{L}T$, logo os escores serão obtidos, fazendo $\hat{F}_j^* = T\hat{F}_j$.

2.6 A ANÁLISE DE AGRUPAMENTO

A análise de agrupamento é uma técnica multivariada amplamente utilizada para diversos fins em ciência florestal (MARTINS et al., 2003; FERRAZ et al., 2004; SANTOS et al. 2004, ARAÚJO et al., 2004, ALBUQUERQUE et al., 2006).

A análise de agrupamento tem por finalidade reunir, por algum critério de classificação às unidades amostrais em grupos, de tal forma que exista homogeneidade dentro do grupo e heterogeneidade entre grupos (CRUZ e REGAZZI, 1994; JOHNSON e WICHERN, 1998).

O processo de agrupamento se inicia tomando-se p medidas sobre cada um de n objetos. A matriz ($n \times p$) dos dados é então transformada em matriz ($n \times n$) de similaridades ou, alternativamente, medida de distância, onde as similaridades ou distâncias são computadas entre pares de objetos dentre as p variáveis. Em seguida, um algoritmo de agrupamento é selecionado, que define as regras relativas a como agrupar os objetos em subgrupos com base nas similaridades entre os objetos (DILLON e GOLDSTEIN, 1984).

Um grande número de medidas de similaridade ou de dissimilaridade tem sido proposto e utilizado em análise de agrupamento, sendo a escolha entre elas baseada na preferência e/ou na conveniência do pesquisador (BUSSAB et al., 1990).

Segundo Souza et al. (1997), com a definição da medida de dissimilaridade a ser utilizada, a etapa seguinte é a adoção de uma técnica de agrupamento para formação dos grupos.

Vários são os tipos de técnicas de agrupamento encontradas na literatura (MARDIA et al., 1997), das quais o pesquisador tem de decidir qual a mais adequada ao seu propósito, uma vez que as diferentes técnicas podem levar as diferentes soluções.

As técnicas de análise de agrupamento exigem de seus usuários a tomada de uma série de decisões independentes, que requerem o conhecimento das propriedades dos diversos algoritmos à disposição e que podem representar diferentes agrupamentos. Além disso, o resultado dos agrupamentos pode ser influenciado pela escolha da medida de dissimilaridade, bem como pela definição do número de grupos (GOWER e LEGENDRE, 1986; JACKSON et al., 1989; DUARTE et al., 1999).

As próprias técnicas de agrupar podem ser “classificadas” em grupos, e diferentes autores produzem diferentes classificações. Cormack (1971) propõe a seguinte:

1) A técnica hierárquica de agrupamento consiste em uma série de sucessivos agrupamentos ou sucessivas divisões de elementos, em que os elementos são agregados ou desagregados. As técnicas hierárquicas são subdivididas em aglomerativas e divisivas.

Os grupos, na técnica hierárquica, são geralmente representados por um diagrama bi-dimensional chamado de dendrograma ou diagrama de árvore. Nesse diagrama, cada ramo representa um elemento, enquanto a raiz representa o agrupamento de todos os elementos.

2) As técnicas não-hierárquicas, ou por particionamento, foram desenvolvidos para agrupar elementos em K grupos, em que K é a quantidade de grupos definida previamente.

A idéia central da maioria dos métodos por particionamento é escolher uma partição inicial dos elementos e, em seguida, alterar os membros dos grupos para obter-se a melhor partição (ANDERBERG, 1973).

Quando comparado com a técnica hierárquica, a técnica não-hierárquica ou por particionamento é mais rápido porque não é necessário calcular e armazenar, durante o processamento, a matriz de similaridade ou dissimilaridade (JOHNSON e WICHERN, 1998).

Em geral, os métodos por particionamento diferem entre si pela maneira que constituem a melhor partição. Como qualquer classificação, existirá tipos que serão difíceis de classificar, ou que poderão caber em mais de um grupo.

Dentre as técnicas de agrupamento usadas na literatura, no presente estudo usou-se a técnica de classificação não-hierárquica por k-médias.

2.6.1 Análise de Agrupamento por k-médias

O método do k-médias permite que objetos deixem um grupo e se associem a outro, de modo que esses reposicionamentos sucessivos possam contribuir para melhorar os resultados finais.

No agrupamento é necessário definir, a priori, o vetor central dos agrupamentos, ou centróides, inserindo os objetos mais próximos a eles. A partir do

agrupamento obtido, busca-se avaliar a hipótese da solução encontrada pelo agrupamento.

O algoritmo básico para um número inicial de k-agrupamentos é mostrado segundo AFIF e CLARK (1996):

- Os dados são divididos em k agrupamentos iniciais, definidos previamente;
- Calcula-se a média, ou centróide em cada um do k-agrupamentos;
- Para um dado individuo (ou objeto) é calculada a sua distância para cada um do k-centróides.
- Se o individuo estiver mais perto do centróide de seu próprio agrupamento, permanecerá no próprio agrupamento; caso contrário será deslocado para o grupo cujo centróide estiver mais perto dele.

O processo repete-se em todos os indivíduos até que não seja feita nenhuma troca dos elementos para outros agrupamentos.

2.7 A ANÁLISE DISCRIMINANTE

Para o estudo das diferenças entre os grupos formados, evidenciando as variáveis responsáveis pela diferenciação destes, empregam-se, segundo Mallo (1985), as técnicas de análise discriminante.

A função discriminante de Fisher tem numerosas aplicações, como, por exemplo, permite descobrir as ligações existentes entre um carácter qualitativo a ser explicado e um conjunto de caracteres quantitativos explicativos. Essa análise permite ainda, com a ajuda de um plano fatorial apropriado, descrever as ligações entre o carácter a ser explicado e os caracteres explicativos.

Para Johnson e Wichern (1998), a técnica multivariada, conhecida como análise discriminante ou de classificação, trata dos problemas relacionados em separar conjuntos distintos de objetos (observações) e alocar novos objetos (observações) em grupos previamente definidos.

Segundo Mardia et al. (1997), pode-se, por meio da análise discriminante, testar a suficiência de uma serie de variáveis discriminantes. A suficiência é testada pela análise discriminante *stepwise*, de grande utilidade, principalmente, quando há mais variáveis do que fossem necessárias para se obter uma classificação satisfatória.

Como procedimento separatório, a análise discriminante é empregada com a finalidade de investigar as diferenças observadas, caso estas não sejam bem entendidas. Para Marriott (1974), a análise discriminante consiste em investigar como e quando é possível fazer distinções entre os membros de “g” agrupamentos, com base nas observações feitas sobre eles.

Segundo o Souza (1990), os objetivos da análise discriminante são:

- Testar diferenças estatísticas, significantes a um dado nível de probabilidade entre “g” agrupamentos;
- Determinar o número de funções discriminantes;
- Construir regras de alocações para identificar um indivíduo como membro de um dos “g” agrupamentos e estimar as probabilidades de classificações corretas.

Ainda a análise se presta para testar a suficiência de uma serie de variáveis discriminantes (RAO e MITRA, 1973; LACHENBRUCH, 1979; MARDIA et al., 1997).

Uma das finalidades da Análise Discriminante é obter funções que permita classificar um indivíduo X, com base em medidas de um número p de características do mesmo, em uma de várias populações π_i , ($i=1,2,\dots,g$), distintas, buscando minimizar a probabilidade de classificar erroneamente um indivíduo, em uma população π_i , quando ele realmente pertence à população π_j , ($i \neq j=1,2,\dots,g$).

Para a classificação em uma de varias populações, recomenda-se o uso da função discriminante de Anderson (1984), no qual se tem uma observação \mathbf{x} e deseja-se classificá-la em uma dentre g populações, sendo $g > 2$.

Admitiu-se que a discriminação a ser feita envolvia custos idênticos de má classificação e probabilidades de má classificação, conhecidas para as varias populações.

Considerando que a densidade associada a cada população π_i tem distribuição normal multivariada de parâmetros μ_i e Σ_i , ($i = 1,2,\dots,g$), então a função discriminante (ou escore quadrático de classificação) é dada por:

$$D_i(\mathbf{x}) = \frac{1}{2} \ln |\Sigma_i| - \frac{1}{2} (\mathbf{x} - \mu_i)' \Sigma_i^{-1} (\mathbf{x} - \mu_i) + \ln(p_i) \quad (27)$$

No caso em que as matrizes de covariância de todas as populações são iguais ($\Sigma_i = \Sigma, i = 1, 2, \dots, g$), os termos que dependem de Σ_i e não de μ_i são constantes para as g populações e podem ser ignorados. Logo, o escore passa a ser linear e função discriminante é dada por:

$$D_i(\mathbf{x}) = \ell'_i \mathbf{x} - \frac{1}{2} \ell'_i \mu_i + \ln(p_i) \quad (28)$$

em que: $\ell'_i = \Sigma^{-1} \mu_i$.

A regra de decisão quanto à população de \mathbf{x} , consiste em classificar \mathbf{x} em π_i se e somente se:

$$D_i(\mathbf{x}) = \text{Max}(D_1(\mathbf{x}), D_2(\mathbf{x}), \dots, D_g(\mathbf{x})), \text{ para } g \geq 2.$$

As estimativas são dadas por: $D_i(\mathbf{x}) = \hat{\ell}'_i \mathbf{x} - \frac{1}{2} \hat{\ell}'_i \bar{x}_i + \ln(p_i)$

no qual temos $\hat{\ell}'_i = S_c^{-1} \bar{x}_i$, sendo $S_c = \frac{\sum_{i=1}^g (n_i - 1)}{n - g}$ e $\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}, i = 1, 2, \dots, g$, a variância combinada e a média de cada grupo g , respectivamente.

A taxa de erro aparente é definida pela seguinte expressão (JOHNSON e WICHERN, 1998):

$$T(\hat{R} | \hat{f}) = \frac{\sum_{i=1}^g m_i}{\sum_{i=1}^g n_i}$$

m_i é o número de observações de má classificação em π_i

n_i o número de observações em π_i .

Ou seja, a partir das estimativas das matrizes de covariância e de médias, obteremos as regras de discriminação com a finalidade de verificar a probabilidade de má classificação para cada agrupamento.

O resultado da classificação dos indivíduos levantadas em cada grupo indica a precisão do agrupamento por meio de uma Tabela, o qual permite, além da identificação das árvores correta e incorretamente classificadas, também aponta os grupos com maior probabilidade de alocação, ou seja, aqueles de cujos centróides estão mais próximos, conforme (FERREIRA e LIMA, 1978).

3. MATERIAL E MÉTODOS

3.1 DADOS UTILIZADOS

Foram utilizados dados de altura de um povoamento de *Leucaena leucocephala* de Wit. (Hawai), variedade K8, de procedência australiana. O experimento foi implantado em dezembro de 1989, na Estação Experimental da Empresa Pernambucana de Pesquisa Agropecuária (IPA) em Caruaru – PE, considerando 3 fatores com diferentes níveis o que resultou em 24 tratamentos repetidos inicialmente 24 vezes (Meunier, 1991). A Estação está situada a 08°14'18" latitude sul e 35°55'20" de longitude oeste e 537 m de altitude. Pela classificação climática de Thornthwaite esta área está enquadrada no tipo Dd'a' (semi-árido megatérmico).

As alturas inicialmente foram mensuradas em 19 ocasiões: aos 11; 37; 65; 98; 128; 159; 189; 220; 255; 287; 318; 361; 479; 571; 935; 1277; 1466; 1837 e 3256 dias após o plantio. Foram estudadas as fontes de variação do uso do composto de resíduo urbano, dentre os níveis de adubação com superfosfato simples (18% de P₂O₅) e estirpes de bactérias fixadoras de dinitrogênio atmosférico.

Durante o experimento observou-se, ao longo do tempo, a perda de 42 árvores, sendo o mesmo concluído com 534 das 576 mudas iniciais. Assim sendo, os 24 tratamentos que inicialmente tinham 24 repetições, foram alterados para um número que variou entre 17 e 24 (Tabela 3).

O número de plantas analisadas foi de 265, para o tratamento sem o composto orgânico e 269 para as plantas que receberam o composto de resíduo urbano, proveniente da usina de tratamento de lixo da Prefeitura da Cidade do Recife.

Segundo Meunier (1991), a fase de maior crescimento em altura das mudas iniciou-se por volta dos 55 dias após a semeadura, tendendo as taxas de crescimento a se estabilizarem mais cedo quando na ausência do composto orgânico, mostrando que o efeito do composto orgânico foi o que realmente influenciou o crescimento das mudas e das plantas até o quarto ano, o que também foi verificado por Souza (2003).

Com base nestas conclusões, as análises foram feitas levando-se em consideração dois grupos distintos de tratamentos: os que não receberam tratamento com composto orgânico e os que receberam composto orgânico, conforme Tabela 3.

Tabela 3 - Tratamentos aplicados no povoamento de *Leucaena leucocephala* (Lam.) de Wit., em Caruaru, PE

Tratamento	Nitrogênio	Fósforo	Sobrevivência
0	0	0	21
0	0	1	24
0	0	2	22
0	1	0	23
0	1	1	22
0	1	2	23
0	2	0	22
0	2	1	23
0	2	2	22
0	3	0	22
0	3	1	22
0	3	2	23
1	0	0	22
1	0	1	24
1	0	2	23
1	1	0	17
1	1	1	24
1	1	2	20
1	2	0	22
1	2	1	20
1	2	2	24
1	3	0	24
1	3	1	22
1	3	2	23

Composto orgânico: 0 = sem composto orgânico; 1 = com composto orgânico (100 ml/54 g);
 Nitrogênio: 0 = sem nitrogênio; 1 = sem inoculação (1,0g de uréia/recipiente); 2 = inoculação com rizóbio NFB 473; 3 = inoculação com rizóbio NFB 466.
 Fósforo: 0 = sem fósforo; 1 = 0,5g de Superfosfato simples/recipiente (82,8 g/m³); 2 = 1,0 g de Superfosfato simples/recipiente (165,6 g/m³).

Para o estudo do comportamento das leucenas ao longo dos anos, foi necessário interpolar novas medições, de forma a reduzir os grandes espaçamentos para aproximadamente 120 dias (entre 3 e 4 meses). Com isso, os dados que

continham as 19 medições receberam 18 medições, totalizando 37 medidas para o estudo, de acordo com as novas datas, a idade da planta correspondente (em mês) e os dias após o plantio (Tabela 4).

Tabela 4: Medições em altura, segundo a data, idade em meses e dias após o plantio.

Medição	H_i	Data	Idade (meses)	Nº. de Dias
1	H_1	01/01/1990	0,4	11
2	H_2	07/02/1990	1,2	37
3	H_3	07/03/1990	2,2	65
4	H_4	09/04/1990	3,3	98
5	H_5	09/05/1990	4,3	128
6	H_6	09/06/1990	5,3	159
7	H_7	09/07/1990	6,3	189
8	H_8	09/08/1990	7,3	220
9	H_9	13/09/1990	8,5	255
10	H_{10}	15/10/1990	9,6	287
11	H_{11}	15/11/1990	10,6	318
12	H_{12}	28/12/1990	12,0	361
13	H_{13}	25/04/1991	16,0	479
14	H_{14}	26/07/1991	19,0	571
15	H_{15}	25/11/1991	23,1	693
16	H_{16}	26/03/1992	27,2	815
17	H_{17}	24/07/1992	31,2	935
18	H_{18}	12/11/1992	34,9	1046
19	H_{19}	03/03/1993	38,6	1157
20	H_{20}	01/07/1993	42,6	1277
21	H_{21}	29/10/1993	46,6	1397
22	H_{22}	06/01/1994	48,9	1466
23	H_{23}	06/05/1994	52,9	1586
24	H_{24}	03/09/1994	56,9	1706
25	H_{25}	12/01/1995	61,2	1837
26	H_{26}	10/05/1995	65,2	1955
27	H_{27}	05/09/1995	69,1	2073
28	H_{28}	01/01/1996	73,0	2191
29	H_{29}	28/04/1996	77,0	2309
30	H_{30}	24/08/1996	80,9	2427
31	H_{31}	20/12/1996	84,8	2545
32	H_{32}	17/04/1997	88,8	2663
33	H_{33}	13/08/1997	92,7	2781
34	H_{34}	09/12/1997	96,6	2899
35	H_{35}	06/04/1998	100,6	3017
36	H_{36}	02/08/1998	104,5	3135
37	H_{37}	01/12/1998	108,5	3256

Obs. As medições em negrito, representam as interpolações, do modelo de Weibull e também linear, em função dos espaçamentos, variando de 4 em 4 meses.

3.2 O MÉTODO ESTATÍSTICO

3.2.1 Interpolação de Medidas com o Modelo de Weibull

A interpolação entre as medidas foi estimada por meio do modelo matemático de crescimento de Weibull (Tabela 5), ajustado por Brito (2005), para os tratamentos com composto orgânico e sem composto orgânico:

$$W_f = W_i - \beta \left[e^{-KT_f^\theta} - e^{-KT_i^\theta} \right] + \varepsilon_i \quad (30)$$

sendo:

W_f : a altura ou comprimento final

W_i : a altura ou comprimento inicial observado

T_f : tempo final das medições das alturas de leucena

T_i : tempo inicial das medições das alturas de leucena

ε_i : o erro aleatório, ou a parte estocástica

e : o antilogaritmo natural.

β , θ e K são os parâmetros do modelo.

Tabela 5: Parâmetros estimados e Índice de Ajuste, segundo o modelo de Weibull aplicado nos tratamentos sem composto e com composto orgânico (BRITO, 2005).

Parâmetros	Tratamento	
	Sem Composto Orgânico	Com Composto Orgânico
K	0,041556412	0,078534725
β	4,837531361	5,520878931
θ	0,907562809	0,755432332
I.A. (%)	99,184	99,112

Em casos nos qual a aplicação das equações usadas por Brito (2005) extrapolou a altura final das plantas, considerou-se o incremento periódico, com interpolações lineares.

3.2.2 Análise Multivariada

Após as interpolações das medições foi realizada a análise fatorial visando à redução da dimensionalidade das medições originais, identificando as mais relevantes ao modelo, aquelas que melhor descreveram o comportamento da leucena ao longo do tempo, a fim de classificar e detectar grupos que representem sua fase de crescimento.

As medidas mais significativas ao modelo (os fatores) foram extraídas de forma a representar as medições por meio de seu percentual de explicação da variabilidade de cada fator (as comunalidades).

O método de extração das cargas fatoriais foi a da análise por componentes principais, usando a matriz de correlação R , o qual é especificado em função dos pares de autovalores e autovetores $(\hat{\lambda}_i, \hat{e}_i)$, $i = 1, \dots, p$, em que $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_p$, pois $Var(Y_1) \geq Var(Y_2) \geq \dots \geq Var(Y_p)$.

A matriz de cargas fatoriais estimadas $L = \{\ell_{ij}\}_{p \times m}$ é descrita, conforme Johnson e Wichern (1998):

$$L = \{\sqrt{\hat{\lambda}_1} \hat{e}_1; \sqrt{\hat{\lambda}_2} \hat{e}_2; \dots; \sqrt{\hat{\lambda}_m} \hat{e}_m\} = \hat{P} \hat{\Lambda}^{1/2} \quad (31)$$

em que, \hat{P} é uma matriz ($p \times m$) dos autovetores amostrais e $\hat{\Lambda}$ é a matriz diagonal ($m \times m$) dos autovalores amostrais de R . Ou ainda:

$$\ell_{ij} = a_{ij} \sqrt{\hat{\lambda}_i} \quad (32)$$

no qual:

ℓ_{ij} : a carga fatorial da medição i , extraída no fator j .

a_{ij} : o autovetor normalizado, isto é, o vetor obtido tal que $a'_{ij} \cdot a_{ij} = 1$.

$\sqrt{\hat{\lambda}_i}$: a raiz quadrada do autovalor da matriz de correlação R .

Os estimadores das variâncias específicas são dados por:

$$\hat{\psi}_i = 1 - \hat{h}_i^2 = 1 - \sum_{j=1}^m \hat{\ell}_{ij}^2 \quad (33)$$

Sendo $m = p$, o número de fatores retidos no modelo os que apresentaram autovalores maiores ou iguais a um (KAISER, 1958). Logo, a proporção do total da variação explicada, devido ao j -ésimo fator comum obtidas pela matriz R , é dada por:

$$\%VarExpl = \frac{\hat{\lambda}_j}{\sum_{j=1}^m \hat{\lambda}_j} = \frac{\hat{\lambda}_j}{Traço(R)} = \frac{\hat{\lambda}_j}{p}, \quad j = 1, \dots, m. \quad (34)$$

A fim de melhorar a interpretabilidade das medições nos fatores e diminuir a correlação entre os fatores pelas suas cargas fatoriais, usou-se uma rotação ortogonal nos eixos coordenados do tipo Varimax, baseado no critério de Kaiser.

Conforme Barroso (2003), isso tenderá a fazer com que as cargas fatoriais (em módulo) sejam ou muito altas ou muito baixas, o que facilita a interpretação dos fatores, na medida em que, um conjunto de variáveis (ou medidas) teria correlação alta com o fator, enquanto que as demais tenderiam a ter correlações baixas.

Em seguida foi testada a validação do modelo por meio dos critérios da medida de adequabilidade da amostra de Kaiser-Meyer-Olkin (KMO), onde o KMO varia entre -1 e +1 comparando as correlações simples com as correlações parciais entre as variáveis, e o teste de esfericidade de Bartlett, utilizando-se a distribuição χ^2 , para testar se a matriz de correlação poderia ser considerada uma matriz do tipo identidade.

Os escores foram obtidos usando o método da regressão, como uma combinação linear dos fatores rotacionados e das medições padronizadas, conforme Manly (1997):

$$F^* = (G'G)^{-1} G'X \quad (35)$$

em que:

F^* : é o vetor de dimensão (mx1) de escores fatoriais

G : a matriz de dimensão (px1) das cargas fatoriais finais rotacionados

X : o vetor de dimensão (px1) de médias dos caracteres da k-ésima família (medições padronizadas).

As medições originais foram substituídas no modelo através dos escores fatoriais, quantificando o efeito de cada fator comum na expressão dos caracteres e permitindo classificar as leucenas em grupos estabelecidos por meio da análise de agrupamento, usando o algoritmo do k-médias, sendo testado para de k variando de 3 a 10 grupos pré-estabelecidos.

Na prática, a criação de grupos de unidades a partir da análise de agrupamento pelo algoritmo de agrupamento k-médias é resultante de uma informação a priori do número de classes que se deseja estabelecer.

Em classificação de sítios, Tonini et al (2002) citam que alguns analistas do assunto fixam o número de grupos com base nas características do povoamento a ser estudado, ou por algumas técnicas não multivariadas de ajustamento de curvas, usando o índice de sítio.

O grupo pré-estabelecido com maior poder de discriminação e do valor que maximiza a probabilidade geral de classificação correta; tanto nos dados com composto, como nos dados sem composto orgânico, é encontrado pela Análise Discriminante.

Todo o processamento dos dados e ajustamento dos modelos foi realizado através do software estatístico Statistical Package for Society Sciences, versão 13.0 (SPSS, 2004).

4. RESULTADOS E DISCUSSÃO

4.1 ANÁLISE DOS INDIVÍDUOS NA AUSÊNCIA DE COMPOSTO ORGÂNICO

Em 269 plantas que não receberam composto orgânico, observou-se que o crescimento em altura nas medições puderam ser descritas por combinações lineares em função de três fatores retidos, com autovalores maiores que um, obtidos pelo método das componentes principais (Tabela 6).

Os testes KMO e Bartlett não foram aplicados ao ajuste do modelo, pois se observou que o valor do determinante da matriz de correlação foi bem próximo à zero, indicando que a variabilidade entre cada medição em altura foi relativamente desprezível.

O modelo fatorial ajustado teve explicação de, aproximadamente, 94,60% da variabilidade total explicada pelos três fatores, por meio da rotação ortogonal varimax. As cargas fatoriais entre os fatores e as medições das alturas tiveram valores acima de 0,7 não sendo necessário o descarte de medida (Tabela 7).

Tabela 6: Autovalores e Percentual de Explicação da Variância dos Fatores por Componentes Principais e na Rotação Varimax, no Tratamento Sem Composto Orgânico.

Nº. de Fatores	Extração Inicial dos Fatores			Extração dos Fatores na Rotação Varimax		
	Autovalor	% da Variância	% Acumulado	Autovalor	% da Variância	% Acumulado
1	27,9014	75,4092	75,4092	18,6721	50,4652	50,4652
2	5,1016	13,7881	89,1973	12,1610	32,8676	83,3328
3	1,9989	5,4025	94,5998	4,1688	11,2670	94,5998

A matriz das cargas fatoriais estimadas, as comunalidades, as variâncias específicas em cada medição no tratamento sem o composto orgânico são exibidas na Tabela 7. As comunalidades foram maiores ou próximas de um, indicando uma alta qualidade de explicação da variância de cada medição atribuída ao fator.

Tabela 7: Matriz de cargas fatoriais, comunalidades e variâncias específicas segundo a idade da *Leucena leucocephala* (Lam.) de Wit., obtidos na Rotação Varimax, no Tratamento Sem Composto.

Idade (meses)	Fator			Comunalidade	Variância Específica
	1	2	3		
0,4	-0,0081	0,1367	0,9164	0,8585	0,1415
1,2	0,0424	0,2106	0,9605	0,9686	0,0314
2,2	0,0849	0,2923	0,9371	0,9708	0,0292
3,3	0,1332	0,5158	0,8051	0,9320	0,0680
4,3	0,1882	0,7552	0,5149	0,8709	0,1291
5,3	0,2470	0,8391	0,3563	0,8921	0,1079
6,3	0,3264	0,8645	0,2574	0,9201	0,0799
7,3	0,3459	0,8784	0,2220	0,9405	0,0595
8,5	0,3859	0,8749	0,2082	0,9577	0,0423
9,6	0,4019	0,8751	0,1920	0,9642	0,0358
10,6	0,4106	0,8737	0,1903	0,9682	0,0318
12,0	0,4322	0,8670	0,1885	0,9740	0,0260
16,0	0,4622	0,8491	0,1813	0,9675	0,0325
19,0	0,5743	0,7440	0,1554	0,9074	0,0926
23,1	0,5923	0,7392	0,1533	0,9206	0,0794
27,2	0,6284	0,7184	0,1461	0,9324	0,0676
31,2	0,7129	0,6092	0,0933	0,8881	0,1119
34,9	0,7277	0,6052	0,0900	0,9040	0,0960
38,6	0,7457	0,5939	0,0851	0,9161	0,0839
42,6	0,7665	0,5684	0,0732	0,9160	0,0840
46,6	0,7774	0,5582	0,0723	0,9212	0,0788
48,9	0,8193	0,4883	0,0699	0,9147	0,0853
52,9	0,8250	0,4823	0,0713	0,9183	0,0817
56,9	0,8346	0,4729	0,0696	0,9250	0,0750
61,2	0,9332	0,3206	0,0713	0,9788	0,0212
65,2	0,9351	0,3188	0,0721	0,9813	0,0187
69,1	0,9384	0,3147	0,0713	0,9847	0,0153
73,0	0,9418	0,3086	0,0711	0,9873	0,0127
77,0	0,9442	0,3034	0,0714	0,9887	0,0113
80,9	0,9459	0,2988	0,0718	0,9892	0,0108
84,8	0,9469	0,2949	0,0721	0,9888	0,0112
88,8	0,9475	0,2913	0,0722	0,9878	0,0122
92,7	0,9475	0,2893	0,0727	0,9867	0,0133
96,6	0,9475	0,2869	0,0729	0,9853	0,0147
100,6	0,9473	0,2847	0,0731	0,9838	0,0162
104,5	0,9471	0,2826	0,0732	0,9823	0,0177
108,5	0,9219	0,2677	0,0835	0,9286	0,0714
% Variância Explicada	75,4092	13,7881	5,4025	-	-

Obs. A comunalidade é uma medida da qualidade de explicação de uma variável e assume valores de 0 a 1, sendo alta de 0,700 a 1,000, média de 0,500 a 0,699 e baixa se abaixo de 0,500.

Usando 3 funções discriminantes, os agrupamentos de melhor classificação e discriminado pela análise discriminante, foram os de três e de cinco grupos, ambos com 95,2% de percentual de classificação correta. Em seguida temos o agrupamento com 6 grupos, com 93,3% de probabilidade de classificação correta (Tabela 9).

Tabela 9: Resultado da Análise Discriminante, segundo Percentual de Classificação Correta dos Grupos, para a *Leucena leucocephala* (Lam.) de Wit., no tratamento sem Composto Orgânico.

Número de Grupos	Percentual de Classificação Correta
3	95,2%
4	90,0%
5	92,6%
6	93,3%
7	92,6%
8	92,2%
9	95,2%
10	91,8%

Na Tabela 10, estão as estatísticas em termos da média e do desvio padrão das alturas em cada idade da planta, no agrupamento com 3 grupos.

Na idade de 0,4 a 3,3 meses, verifica-se que o grupo 2 teve maior crescimento nesta fase, variando de 0,87m a 1,01m, em média, enquanto no grupo 3, as menores alturas, que variou de 0,51m a 0,62m na sua média.

No período de 4,3 a 38,6 meses, ou seja, do 4º mês do 1º ano ao 3º mês do 3º ano de crescimento da planta, o grupo 1 obteve maior taxa de crescimento, com variação, em média, de 1,32m a 3,91m, enquanto que o grupo 3 apresentou crescimento entre 0,74m e 3,32m. Entretanto, o grupo 2 teve maior diferença entre os grupos 1 e 3, com amplitude de 2,67m na média.

Na fase do 7º mês do 3º ano ao 12º mês do 9º ano, que corresponde a 42,6 a 108,5 meses, o grupo 2, com variação de 3,95m a 5,54m foi o de maior crescimento em altura neste período, seguido do grupo 1, que variou de 4,0m a 5,03m na média. O grupo 3 foi o de menor crescimento, em média, de 3,42m a 4,86m em altura.

Tabela 10: Médias e Desvios-padrão das alturas da *Leucaena leucocephala* (Lam.) de Wit., agrupados em 3 grupos, de acordo com a idade, para o tratamento sem composto orgânico.

Idade (meses)	Grupo 1		Grupo 2		Grupo 3	
	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão
0,4	0,78	0,16	0,87	0,17	0,51	0,15
1,2	0,85	0,14	0,93	0,15	0,54	0,14
2,2	0,89	0,15	0,96	0,15	0,57	0,13
3,3	1,00	0,18	1,01	0,16	0,62	0,14
4,3	1,32	0,28	1,16	0,24	0,74	0,20
5,3	1,57	0,39	1,30	0,31	0,90	0,27
6,3	2,01	0,51	1,63	0,42	1,25	0,38
7,3	2,39	0,62	1,94	0,49	1,58	0,40
8,5	2,66	0,69	2,18	0,55	1,79	0,47
9,6	2,72	0,70	2,24	0,56	1,85	0,48
10,6	2,74	0,71	2,26	0,55	1,87	0,49
12,0	2,81	0,72	2,35	0,57	1,94	0,50
16,0	2,89	0,73	2,46	0,58	2,03	0,53
19,0	3,11	0,79	2,81	0,72	2,26	0,62
23,1	3,37	0,79	3,07	0,72	2,54	0,62
27,2	3,54	0,81	3,28	0,73	2,74	0,63
31,2	3,70	0,82	3,59	0,77	3,09	0,72
34,9	3,84	0,83	3,73	0,77	3,24	0,73
38,6	3,91	0,85	3,83	0,76	3,32	0,73
42,6	4,00	0,86	3,95	0,76	3,42	0,76
46,6	4,10	0,87	4,08	0,76	3,52	0,77
48,9	4,24	0,91	4,34	0,72	3,73	0,85
52,9	4,34	0,92	4,46	0,72	3,85	0,85
56,9	4,42	0,93	4,56	0,72	3,95	0,84
61,2	4,69	1,04	5,10	0,83	4,42	0,82
65,2	4,76	1,03	5,18	0,83	4,49	0,82
69,1	4,80	1,03	5,23	0,82	4,55	0,83
73,0	4,83	1,03	5,28	0,82	4,59	0,84
77,0	4,86	1,03	5,31	0,82	4,62	0,85
80,9	4,88	1,03	5,34	0,82	4,65	0,86
84,8	4,90	1,03	5,36	0,82	4,67	0,86
88,8	4,91	1,03	5,38	0,83	4,69	0,87
92,7	4,92	1,03	5,40	0,83	4,71	0,87
96,6	4,94	1,03	5,42	0,83	4,72	0,87
100,6	4,95	1,03	5,44	0,83	4,74	0,88
104,5	4,96	1,03	5,45	0,83	4,75	0,88
108,5	5,03	1,04	5,54	0,87	4,86	0,91
Total de plantas	132		93		44	

Na tabela 11, temos as estatísticas em termos da média e do desvio padrão das alturas em cada idade da planta, no agrupamento com 9 grupos.

Tabela 11: Médias e Desvios-padrão das alturas da *Leucaena leucocephala* (Lam.) de Wit., agrupados em 9 grupos, de acordo com a idade, para o tratamento sem composto orgânico.

Idade (meses)	Grupo 1		Grupo 2		Grupo 3		Grupo 4		Grupo 5	
	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão
0,4	0,68	0,16	0,29	0,12	0,89	0,12	1,03	0,13	0,74	0,14
1,2	0,73	0,15	0,31	0,13	0,96	0,07	1,07	0,11	0,81	0,10
2,2	0,76	0,14	0,32	0,13	1,01	0,06	1,11	0,11	0,85	0,10
3,3	0,83	0,17	0,33	0,12	1,18	0,15	1,17	0,09	0,89	0,11
4,3	1,07	0,28	0,38	0,14	1,59	0,28	1,39	0,16	1,05	0,19
5,3	1,26	0,34	0,44	0,18	1,95	0,32	1,58	0,21	1,21	0,26
6,3	1,61	0,41	0,63	0,21	2,56	0,34	2,01	0,28	1,57	0,34
7,3	1,92	0,44	0,96	0,36	2,99	0,29	2,41	0,32	1,87	0,37
8,5	2,16	0,50	1,13	0,43	3,31	0,28	2,74	0,37	2,13	0,40
9,6	2,20	0,50	1,23	0,39	3,37	0,31	2,81	0,39	2,21	0,39
10,6	2,22	0,50	1,23	0,39	3,43	0,32	2,83	0,40	2,24	0,39
12,0	2,29	0,52	1,40	0,47	3,48	0,29	2,93	0,37	2,33	0,37
16,0	2,36	0,54	1,53	0,47	3,63	0,34	3,03	0,36	2,46	0,39
19,0	2,59	0,64	1,84	0,54	4,27	0,55	3,32	0,39	2,90	0,55
23,1	2,85	0,64	2,13	0,54	4,55	0,56	3,59	0,37	3,17	0,52
27,2	3,03	0,64	2,32	0,44	4,78	0,60	3,79	0,36	3,37	0,49
31,2	3,23	0,66	3,10	0,27	5,01	0,65	4,00	0,37	3,74	0,53
34,9	3,37	0,66	3,28	0,27	5,17	0,64	4,14	0,37	3,89	0,51
38,6	3,44	0,66	3,43	0,29	5,27	0,60	4,21	0,36	4,00	0,49
42,6	3,53	0,67	3,50	0,26	5,38	0,51	4,29	0,37	4,15	0,48
46,6	3,64	0,67	3,64	0,26	5,52	0,51	4,39	0,37	4,28	0,47
48,9	3,78	0,68	3,89	0,31	5,81	0,36	4,58	0,39	4,57	0,43
52,9	3,88	0,69	4,00	0,31	5,92	0,36	4,69	0,39	4,68	0,42
56,9	3,96	0,69	4,11	0,31	6,03	0,36	4,79	0,40	4,78	0,42
61,2	4,27	0,71	4,31	0,32	6,49	0,70	5,28	0,45	5,41	0,48
65,2	4,34	0,71	4,39	0,31	6,57	0,69	5,35	0,45	5,49	0,48
69,1	4,39	0,71	4,44	0,30	6,61	0,69	5,41	0,46	5,54	0,47
73,0	4,42	0,71	4,47	0,30	6,63	0,68	5,45	0,46	5,59	0,47
77,0	4,45	0,71	4,49	0,31	6,65	0,67	5,48	0,47	5,62	0,47
80,9	4,47	0,71	4,51	0,32	6,67	0,66	5,50	0,48	5,65	0,47
84,8	4,49	0,71	4,53	0,33	6,68	0,66	5,53	0,49	5,68	0,47
88,8	4,51	0,72	4,54	0,34	6,70	0,66	5,55	0,49	5,70	0,47
92,7	4,52	0,72	4,56	0,35	6,71	0,66	5,56	0,50	5,72	0,47
96,6	4,54	0,72	4,57	0,36	6,72	0,65	5,58	0,50	5,74	0,47
100,6	4,55	0,72	4,58	0,37	6,73	0,65	5,60	0,51	5,75	0,47
104,5	4,56	0,72	4,60	0,37	6,74	0,65	5,61	0,51	5,77	0,47
108,5	4,65	0,78	4,63	0,40	6,92	0,79	5,68	0,57	5,83	0,46
Total de plantas	91		4		8		40		54	

Continua...

Tabela 11: Continuação

Idade (meses)	Grupo 6		Grupo 7		Grupo 8		Grupo 9	
	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão
0,4	0,37	0,12	0,87	0,10	0,95	0,15	0,71	0,12
1,2	0,42	0,10	0,90	0,11	1,08	0,12	0,79	0,11
2,2	0,48	0,09	0,92	0,11	1,17	0,15	0,86	0,09
3,3	0,58	0,16	0,95	0,11	1,34	0,14	1,09	0,13
4,3	0,82	0,24	1,01	0,13	1,77	0,17	1,58	0,19
5,3	1,16	0,33	1,08	0,14	2,13	0,32	2,00	0,22
6,3	1,75	0,31	1,30	0,22	2,66	0,42	2,62	0,27
7,3	2,16	0,21	1,51	0,31	3,20	0,50	3,17	0,24
8,5	2,46	0,29	1,62	0,36	3,53	0,62	3,46	0,24
9,6	2,57	0,27	1,65	0,35	3,61	0,61	3,54	0,26
10,6	2,61	0,28	1,67	0,35	3,66	0,64	3,58	0,25
12,0	2,67	0,30	1,69	0,36	3,72	0,62	3,65	0,25
16,0	2,79	0,32	1,76	0,36	3,76	0,64	3,70	0,25
19,0	3,11	0,39	1,87	0,38	3,90	0,70	3,83	0,32
23,1	3,38	0,40	2,14	0,38	4,14	0,71	4,09	0,31
27,2	3,60	0,43	2,29	0,41	4,29	0,69	4,27	0,31
31,2	3,97	0,71	2,51	0,52	4,43	0,68	4,39	0,33
34,9	4,15	0,71	2,63	0,53	4,56	0,69	4,55	0,34
38,6	4,27	0,67	2,70	0,55	4,57	0,68	4,65	0,35
42,6	4,55	0,63	2,74	0,56	4,59	0,68	4,75	0,37
46,6	4,68	0,63	2,84	0,56	4,68	0,70	4,86	0,36
48,9	4,94	0,61	3,04	0,56	4,77	0,72	5,05	0,40
52,9	5,05	0,61	3,14	0,56	4,88	0,72	5,16	0,40
56,9	5,15	0,60	3,22	0,57	4,96	0,72	5,25	0,40
61,2	5,64	0,26	3,47	0,63	5,18	0,60	5,54	0,48
65,2	5,71	0,25	3,54	0,63	5,25	0,60	5,61	0,47
69,1	5,77	0,24	3,59	0,63	5,29	0,60	5,66	0,47
73,0	5,80	0,23	3,63	0,64	5,32	0,60	5,69	0,46
77,0	5,83	0,21	3,65	0,65	5,34	0,59	5,72	0,46
80,9	5,86	0,20	3,68	0,65	5,36	0,59	5,73	0,45
84,8	5,88	0,19	3,69	0,65	5,38	0,58	5,75	0,45
88,8	5,89	0,19	3,71	0,66	5,39	0,58	5,77	0,45
92,7	5,90	0,19	3,73	0,66	5,41	0,58	5,78	0,45
96,6	5,92	0,19	3,74	0,66	5,42	0,58	5,80	0,45
100,6	5,93	0,19	3,75	0,66	5,44	0,58	5,81	0,45
104,5	5,94	0,19	3,77	0,67	5,45	0,58	5,82	0,45
108,5	5,97	0,20	3,91	0,78	5,53	0,58	5,87	0,45
Total de plantas	7		34		7		24	

Analisando a tabela 11, o grupo 2, teve menor crescimento em altura, de 0,29m a 0,33m, em média, seguido do grupo 6, com variação de 0,37m a 0,58m, do 1º mês ao 3º mês do primeiro ano de fase da planta. O grupo 8 apresentou maior crescimento, com variação de 0,95m no 1º mês a 1,34m no 3º mês de idade.

Na fase, que segue do 4º mês do 1º ano ao 3º mês do 3º ano após o plantio, o agrupamento com 8 grupos, em que variou entre 1,77m a 4,57m teve maior desenvolvimento em altura, seguido do grupo 9, este, com variação de 1,58m a 4,65m na média. O grupo de menor crescimento da leucena neste período foi o grupo 7, com 34 plantas agrupadas e variação de 1,01m a 2,70m em média.

Na idade do 7º mês do 3º ano a 12º mês do 9º ano, observa-se que o grupo 7 teve crescimento inferior aos demais, concluindo que este agrupamento foi o de menor crescimento em todas as fases da leucena, com variação média de 2,74 a 3,91m, embora que no 1º mês de medição, já possuía altura média de 0,87m.

O grupo de maior crescimento em altura, foi o grupo 3, com 8 plantas agrupadas, embora na 1ª fase, este mesmo grupo variou menos que o grupo 8, de 0,89m a 1,18m, na 3ª fase de crescimento, as alturas variaram de 5,38m a 6,92m em média.

Em comparação ao agrupamento com 3 grupos, observa-se que as maiores alturas médias foram mais discriminadas no 2º agrupamento (Figuras 2 e 3).

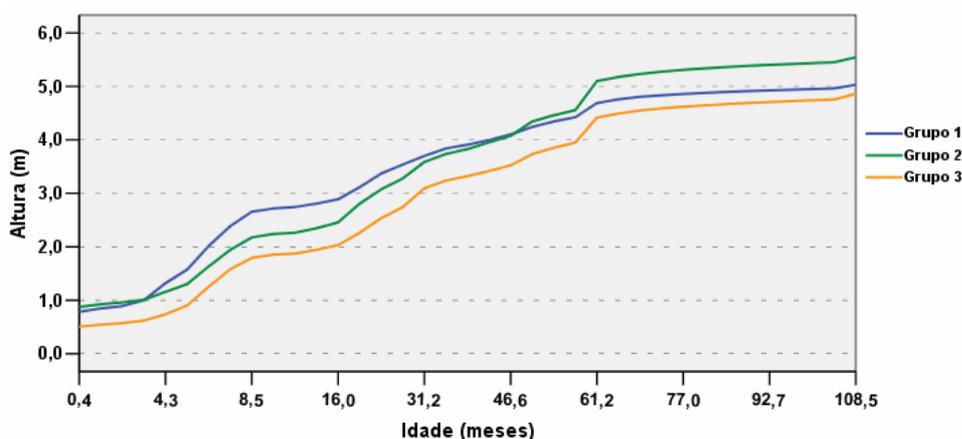


Figura 2: Crescimento em altura da *Leucaena leucocephala* (Lam.) de Wit, ao longo do tempo, no tratamento sem composto orgânico, no agrupamento com 3 grupos.

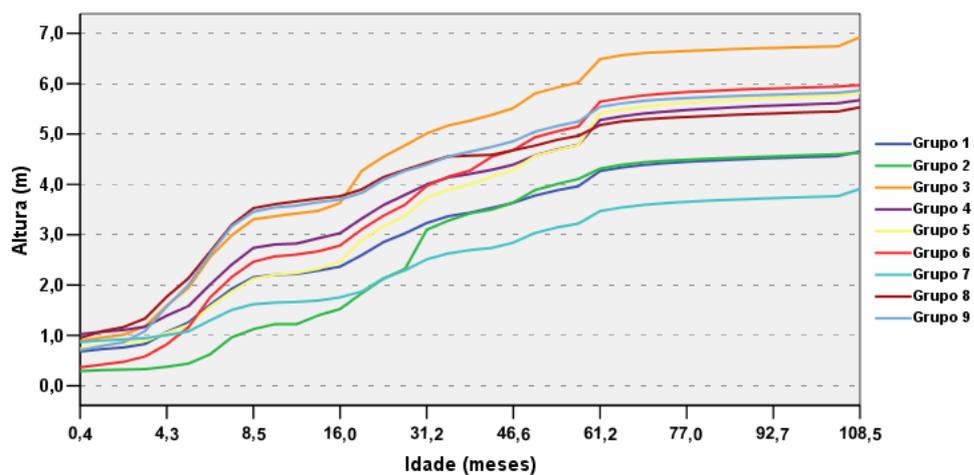


Figura 3: Crescimento em altura da *Leucaena leucocephala* (Lam.) de Wit, ao longo do tempo, no tratamento sem composto orgânico, no agrupamento com 9 grupos.

4.2 ANÁLISE DOS INDIVÍDUOS NA PRESENÇA DE COMPOSTO ORGÂNICO

A matriz de carregamentos fatoriais foi formada e os fatores foram, então, sem alteração do grau de explicação, rotacionados pelo método varimax normal em 37 medições para as 265 plantas de leucena que receberam composto orgânico, usando o método das componentes principais, estão apresentados nas Tabelas 12 e 13.

Como no modelo sem composto orgânico, também se observa que foram retidos três fatores, com autovalores maiores que um e que, juntos, explicam cerca de 94,30% da variabilidade total, o que permitiu a redução da dimensão original para três dimensões.

Segundo Furtado et al. (2003), as variáveis que apresentaram altos carregamentos fatoriais num mesmo fator podem ser agrupadas, procurando estabelecer alguma relação técnica que justifique o agrupamento, uma vez que a interpretação do agrupamento das variáveis é subjetiva.

A rotação varimax serviu como ferramenta ao auxílio de interpretação, determinando os possíveis agrupamentos de acordo com a similaridade das variáveis em cada fator.

As comunalidades foram relativamente altas e, conseqüentemente, o erro devido a aleatoriedade foram pequenos e próximos de zero, antes e após a rotação, o que torna imprescindível para a verossimilhança da classificação dos agrupamentos, mostrado na Tabela 13.

Tabela 12: Autovalores e percentual de explicação da variância na extração dos fatores por componentes principais e na rotação varimax, no tratamento com composto orgânico.

Nº. de Fatores	Extração Inicial dos Fatores			Extração dos Fatores na Rotação Varimax		
	Autovalor	% da Variância	% Acumulado	Autovalor	% da Variância	% Acumulado
1	27,2916	73,7611	73,7611	17,4970	47,2891	47,2891
2	5,4199	14,6484	88,4095	13,7086	37,0504	84,3395
3	2,1809	5,8944	94,3039	3,6868	9,9644	94,3039

Tabela 13: Matriz de cargas fatoriais, comunalidades e variâncias específicas segundo a idade da *Leucena leucocephala* (Lam.) de Wit., obtidos na Rotação Varimax, no Tratamento Com Composto.

Idade (meses)	Fator			Comunalidade	Variância Específica
	1	2	3		
0,4	-0,0930	0,0672	0,9430	0,9024	0,0976
1,2	-0,0240	0,1864	0,9609	0,9586	0,0414
2,2	0,0228	0,3324	0,9196	0,9568	0,0432
3,3	0,1388	0,6281	0,6988	0,9021	0,0979
4,3	0,2046	0,8123	0,4277	0,8845	0,1155
5,3	0,2470	0,8557	0,3095	0,8890	0,1110
6,3	0,2897	0,8720	0,2499	0,9068	0,0932
7,3	0,3241	0,8935	0,1956	0,9417	0,0583
8,5	0,3509	0,8948	0,1775	0,9554	0,0446
9,6	0,3607	0,8964	0,1559	0,9580	0,0420
10,6	0,3613	0,8973	0,1591	0,9611	0,0389
12,0	0,3757	0,8962	0,1475	0,9660	0,0340
16,0	0,4122	0,8780	0,1323	0,9582	0,0418
19,0	0,5063	0,8008	0,1033	0,9083	0,0917
23,1	0,5263	0,7961	0,0918	0,9192	0,0808
27,2	0,5618	0,7819	0,0643	0,9310	0,0690
31,2	0,6192	0,7089	-0,0317	0,8870	0,1130
34,9	0,6435	0,6964	-0,0287	0,8998	0,1002
38,6	0,6710	0,6791	-0,0240	0,9119	0,0881
42,6	0,7149	0,6317	-0,0082	0,9101	0,0899
46,6	0,7222	0,6243	-0,0061	0,9114	0,0886
48,9	0,7795	0,5377	0,0021	0,8968	0,1032
52,9	0,7827	0,5356	0,0024	0,8995	0,1005
56,9	0,7914	0,5270	-0,0006	0,9040	0,0960
61,2	0,9305	0,3342	0,0419	0,9793	0,0207
65,2	0,9320	0,3344	0,0419	0,9823	0,0177
69,1	0,9356	0,3304	0,0419	0,9862	0,0138
73,0	0,9389	0,3258	0,0401	0,9892	0,0108
77,0	0,9418	0,3203	0,0375	0,9910	0,0090
80,9	0,9440	0,3148	0,0350	0,9915	0,0085
84,8	0,9456	0,3097	0,0326	0,9911	0,0089
88,8	0,9465	0,3051	0,0307	0,9899	0,0101
92,7	0,9470	0,3018	0,0295	0,9887	0,0113
96,6	0,9471	0,2987	0,0284	0,9871	0,0129
100,6	0,9471	0,2957	0,0273	0,9853	0,0147
104,5	0,9469	0,2933	0,0263	0,9834	0,0166
108,5	0,9266	0,2634	-0,0002	0,9280	0,0720
% Variância Explicada	73,7611	14,6484	5,8944	-	-

Com a aplicação da análise discriminante, observou-se que a melhor discriminação de grupos está no agrupamento com 10 grupos, resultando em 95,1% de percentual de classificação correta das leucenas em cada agrupamento, dos oito modelos gerados pela análise discriminante; em seguida, os agrupamentos com 3 e 5 grupos tiveram percentual de classificação equivalente a 94,7%, conforme a Tabela 15.

Tabela 15: Percentual de Classificação dos Grupos Obtidos, usando Análise Discriminante para as leucenas com Composto Orgânico.

Número de Grupos Classificados	Percentual de Classificação
3	94,7%
4	93,2%
5	94,7%
6	92,5%
7	92,1%
8	94,3%
9	89,4%
10	95,1%

Na Tabela 16, mostram as estatísticas descritivas, em termos de média e desvio padrão, nas idades de crescimento da leucena, no agrupamento mais bem discriminado e classificado na análise discriminante.

Nas plantas que receberam composto orgânico, observa-se que na fase inicial de crescimento, os grupos 1 e 8 tiveram as menores variações de altura, de 0,50m a 0,82m e de 0,46m a 0,65m, respectivamente. Entre os que apresentaram maior crescimento, o grupo 4 variou na média entre 1,16m a 1,37m, do 1º mês ao 3º mês do 1º ano após o plantio, em 7 plantas agrupadas.

No 4º mês do 1º ano ao 3º mês do 2º ano, o qual representa a 2ª fase, o período de 4,3 a 38,6 meses de crescimento da planta, 15 plantas do grupo 10 tiveram alturas, variando de 1,77m a 5,06m em média. O grupo de menor crescimento nesta fase foi o grupo 2, variando de 0,85m a 2,39m, com 4 plantas agrupadas.

Para a fase que corresponde a idade de 42,6 meses a 108,5 meses (7º mês do 3º ano ao 12º mês do 9º ano da planta), percebe-se que o grupo 2 é o que possui menor crescimento em altura, de 2,40m a 3,09m na 37ª medição. Entre as maiores

alturas observadas, temos os grupos 1 e 9, com medições estimadas de 4,57 a 6,30m e de 4,04m a 6,23m, respectivamente.

Tabela 16: Médias e Desvios-padrão das alturas da *Leucaena leucocephala* (Lam.) de Wit., agrupados em 10 grupos, de acordo com a idade, para o tratamento com composto orgânico.

Idade (meses)	Grupo 1		Grupo 2		Grupo 3		Grupo 4		Grupo 5	
	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão
0,4	0,50	0,09	0,63	0,06	0,85	0,12	1,16	0,08	0,92	0,10
1,2	0,59	0,09	0,65	0,07	0,91	0,11	1,23	0,08	0,98	0,09
2,2	0,67	0,09	0,68	0,07	0,96	0,11	1,29	0,10	1,05	0,11
3,3	0,82	0,20	0,71	0,07	1,11	0,15	1,37	0,13	1,22	0,16
4,3	1,10	0,29	0,85	0,21	1,47	0,24	1,68	0,19	1,65	0,26
5,3	1,38	0,35	0,94	0,27	1,75	0,30	1,95	0,28	1,98	0,40
6,3	1,83	0,45	1,12	0,34	2,24	0,38	2,48	0,32	2,50	0,48
7,3	2,21	0,44	1,26	0,46	2,64	0,43	2,85	0,26	2,94	0,53
8,5	2,51	0,44	1,36	0,50	2,94	0,45	3,11	0,29	3,26	0,49
9,6	2,59	0,46	1,41	0,49	3,00	0,46	3,15	0,30	3,32	0,48
10,6	2,61	0,46	1,43	0,51	3,03	0,46	3,22	0,32	3,35	0,49
12,0	2,71	0,43	1,45	0,53	3,12	0,45	3,28	0,30	3,46	0,53
16,0	2,81	0,41	1,49	0,52	3,23	0,45	3,37	0,32	3,51	0,54
19,0	3,26	0,34	1,73	0,38	3,50	0,42	3,69	0,45	3,67	0,59
23,1	3,54	0,34	2,01	0,38	3,77	0,42	3,95	0,44	3,89	0,55
27,2	3,78	0,33	2,17	0,38	3,95	0,41	4,09	0,45	3,98	0,53
31,2	4,28	0,34	2,30	0,44	4,14	0,43	4,16	0,44	4,06	0,50
34,9	4,42	0,34	2,36	0,48	4,28	0,42	4,32	0,42	4,17	0,50
38,6	4,50	0,34	2,39	0,52	4,35	0,41	4,43	0,41	4,21	0,51
42,6	4,57	0,37	2,40	0,53	4,44	0,41	4,54	0,42	4,22	0,51
46,6	4,66	0,38	2,42	0,54	4,55	0,41	4,63	0,43	4,32	0,52
48,9	4,92	0,48	2,43	0,54	4,75	0,49	4,87	0,48	4,38	0,51
52,9	5,03	0,48	2,52	0,54	4,86	0,49	4,98	0,48	4,47	0,51
56,9	5,15	0,51	2,57	0,53	4,95	0,49	5,07	0,49	4,54	0,52
61,2	5,73	0,41	2,58	0,53	5,36	0,53	5,60	0,39	4,65	0,58
65,2	5,81	0,41	2,63	0,51	5,43	0,53	5,67	0,40	4,72	0,58
69,1	5,87	0,41	2,66	0,51	5,48	0,53	5,72	0,40	4,77	0,58
73,0	5,92	0,41	2,70	0,51	5,52	0,53	5,77	0,40	4,81	0,59
77,0	5,96	0,40	2,73	0,52	5,55	0,53	5,80	0,41	4,84	0,59
80,9	6,00	0,40	2,75	0,52	5,58	0,52	5,82	0,41	4,86	0,58
84,8	6,03	0,39	2,78	0,53	5,60	0,52	5,84	0,41	4,88	0,58
88,8	6,06	0,39	2,79	0,54	5,62	0,52	5,85	0,41	4,90	0,58
92,7	6,08	0,39	2,81	0,54	5,64	0,53	5,87	0,42	4,91	0,58
96,6	6,10	0,39	2,82	0,55	5,66	0,53	5,88	0,42	4,93	0,58
100,6	6,12	0,39	2,84	0,56	5,67	0,53	5,89	0,42	4,94	0,58
104,5	6,14	0,39	2,85	0,56	5,69	0,53	5,91	0,42	4,95	0,58
108,5	6,30	0,50	3,09	0,92	5,76	0,56	5,96	0,46	5,02	0,63
Total de plantas	18		4		124		7		21	

Continua...

Tabela 16: Continuação

Idade (meses)	Grupo 6		Grupo 7		Grupo 8		Grupo 9		Grupo 10	
	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão
0,4	0,96	0,09	0,62	0,13	0,46	0,12	0,92	0,12	0,68	0,13
1,2	1,00	0,11	0,65	0,12	0,50	0,13	0,96	0,13	0,80	0,11
2,2	1,02	0,11	0,69	0,12	0,56	0,17	0,98	0,13	0,93	0,09
3,3	1,06	0,12	0,73	0,12	0,65	0,25	1,01	0,14	1,25	0,17
4,3	1,14	0,15	0,87	0,19	0,90	0,39	1,11	0,21	1,77	0,22
5,3	1,25	0,19	1,01	0,23	1,12	0,42	1,22	0,29	2,24	0,30
6,3	1,56	0,28	1,37	0,26	1,46	0,48	1,53	0,37	2,92	0,33
7,3	1,80	0,32	1,69	0,20	1,81	0,36	1,78	0,39	3,46	0,38
8,5	1,97	0,37	1,90	0,22	2,10	0,35	2,06	0,40	3,78	0,33
9,6	2,02	0,37	1,95	0,23	2,17	0,32	2,14	0,38	3,87	0,34
10,6	2,04	0,38	1,97	0,22	2,18	0,31	2,16	0,36	3,92	0,33
12,0	2,08	0,40	2,04	0,25	2,33	0,29	2,25	0,32	3,99	0,34
16,0	2,13	0,41	2,16	0,29	2,47	0,33	2,41	0,29	4,11	0,30
19,0	2,27	0,45	2,40	0,36	2,68	0,39	2,92	0,44	4,32	0,35
23,1	2,53	0,42	2,68	0,36	2,95	0,39	3,20	0,43	4,57	0,35
27,2	2,68	0,43	2,90	0,37	3,14	0,40	3,38	0,38	4,70	0,40
31,2	2,84	0,50	3,22	0,48	3,36	0,38	3,59	0,29	4,87	0,55
34,9	2,94	0,51	3,37	0,48	3,49	0,41	3,76	0,29	4,99	0,51
38,6	2,99	0,51	3,46	0,45	3,55	0,46	3,86	0,27	5,06	0,49
42,6	3,01	0,51	3,61	0,42	3,59	0,52	4,04	0,31	5,12	0,47
46,6	3,11	0,52	3,72	0,41	3,69	0,52	4,16	0,31	5,20	0,46
48,9	3,23	0,55	3,97	0,47	3,78	0,48	4,45	0,40	5,28	0,47
52,9	3,33	0,54	4,08	0,47	3,86	0,47	4,56	0,40	5,39	0,47
56,9	3,41	0,54	4,17	0,47	3,91	0,46	4,66	0,40	5,49	0,47
61,2	3,67	0,58	4,51	0,48	4,12	0,53	5,60	0,37	5,72	0,46
65,2	3,74	0,58	4,58	0,48	4,19	0,54	5,66	0,37	5,79	0,45
69,1	3,79	0,59	4,63	0,47	4,23	0,55	5,72	0,36	5,84	0,43
73,0	3,82	0,59	4,66	0,48	4,27	0,55	5,78	0,36	5,87	0,43
77,0	3,84	0,59	4,70	0,49	4,30	0,55	5,83	0,36	5,89	0,43
80,9	3,85	0,59	4,72	0,49	4,32	0,56	5,87	0,35	5,90	0,43
84,8	3,86	0,60	4,75	0,50	4,34	0,56	5,91	0,35	5,92	0,43
88,8	3,87	0,60	4,77	0,50	4,36	0,57	5,94	0,35	5,93	0,43
92,7	3,88	0,60	4,79	0,51	4,38	0,57	5,97	0,35	5,95	0,43
96,6	3,89	0,60	4,80	0,51	4,40	0,58	5,99	0,35	5,96	0,43
100,6	3,90	0,60	4,82	0,51	4,41	0,58	6,02	0,35	5,97	0,43
104,5	3,91	0,60	4,83	0,52	4,43	0,59	6,04	0,35	5,98	0,43
108,5	3,92	0,61	4,96	0,63	4,52	0,57	6,23	0,51	6,05	0,47
Total de plantas	28		27		6		15		15	

Em comparação ao tratamento sem o composto orgânico, nos 9 grupos pré-determinados, o crescimento em altura, em média, foi superior ao tratamento que recebera composto orgânico, em mais de 6,92m no último mês do 9º ano do plantio, a 108,5 meses de idade (Figura 4).

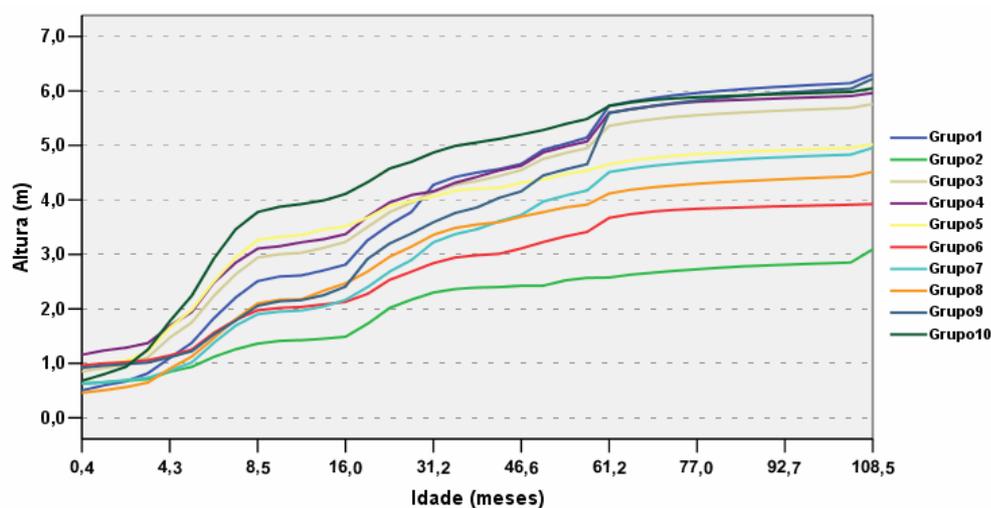


Figura 4: Crescimento em altura da *Leucaena leucocephala* (Lam.) de Wit, ao longo do tempo, no tratamento com composto orgânico, no agrupamento com 10 grupos.

CONCLUSÕES

Com as técnicas de análise multivariada, foi possível descrever o comportamento do crescimento em altura da *Leucaena leucocephala* (Lam.) de Wit. ao longo do tempo, na Estação Experimental de Caruaru, em PE.

O estudo mostrou que o modelo fatorial com 3 fatores representou as 3 fases de crescimento da leucena, a forma sigmóide nos gráficos de crescimento, podendo ser indicado para posteriores trabalhos.

O modelo das medições, em cujo tratamento não recebeu composto orgânico teve melhor descrição e classificação, do que aquelas que tiveram o tratamento sem o composto orgânico, constatado pelo Agrupamento do k-médias e pela Análise Discriminante, no qual, outros modelos não multivariados para ajustamento de curvas de crescimento e índice de sítio, obtiveram as mesmas conclusões (BRITO, 2005; SANTOS JUNIOR, 2005).

REFERÊNCIAS

AFIF, A.A.; CLARK, V. **Computer-aided multivariate analysis**. London: Chapman e Hall, 1996. 455p.

ALBUQUERQUE, M.A. et al. Estabilidade em análise de agrupamento: estudo de caso em ciência florestal. **Revista Árvore**, Viçosa-MG, v.30, n.2, 257-265, 2006.

ANDERBERG, M. R. **Cluster analysis for applications**. New York: Academic Press, 1973. 359p.

ANDERSON, T. W. **An introduction to multivariate statistical analysis**, New York: John Wiley e Sons, 1984. 675 p.

ANDRADE et al. Classificação ecológica do Estado da Paraíba 2. Delimitação e caracterização de regiões ecológicas a partir de variáveis climáticas. **Revista Árvore**, Viçosa-MG, v. 23, n. 2, p. 139-149, 1999.

ANDRADE, et al. Classificação ecológica do Estado da Paraíba. 3. Delimitação e caracterização de sub-regiões ecológicas a partir de variáveis biopedológicas. **Revista Árvore**, Viçosa-MG, v. 24, n. 2, p. 207-214, 2000.

ARAÚJO, M.M.et al. Análise de agrupamento da vegetação de um fragmento de floresta estacional decidual aluvial, Cachoeira do Sul, RS, Brasil. **Ciência Florestal**, Santa Maria, v. 14, n. 1, p. 133-147, 2004.

ASPIAZÚ, C. Classificação de sítios florestais mediante três métodos de análise da vegetação natural. **Revista Árvore**, Viçosa-MG, v.3, n.1, p. 1-15, 1979.

ASPIAZÚ, C. et al. Análise dos componentes principais aplicada na classificação climática do Estado de Minas Gerais - Teste metodológico. **Revista Árvore**, Viçosa-MG, v.14, n.1, p.1-15, 1990.

BARROS, N.F. de **Contribuição ao relacionamento de características pedológicas e topográficas com altura de *Eucalyptus alba* na região de Santa Bárbara, MG**. 1974. 88f. Dissertação (Mestrado em Ciência do Solo) – Universidade Federal de Viçosa, Viçosa-MG, 1974.

BARROSO, L. P.; ARTES, R. Análise multivariada. Lavras: Universidade Federal de Lavras, 2003. 157p.

BARTLETT, M. S. A note on multiplying factors for various chi squared approximations. **Journal of the Royal Statistical Society (B)**, London, 16, 1954. 296-298 p.

BARTLETT, M.S. The statistical conception of mental factors. **British Journal of Psychology**, London, v.28, p.97-104, 1937.

BRAGA, F. A. **Uso da análise discriminante na identificação de características ambientais determinantes da capacidade produtiva de povoamentos de eucalipto**. 1997. 65f. Tese (Doutorado em Ciência Florestal) - Universidade Federal de Viçosa, Viçosa-MG, 1997.

BREWBAKER, J.L. Leucaenas for the highland tropics. **Leucaena Research Reports**, Taipei, v.7, p.14-16, 1987.

BURTON et al. Use of multivariate methods in forest research site selection. **Canadian Journal Forest Research**, Ottawa, v.21, p.1573-1580, 1991.

BUSSAB, W.O.; MIYASAKI, E. S.; ANDRADE, D. F. **Introdução à análise de agrupamentos**. São Paulo: IME-USP, 1990. 105p.

BRITO, C.C.R. **Novos modelos de crescimento resultantes da combinação e das variações dos modelos de Chapman-Richards e Silva-Bailey aplicados em leucena [*Leucaena leucocephala* (Lam) de Wit] e tubarão-junteiro (*Carchathinus porosus*)**. 2005. 138f. Dissertação (Mestrado em Biometria) – Universidade Federal Rural de Pernambuco, Recife, 2005.

CARMEAN, H.W. Forest site quality valuation in the United States. **Advances in Agronomy**, San Diego, v.27, p. 209-69, 1975.

CLUTTER, J. L. et al. **Timber management: a quantitative approach**. New York: John Wiley & Sons, 1983. 333 p.

COMARCK, R. M. A review of classifications. **Journal of the Royal Statistical Society**, London, v.134, p. 321-353, 1971.

COOLEY, W.W.; LOHNES, R. R. **Multivariate data analysis**. New York: John Wiley & Sons, 1971. 364 p.

COSTA, N. de L. **Recomendações técnicas para o cultivo de leucena**. Porto Velho: EMBRAPA-UEPAE, 1987. 8 p. (Comunicado Técnico, 50).

CUNHA NETO, F. R. C. et al. Uso da diferença algébrica para construção de curvas de índice de sítio para *Eucalyptus grandis* e *Eucalyptus urophylla* na região de Luiz Antonio – SP. **Cerne**, Lavras, v.2, n.2, p.119-141, 1996.

CRUZ, C. D.; REGAZZI, A. J. Divergência genética. In: CRUZ, C. D.; REGAZZI, A. J. **Métodos biométricos aplicados ao melhoramento genético**. Viçosa, UFV: Imprensa Universitária, 1994. cap. 6, p. 287-323.

DILLON, W. R.; GOLDSTEIN, M. **Multivariate data analysis: methods and applications**. New York: John Wiley & Sons, 1984.

DUARTE, M. C.; SANTOS, J. B.; MELO, L. C. Comparison of similarity coefficients based on RAPD markers in the common bean. **Genetics and Molecular Biology**, Ribeirão Preto, v.22, n.3, p.427-432, 1999.

DRUMOND, M. A. **Leucena – uma arbórea de uso múltiplo, para a região semi-árida do nordeste brasileiro**. In: SIMPÓSIO BRASILEIRO DE CAPTAÇÃO DE ÁGUA DE CHUVA NO SEMI-ÁRIDO, 3. 2001, Petrolina. **Anais...** Petrolina: EMBRAPA, 2001.4p.

FERRAZ, E.M.N.; ARAÚJO, E.L. de; SILVA, S.I. da. Floristic similarities between lowland and montane areas of Atlantic Coastal Forest in Northeastern Brazil. **Plant Ecology**, Dordrecht, v.174, n.1, p.59–70, 2004.

FERREIRA, M. L.; LIMA, O. M. B. Processo de classificação. In: FAISSOL, S. et al. (Ed). **Tendências atuais na geografia urbano regional: teorização e quantificação**. Rio de Janeiro: IBGE, 1978. p. 113-30.

FERREIRA, D.F. **Análise multivariada**. Lavras: Universidade Federal de Lavras, 1996. 400 p.

FERREIRA, R. L. C.; SOUZA, A. L. **Técnicas de análise multivariada aplicadas ao manejo florestal no Brasil**. Sociedade de Investigações Florestais. Viçosa, MG: SIF, 1997. 21p. (Documento SIF, 14).

FINCHER, J., SMITH, M.L. **A discriminant function approach to ecological site classification in Northern New England**. Radnor: USDA, 1994. 12p. (Research paper NE 686).

FRANCO, A. A; SOUTO, S. M. **Leucaena leucocephala – uma leguminosa com múltiplas utilidades para os trópicos**. Rio de Janeiro: EMBRAPA-UAPNBS, 1986. 7p. (EMBRAPA-UAPNBS. Comunicado técnico, 2).

FREITAS, A. R. D. et al. **Leucaena leucocephala Lam. (de Witt): cultura e melhoramento**. São Carlos: Embrapa-UEPAE, 1991. 93p. (Documento 12).

FURTADO, E.M. et al. Ranqueamento de faxinais do Estado do Paraná através da análise fatorial. **Revista Ciências Exatas e Naturais**, Guarapuava, v.5, n.1, p. 85-98, jan./jun. 2003.

GERDING, V.; SCHLATTER, J.E. Variables Y factores del sitio de importancia para la productividad de Pinus radiata. D. Don em Chile. **Bosque**, Valdivia, v.16, n.2, p.39-56, 1995.

GERHARDT, E. J. et al. Contribuição da análise multivariada na classificação de sítios em povoamentos de Araucaria angustifolia (Bert) O. Ktze baseada nos fatores físicos e morfológicos do solo e no conteúdo de nutrientes da serapilheira. **Ciência Florestal**, Santa Maria, v. 11, n. 2, p. 41-57, 2001.

GOWER, J. C.; LEGENDRE, P. Metric and euclidean properties of dissimilarity coefficients, **Journal of Classification**, New York, v. 3, p. 5-48, 1986.

HAIR, J.F. et al. **Multivariate data analysis**. 5th ed. Upper Saddle River: Prentice Hall, 1988.

HAIR, J. F. JR. et al. **Multivariate data analysis (with readings)**. 4.ed., Englewood Cliffs: Prentice Hall, 1995. 745p.

HARDING, R.B.; GRIGAL, D.F.; WHITE E.H. Site quality evaluation for white Spruce plantations using discriminant analysis. **Soil Science Society of American Journal**, Madison, v.49, n.1, p.229-232, 1985.

HARMAN, H. H. **Modern factor analysis**. 3.rd ed., Chicago: University of Chicago Press, 1976. 467p.

HUSCH, B.; MILLER, C. I.; BEERS, T. W. **Forest mensuration**. 3.ed. New York: John Wiley & Sons, 1993. 402 p.

JACKSON, A. A.; SOMERS, K. M.; HARVERY, H. H. Similarity coefficients: measures for co-occurrence and association or simply measures of occurrence? **American Naturalist**, Chicago, v.133, p. 436-453, 1989.

JOHNSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis**. 4 ed. New Jersey: Prentice Hall, 1998, 816 p.

JOHNSTON, D. R.; GRAYSON, A. J.; BRADLEY. R. T. **Planeamento florestal**. Lisboa: Fundação Calouste Gulbenkain, 1977. 798 p.

KAYSER, Henry F. The varimax criterion for analytic rotation in factor analysis. **Psychometrika**, Williamsburg, v. 23, n. 3, p.187-200, 1958.

KILLIAN, W. Site classification and mapping: principles and trends. In: SYMPOSIUM ON SITE AND PROOUCTIVITY OF FAST GROWING PLANTATIONS, Pretoria, 1984. **Proceedings**. Pretoria, IUFRO, 1984. v.1, p.51-68.

LAAR, A. van. Multivariate analysis-a way to better understanding of complexity. **South African Forestry Journal**, Pretoria, v.141, p.34-40, 1987.

LACHENBRUCH, P.A. **Discriminant analysis**. New York: Hatner Press, 1979. 128p.

LAVEN, R.D. Establishing homogeneity in studies of forest succession. **Forest Ecology Management**, Amsterdam, v.4, p.161-77, 1982.

LEPSCH, I.F. et al. **Manual para levantamento utilitário do meio físico e classificação de terras no sistema de capacidade de uso**. Campinas, Sociedade Brasileira de Ciência do Solo, 1983. 175p.

MALLO, F. **Analisis de componentes principales y técnicas factoriales relacionadas: teoria, computacion y aplicaciones**. Leon: Universidad de Leon, 1985. 523p.

MANLY, B.F.J. **Multivariate statistical methods: a primer**. 2.ed. London: Chapman & Hall, 1997. 215p.

MARDIA, K.V.; KENT, J.T.; BIBBY, J.M. **Multivariate analysis**. London: Academy Press, 1997. 117p.

MARRIOTT, F.H.C. **The interpretation of multivariate observation**. London: Academic Press, 1974. 117p.

MARTINS, S.V. et al. Classificação ecológica do Espírito Santo, baseado em condições climáticas, **Revista Árvore**, Viçosa, MG, v.16, n.3, p.272-286, 1992.

MARTINS, S.V. et al. Distribuição de espécies arbóreas em um gradiente topográfico de Floresta Estacional Semidecidual em Viçosa-MG. **Scientia Forestalis**, Piracicaba, n. 64, p. 172-181, dez. 2003.

McBOYLE, G. Climatic classification of Australia by computer. In: McBOYLE, G. (Ed.). **Climatic in review**. Boston: Houghton Mifflin, 1973. p.110-118.

MEUNIER, I. M. J. **Crescimento de mudas de *Leucaena leucocephala* (Lam.) de Wit. Em função do uso de composto de resíduo urbano, adubação fosfatada e inoculação com *Rhizobium loti***. 1991. 110 f. Dissertação (Mestrado em Agronomia, Ciência do Solo) – Universidade Federal Rural de Pernambuco, Recife, 1991.

MEYER, L.F.F.; BRAGA, M.J. Elementos para uma tipologia do uso do solo agrícola no Estado do Pará: uma discussão sobre análise fatorial. In: CONGRESSO BRASILEIRO DE ECONOMIA E SOCIOLOGIA RURAL, 1999, Foz do Iguaçu. **Anais...** Foz do Iguaçu: SOBER, 1999.

NASCIMENTO, M.P.S.C.B. Germinação de leguminosas forrageiras nativas submetidas a tratamentos para quebra da impermeabilidade do tegumento. Teresina: EMBRAPA, 1982. 37p. (Boletim de pesquisa, 5).

NEWNHAM, R. M. Cluster analysis: an application in forest management planning. **Forestry Chronicle**, Ottawa, v.68, n.5, p.628-633, 1992.

OLIVEIRA, R. M. B.; ARLINDO, D. M.; PEREIRA, I. E. Avaliação de diferentes tamanhos de sacos de polietileno sobre o desenvolvimento de mudas de Leucena (*Leucaena leucocephala* (Lam). de Wit). **Revista de Biologia e Ciências da Terra**, Campina Grande, PB, v.4, n.2, 4p. 2º Semestre 2004.

OMI, P. N.; WENSEL, L. C.; MURPHY, J. L. application An of multivariate statistics to land use planning: classifying land units into homogeneous zones. **Forest Science**, Bethesda, v.25, n.3, p.399-414, 1979.

PEREIRA, R. C. et al. Efeitos da inclusão de forragem de leucena (*Leucaena leucocephala* (Lam.) de Wit) na qualidade da silagem de milho (*Zea mays* L.). **Ciência a Agrotecnologia**, Lavras, v. 28, n. 4, p. 924-930, jul./ago., 2004.

PREGITZER, K. S.; BARNES, B. V. The use o ground flora to indicate edaphic factors in upland ecosystems of the McCornick Experimental Forestal, upper Michigan. **Canadian Journal of Forest Research**, Ottawa, v.12, n.3, p.661-672, 1982.

QUEIROZ, W.T. **Análise de fatores ("factor analysis") pelo método da máxima verossimilhança: aplicação no estudo da estrutura de florestas tropicais**. 1984. 112f. Tese (Doutorado em Genética e Melhoramento) - Escola Superior de Agricultura Luiz de Queiroz, Piracicaba, 1984.

RAO, C. R.; MITRA, S. K. **Linear statistical inference and its applications**. New York: John Wiley, 1973.

REIS, E. **Estatística multivariada aplicada**. Lisboa: Edições Sílabo, 1997.

ROWE, J.S. Forest site classification in Canada: a current perspective. Prologue. **Forestry Chronicle**, Ottawa, v.68. n.1, p.22-24, 1992.

SALOMÃO, A. L. **Subsídios técnicos para elaboração do plano de manejo da floresta nacional do Rio Preto-ES**. 1998. 151 f. Tese (Doutorado em Ciência Florestal) – Universidade Federal de Viçosa, Viçosa-MG, 1998.

SANTOS, J.H. da S. et al. Distinção de grupos ecológicos de espécies florestais por meio de técnicas multivariadas. **Revista Árvore**, Viçosa-MG, v.28, n.3, p.387-396, 2004.

SANTOS JUNIOR, R. C. B. **Modelagem matemática na estimativa de crescimento em altura de leucena (*Leucaena leucocephala* (Lam.) de Wit.), no agreste de Pernambuco**. 2005. 95f. Dissertação (Mestrado em Ciências Florestais) – Universidade Federal Rural de Pernambuco, Recife.

SELLE, G. L.; SCHNEIDER, P. R.; FINGER, C. A. G. Classificação de sítio para *Pinus taeda* L., através da altura dominante, para a região de Cambará do Sul, RS, Brasil. **Ciência Florestal**, Santa Maria, v.4, n.1, p. 77-95, 1994.

SEIFFERT, N. F.; THIAGO, L. R. **Legumineira: cultura forrageira para a produção de proteína**. Campo Grande: EMBRAPA – CNPGC. 1983. 52 p. (EMBRAPA – CNPGC. Circular Técnica, 13)

SILVA, C. M. M. de. S. de. Avaliação de gênero *Leucena* na região semi-árida de Pernambuco. Petrolina: EMBRAPA – CPATSA, 1992, 21p. (*Boletim de pesquisa*, 44).

SCHEEREN, L. W.; FINGER, C. A. G. Efeito dos elementos químicos do solo e da copa no crescimento em altura de *araucaria angustifolia* (Bert.) O. Ktze., na região de Canela, RS. **Ciência Florestal**, Santa Maria, n.29, out-1998.

SCHNEIDER, P. R. **Introdução ao manejo florestal**. Santa Maria: UFSM, 1993. 348p.

SCHÖNAU, A.P. Problems in using vegetation or soil classification in determining site quality. **South African Forestry Journal**, Pretoria, n.141, p.13-18, 1987.

SCOLFORO, J. R. S. Curvas de índice de sítio para *Pinus caribaea* var. *hondurensis*. **Revista do IPEF**, Piracicaba, v.45, p.40-47, 1992.

SCOLFORO, J.R. **Mensuração florestal 4**: classificação de sítios florestais. Lavras: ESAL/FAEPE, 1993. 138p.

SOUZA, A. L. et al. Análises multivariadas para manejos de floresta natural na reserva florestal de Linhares, Espírito Santo: análises de agrupamento e discriminante. **Revista Árvore**, Viçosa-MG, v.14, n.2, p.85-101, jul/dez, 1990.

SOUZA, A. L.; FERREIRA, R. L. C.; XAVIER, A. Análise de agrupamento aplicada à área florestal. Sociedade de investigações florestais. Viçosa, MG: SIF, 1997. 109p. (Documento SIF, 16).

SOUZA, C. M. **Avaliação do crescimento em altura de leucena *Leucaena leucocephala* (Lam.) de Wit., no agreste de Pernambuco, por meio da análise multivariada de medidas repetidas**. 2003. 123 f. Dissertação (Mestrado em Biometria) – Universidade Federal Rural de Pernambuco, Recife, 2003.

STATISTICAL PACKAGE FOR THE SOCIAL SCIENCES – SPSS. **Base 10.0 User's Guide**. Chicago, 1999.

STATISTICAL PACKAGE FOR THE SOCIAL SCIENCES – SPSS. **Command Syntax Reference**. SPSS® 13.0. Chicago, 2004. p. 593-610.

SPURR, S. H.; BARNES, B. V. **Forest ecology**. 3 ed. New York: John Wiley, 1980. 687p.

TELES, M. M. et al. Métodos para quebra da dormência em sementes de leucena (*Leucaena leucocephala* (Lam.) de Wit.). **Revista Brasileira de Zootecnia**, Viçosa, MG, v.29, n.2, p.387-391, 2000.

TONINI, H. et al. Comparação gráfica entre curvas de índice de sítio para *Pinus elliottii* e *Pinus taeda* desenvolvidas no sul do Brasil. **Ciência Florestal**, Santa Maria, v. 12, n. 1, p. 143-152, 2002.

TONINI, H. et al. Crescimento em altura de *Pinus elliottii* Engelm., na região de Piratini no Rio Grande do Sul, Brasil. **Ciência Rural**, Santa Maria, v.31, n. 3, p.417-423, maio/jun. 2000.

WILLIAMS, B.K. Some observations on the use of discriminant analysis in ecology.
Ecology, Tempe, v.64, n.5, p.1283-1291, 1983.

XAVIER, D.F. *Leucena: procedimentos e cuidados para um bom estabelecimento*.
Coronel Pacheco: EMBRAPA - Gado de Leite, 1989. 3p. (Comunicado Técnico, 4).