

JORGE ROBERTO GROBE

**APLICAÇÕES DA ESTATÍSTICA MULTIVARIADA NA ANÁLISE DE
RESULTADOS EM EXPERIMENTOS COM SOLOS E ANIMAIS**

Dissertação apresentada como requisito parcial à
obtenção do grau de Mestre em Ciências, Curso
de Engenharia Pós –Graduação em Métodos
Numéricos em Engenharia – Programação
Matemática, Setores de Tecnologia e de Ciências
Exatas, Universidade Federal do Paraná.

Orientador: Prof Dr Jair Mendes Marques

CURITIBA

2005

Livros Grátis

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

AGRADECIMENTOS

Agradeço ao professor e orientador Dr. Jair Mendes Marques pelo acompanhamento e revisão do estudo, e aos professores Dr. André Brugnara Soares e Dr. João Alfredo Braida pelas críticas que propiciaram um maior aprofundamento nas questões polêmicas da pesquisa.

SUMÁRIO

LISTAS DE TABELAS	v
LISTAS DE FIGURAS	vii
RESUMO	ix
ABSTRACT	x
1 INTRODUÇÃO	11
1.1 JUSTIFICATIVA	11
1.2 OBJETIVOS	12
1.3 ESTRUTURA DA DISSERTAÇÃO	13
2 REVISÃO DE LITERATURA	14
2.1 USO DA ANÁLISE MULTIVARIADA EM EXPERIMENTOS AGRONÔMICOS	14
2.2 ANÁLISE ESTATÍSTICA MULTIVARIADA	16
2.2.1 Introdução	16
2.2.2 Conceitos	17
2.2.3 Estatísticas Descritivas	18
2.2.4 Densidade Normal Multivariada	20
2.2.4.1 Verificando a normalidade de uma distribuição	21
2.2.5 Inferência acerca do Vetor de Médias $\mu = E(\mathbf{X})$	24
2.2.5.1 Comparações de vetores médias de duas populações	24
2.2.5.2 Comparação entre vetores médios de várias populações	25
2.2.6 Análise de Componentes Principais	28
2.2.6.1 Introdução	28
2.2.6.2 Obtenção das componentes principais	29
2.2.6.3 Propriedades das componentes principais	30
2.2.6.4 Interpretação Das Componentes Principais	31
2.2.7 Análise Fatorial	31
2.2.7.1 Introdução	31
2.2.7.2 Análise fatorial ortogonal	32
2.2.7.3 Método de obtenção de fatores	34
2.2.7.4 Algumas conclusões sobre análise fatorial	41
2.2.8 Análise Discriminante	42
2.2.8.1 Introdução	42
2.2.8.2 Análise discriminante para mais de duas populações	42
2.2.8.3 Método de Fisher	42
2.2.8.4 Problema geral de classificação	46
2.2.8.5 Classificação para populações normais	48
2.2.8.6 Avaliação da função de classificação	49
2.2.9 Análise de Agrupamentos	50
2.2.9.1 Introdução	50
2.2.9.2 Medidas de similaridade e dissimilaridade	51
2.2.9.3 Agrupamentos	53
3 MATERIAL E MÉTODOS	55
3.1 CARACTERIZAÇÃO DA AMOSTRA E DAS VARIÁVEIS	55

4 RESULTADOS E ANÁLISE	57
4.1 DISTRIBUIÇÃO NORMAL MULTIVARIADA	57
4.2 COMPARAÇÃO ENTRE VETORES DE MÉDIAS DE VÁRIAS POPULAÇÕES	63
4.3 ANÁLISE FATORIAL E ANÁLISE DE AGRUPAMENTOS	65
4.4 ANÁLISE DISCRIMINANTE	116
5. CONCLUSÃO	122
REFERÊNCIAS	123
ANEXOS	125

LISTAS DE TABELAS

TABELA 1 - ANÁLISE DA VARIÂNCIA MULTIVARIADA.....	26
TABELA 2 - DISTRIBUIÇÃO DE LÂMBDA DE WILKS	27
TABELA 3 - INTERVALOS DE CONFIANÇA PARA OS EFEITOS DOS TRATAMENTOS DAS POPULAÇÕES 1 E 2; 1 E 3, 1 E 4	64
TABELA 4 - INTERVALOS DE CONFIANÇA PARA OS EFEITOS DOS TRATAMENTOS DAS POPULAÇÕES 2 E 3, 2 E 4, 3 E 4	65
TABELA 5 - FATORES DA ROTAÇÃO VARIMAX E O MÉTODO DAS COMPONENTES PRINCIPAIS DA POPULAÇÃO 1	66
TABELA 6 – MATRIZ DE CORRELAÇÃO DA POPULAÇÃO 1	70
TABELA 7 – MATRIZ DOS RESÍDUOS DA POPULAÇÃO 1	71
TABELA 8 - FATORES DA ROTAÇÃO VARIMAX E O MÉTODO DAS COMPONENTES PRINCIPAIS DA POPULAÇÃO 2	72
TABELA 9 – MATRIZ DE CORRELAÇÃO DA POPULAÇÃO 2	76
TABELA 10 – MATRIZ DOS RESÍDUOS DA POPULAÇÃO 2	77
TABELA 11 - FATORES DA ROTAÇÃO VARIMAX E O MÉTODO DAS COMPONENTES PRINCIPAIS DA POPULAÇÃO 3	78
TABELA 12 – MATRIZ DE CORRELAÇÃO DA POPULAÇÃO 3	82
TABELA 13 – MATRIZ DOS RESÍDUOS DA POPULAÇÃO 3	83
TABELA 14 - FATORES DA ROTAÇÃO VARIMAX E O MÉTODO DAS COMPONENTES PRINCIPAIS DA POPULAÇÃO 4	84
TABELA 15 –MATRIZ DE CORRELAÇÃO DA POPULAÇÃO 4	88
TABELA 16 – MATRIZ DOS RESÍDUOS DA POPULAÇÃO 4	89
TABELA 17 - FATORES PELO MÉTODO DAS COMPONENTES PRINCIPAIS DA POPULAÇÃO 5	90
TABELA 18 – MATRIZ DE CORRELAÇÃO DA POPULAÇÃO 5	92
TABELA 19 – MATRIZ DOS RESÍDUOS DA POPULAÇÃO 5	93
TABELA 20 - FATORES DA ROTAÇÃO VARIMAX E O MÉTODO DAS COMPONENTES PRINCIPAIS DA POPULAÇÃO 6	94
TABELA 21 – MATRIZ DE CORRELAÇÃO DA POPULAÇÃO 6	97
TABELA 22 - MATRIZ DOS RESÍDUOS DA POPULAÇÃO 6	97
TABELA 23 - FATORES DA ROTAÇÃO VARIMAX E O MÉTODO DAS COMPONENTES PRINCIPAIS DA POPULAÇÃO 7	98
TABELA 24 –MATRIZ DE CORRELAÇÃO DA POPULAÇÃO 7	101
TABELA 25 - MATRIZ DOS RESÍDUOS DA POPULAÇÃO 7	101
TABELA 26 - FATORES DA ROTAÇÃO VARIMAX E O MÉTODO DAS COMPONENTES PRINCIPAIS DA POPULAÇÃO 8	102
TABELA 27- MATRIZ DE CORRELAÇÃO DA POPULAÇÃO 8	105
TABELA 28 – MATRIZ DOS RESÍDUOS DA POPULAÇÃO 8	105
TABELA 29 - FATORES DA ROTAÇÃO VARIMAX E O MÉTODO DAS COMPONENTES PRINCIPAIS DA POPULAÇÃO 9	106
TABELA 30 – MATRIZ DE CORRELAÇÃO DA POPULAÇÃO 9	108
TABELA 31 – MATRIZ DOS RESÍDUOS DA POPULAÇÃO 9	109
TABELA 32 - FATORES DA ROTAÇÃO VARIMAX E O MÉTODO DAS COMPONENTES PRINCIPAIS DA POPULAÇÃO 10	109
TABELA 33 – MATRIZ DE CORRELAÇÃO DA POPULAÇÃO 10	112
TABELA 34 – MATRIZ DOS RESÍDUOS DA POPULAÇÃO 10	112
TABELA 35 - FATORES DA ROTAÇÃO VARIMAX E O MÉTODO DAS COMPONENTES PRINCIPAIS DA POPULAÇÃO 11	113

TABELA 36 – MATRIZ DE CORRELAÇÃO DA POPULAÇÃO 11	115
TABELA 37 - MATRIZ DOS RESÍDUOS DA POPULAÇÃO 11.....	116
TABELA 38 - CENTRÓIDES DAS 5 POPULAÇÕES	117
TABELA 39 - FUNÇÕES DISCRIMINANTES	117
TABELA 40 – CLASSIFICAÇÃO DAS AMOSTRAS DA POPULAÇÃO 5	118
TABELA 41 – CLASSIFICAÇÃO DAS AMOSTRAS DA POPULAÇÃO 4	119
TABELA 42 – CLASSIFICAÇÃO DAS AMOSTRAS DA POPULAÇÃO 3	120

LISTAS DE FIGURAS

FIGURA 1 - DETERMINAÇÃO DO NÚMERO APROPRIADO DE COMPONENTES A SEREM RETIDOS	40
FIGURA 2 – DISPERSÃO ENTRE TRÊS INDIVÍDUOS MENSURADOS COM RELAÇÃO A DUAS VARIÁVEIS QUANTITATIVAS CONTÍNUAS	51
FIGURA 3 – LOCAÇÃO DO EXPERIMENTO AGRONÔMICO	55
FIGURA 4 – VERIFICAÇÃO DA NORMALIDADE DA POPULAÇÃO 1	57
FIGURA 5 – VERIFICAÇÃO DA NORMALIDADE DA POPULAÇÃO 2	58
FIGURA 6 – VERIFICAÇÃO DA NORMALIDADE DA POPULAÇÃO 3	58
FIGURA 7 – VERIFICAÇÃO DA NORMALIDADE DA POPULAÇÃO 4	59
FIGURA 8 – VERIFICAÇÃO DA NORMALIDADE DA POPULAÇÃO 5	59
FIGURA 9 – VERIFICAÇÃO DA NORMALIDADE DA POPULAÇÃO 6	60
FIGURA 10 – VERIFICAÇÃO DA NORMALIDADE DA POPULAÇÃO 7	61
FIGURA 11 – VERIFICAÇÃO DA NORMALIDADE DA POPULAÇÃO 8	61
FIGURA 12 – VERIFICAÇÃO DA NORMALIDADE DA POPULAÇÃO 9	62
FIGURA 13 – VERIFICAÇÃO DA NORMALIDADE DA POPULAÇÃO 10	62
FIGURA 14 – VERIFICAÇÃO DA NORMALIDADE DA POPULAÇÃO 11	63
FIGURA 15 – ORDENAÇÃO DAS VARIÁVEIS PELO MÉTODO DAS COMPONENTES PRINCIPAIS DA POPULAÇÃO 1	67
FIGURA 16 – ORDENAÇÃO DAS VARIÁVEIS PELA ROTAÇÃO VARIMAX DA POPULAÇÃO 1	68
FIGURA 17 - DENDROGRAMA DA POPULAÇÃO 1	69
FIGURA 18 – ORDENAÇÃO DAS VARIÁVEIS PELO MÉTODO DAS COMPONENTES PRINCIPAIS DA POPULAÇÃO 2	73
FIGURA 19 – ORDENAÇÃO DAS VARIÁVEIS PELA ROTAÇÃO VARIMAX DA POPULAÇÃO 2	74
FIGURA 20 - DENDROGRAMA DA POPULAÇÃO 2	74
FIGURA 21- ORDENAÇÃO DAS VARIÁVEIS PELO MÉTODO DAS COMPONENTES PRINCIPAIS DA POPULAÇÃO 3	79
FIGURA 22 –ORDENAÇÃO DAS VARIÁVEIS PELA ROTAÇÃO VARIMAX DA POPULAÇÃO 3	80
FIGURA 23 – DENDROGRAMA DA POPULAÇÃO 3	80
FIGURA 24 – ORDENAÇÃO DAS VARIÁVEIS PELO MÉTODO DAS COMPONENTES PRINCIPAIS DA POPULAÇÃO 4	85
FIGURA 25 – ORDENAÇÃO DAS VARIÁVEIS PELA ROTAÇÃO VARIMAX DA POPULAÇÃO 4	86
FIGURA 26 – DENDROGRAMA DA POPULAÇÃO 4	86
FIGURA 27 - DENDROGRAMA DA POPULAÇÃO 5	91
FIGURA 28 –ORDENAÇÃO DAS VARIÁVEIS PELO MÉTODO DAS COMPONENTES PRINCIPAIS DA POPULAÇÃO 6	95
FIGURA 29 – ORDENAÇÃO DAS VARIÁVEIS PELA ROTAÇÃO VARIMAX DA POPULAÇÃO 6	96
FIGURA 30 – DENDROGRAMA DA POPULAÇÃO 6	96
FIGURA 27 –ORDENAÇÃO DAS 11 VARIÁVEIS PELA ROTAÇÃO VARIMAX DA POPULAÇÃO 8	104
FIGURA 28 – DENDROGRAMA DA POPULAÇÃO 8	104
FIGURA 31 – DENDROGRAMA DA POPULAÇÃO 9	108
FIGURA 32 – ORDENAÇÃO DAS VARIÁVEIS PELO MÉTODO DAS COMPONENTES PRINCIPAIS DA POPULAÇÃO 10	110

FIGURA 33 – ORDENAÇÃO DAS VARIÁVEIS PELA ROTAÇÃO VARIMAX DA.....	
POPULAÇÃO 10	111
FIGURA 34 – DENDROGRAMA DA POPULAÇÃO 10.....	111
FIGURA 35 - ORDENAÇÃO DAS VARIÁVEIS PELO MÉTODO DAS COMPONENTES..	
PRINCIPAIS DA POPULAÇÃO 11.....	114
FIGURA 36 - ORDENAÇÃO DAS VARIÁVEIS PELA ROTAÇÃO VARIMAX DA.....	
POPULAÇÃO 11	114
FIGURA 37 - DENDROGRAMA DA POPULAÇÃO 11.....	115
FIGURA 38 - ESPAÇO DISCRIMINANTE DAS 5 POPULAÇÕES	116

RESUMO

Este trabalho tem o objetivo de aplicar o uso da estatística multivariada no resultado das análises de solos e variáveis derivadas de animais. O estudo foi realizado na propriedade do Sr. José Antônio Bueno, localizado no município de Água Doce, SC, na região denominada “Campos de Palmas”. A metodologia de estudo propõe inicialmente no resultado de análises de solo e variáveis derivadas de animais a verificação da normalidade dos dados. E através do teste T^2 de Hotteling verificar a existência de diferenças estatísticas entre vetores de médias das populações referentes aos resultados das análises de solos. Nesta fase preliminar utilizam-se técnicas das componentes principais no resultado de análises de solos, variáveis derivadas de animais com o intuito de resumir o padrão de correlação entre as variáveis e chegar a conjuntos de variáveis que sejam não correlacionados umas com as outras. Outra técnica estatística aplicada no resultado de análises de solos e variáveis derivadas dos animais é a análise fatorial que descreve a estrutura da dependência de um conjunto de variáveis através da criação de fatores que medem aspectos comuns. No resultado de análises de solos também é aplicada outra técnica de estatística multivariada chamada de análise discriminante, que diferencia ou classifica as referidas populações. Para a avaliação de função da classificação das diversas amostras é utilizado o procedimento de validação cruzada, sugerido por Lachembbruch & Mickey. Outra técnica estatística multivariada aplicada a variáveis derivadas de animais e resultados da análise de solos é análise de agrupamentos e análise fatorial que compara os grupos quanto a similaridade e dissimilaridade e correlação. Para determinação dos resultados foram utilizadas funções nos *softwares* MATLAB 7.0 e STATISTIC 5.1.

Palavras-chave: estatística multivariada, componentes principais, análise fatorial, análise discriminante, análise fatorial, análise de agrupamento, variáveis derivadas de animais, análises de solos.

ABSTRACT

This study aims at evaluating the use of multivariate statistics in the analysis of soil and variables derived from animals. The study was performed in the property of Mr. Jose Antonio Bueno, located in Agua Doce – SC, a region named “Campos de Palmas”. The methodology of study proposes firstly the analysis of the soil, the verification of the normality of data, and through the test T^2 of Hotteling, to verify the existence of different statistics among the vectors of the average of population concerning to the soil analysis results. In this preliminary phase, techniques of the main components were used in the analysis of the soil with the aim of condensing the standard of correlation among the variables and reach sets of variables, which are not, correlated each other. Another statistical technique applied in the result of the analysis and derived variables from animals is the factorial analysis, which describes the structure of the dependence of a set of variables through the creation of factors which measure the common aspects. In the result of analysis of the soil, it is also applied another multivariated statistical technique named discriminatory analysis, which differentiate or classify the referred populations. In order to evaluate the function of classification of diverse samples, it is used a process of crossed validation suggested by Lachembruch & Mickey. Another multivariate statistical technique applied to variables derived from animals and results of the analysis of the soil is the grouping and factorial analysis which compare the groups concerning to similarity and dissimilarity, and correlation. In order to determine the result, functions in the *softwares* MATLAB 7.0 and STATISTICS 5.1 were used.

Key words: multivariate statistics, main components, factorial analysis, discriminated, analysis, grouping analysis, variable derived from animals, analysis of the soil.

1 INTRODUÇÃO

1.1 JUSTIFICATIVA

Segundo NETO (2004) na vida cotidiana aparecem vários fatores que podem estabelecer várias decisões. Às vezes, quando não se identifica o fator, toma-se uma decisão a partir de uma intuição. Um grande número de variáveis são identificadas em todos os acontecimentos, sejam eles culturais ou naturais através de diversas ciências, como humanas ou naturais em que o homem pode conhecer a realidade e interpretar os acontecimentos. Para traduzir conhecimentos de um fenômeno analisado há uma necessidade de controlar, manipular, medir as variáveis que são consideradas relevantes e traduzir essas informações. Algumas informações obtidas de conhecimentos podem gerar uma dificuldade, pois a ciência não conhece a realidade, ela representa através de modelos e teorias de diversos estudos. Um ramo do conhecimento que aspira dificuldades da universalidade das explicações científicas e que implica na padronização dos dados é a avaliação estatística das informações. Outra maneira de fazer ciência reduzindo-se poucas variáveis chama-se estatística univariada. Para a estatística univariada existem vantagens e desvantagens quando se tem um estudo frente a várias variáveis e a uma única variável. No caso restrito de variáveis independentes é possível com segurança, interpretar fenômenos usando medidas de tendência central como média, moda, etc. e dispersão dos dados. Na estatística multivariada o fenômeno depende de muitas variáveis, com isso não basta conhecer as variáveis isoladas, mas conhecê-las na sua totalidade, pois uma depende da outra e as informações são fornecidas pelo conjunto e não individualmente. Então a estatística univariada analisa cada variável isoladamente e não o conjunto.

Para a estatística a pesquisa tem significado se conhecermos o verdadeiro problema sobre o universo que elegemos, quanto as variáveis e as metodologias de análises.

BUENO *et. al* (2004) em seu trabalho sobre pastagem nativa melhorada sob distintas intensidades de pastejo na região “CAMPOS DE PALMAS” aplicou nas variáveis derivadas de animais a estatística univariada, o que confirmou que o manejo de pastagem influenciou a produção individual de cada animal. Neste experimento agrônômico foram também realizadas análises de solos, variáveis medidas na pastagem, conforme projeto anexo 1. Nas variáveis de animais, medidas na pastagem e resultado de análises de solos pretende-se aplicar técnicas de estatística multivariada para analisar o comportamento do conjunto dessas variáveis.

Essas técnicas de estatística multivariada serão aplicadas neste experimento agrônomico como: distribuição normal multivariada, inferência sobre o vetor de médias, análise fatorial, análise de agrupamento e análise discriminante. Para a aplicação dessas técnicas pretende-se alcançar alguns objetivos.

1.2 OBJETIVOS

O objetivo geral deste trabalho é explicar o uso da aplicação da análise estatística multivariada em um experimento agrônomico. Para se alcançar estes objetivos é necessário atingir:

- Avaliar a normalidade dos dados.
- Investigar se as populações multivariadas têm o mesmo vetor de médias.
- Identificar quais são os vetores que diferem significativamente.
- Reduzir a dimensão da matriz de dados.
- Analisar quais as variáveis que explicam maior parte da variabilidade total dos dados.
- Obter combinações interpretáveis das variáveis.
- Descrever e analisar a correlação das variáveis.
- Encontrar fatores interpretáveis.
- Determinar funções das variáveis observadas que permitam classificar ou alocar novos objetos ou observações no grupo mais adequado.
- Identificar e classificar uma amostra de indivíduos ou objetos em um pequeno número de grupos mutuamente exclusivos quanto à similaridade e dissimilaridade.

1.3 ESTRUTURA DA DISSERTAÇÃO

A presente dissertação está estruturada da seguinte forma: introdução, em seguida o capítulo II, em que é apresentada a Revisão de Literatura que expõe estudos sobre aplicações da estatística multivariada em experimentos agronômicos e análise estatística multivariada.

No capítulo III, apresenta-se o material e os métodos e no capítulo IV os Resultados e a Análise.

As conclusões do trabalho estão no capítulo V e consta também neste trabalho os resultados de análises de solos, variáveis derivadas de animais e o experimento agronômico que serviu para referência deste estudo.

2 REVISÃO DE LITERATURA

2.1 USO DA ANÁLISE MULTIVARIADA EM EXPERIMENTOS AGRONÔMICOS

VIDIGAL *et. al* (1994) estudou a divergência genética entre cultivares de mandioca, avaliando dez características morfoagronômicas e duas variáveis relacionadas à qualidade das raízes. Foram testados nove cultivares de mandioca adaptadas ao Noroeste do Estado do Paraná através da estatística multivariada, usando bases canônicas e análise de agrupamento, com emprego da distância generalizada de Mahalanobis como medida da dissimilaridade genética. Essa técnica permitiu a formação de grupos de cultivares semelhantes, e ainda, identificar as características que menos contribuíram para a divergência genética, que foram: número médio de raízes, número médio de hastes; diâmetro médio das raízes: teor médio de amido; e diâmetro médio do caule.

Em estudo conduzido para analisar atributos físicos, químicos e mineralógicos de solos do bioma cerrado, GOMES *et. al* (2004) foi aplicada a análise de componentes principais, com intuito de avaliar diferenças de comportamento entre solos das superfícies geomórficas Sul – Americanas e Velhas, sob cobertura vegetal nativa. Essa técnica permitiu agrupar os solos em três grupos. Além disso, a análise de componentes principais auxiliou no entendimento das diferenças e similaridades dos ambientes pedológicos separados no campo.

Na ciência do solo, a estatística multivariada pode ser aplicada na discriminação de unidades de solo. Assim, mediante parâmetros morfométricos de bacias hidrográficas na cidade de Botucatu, SP, foram aplicadas técnicas de análise de agrupamento e análise de componentes principais. O objetivo foi avaliar a ação conjunta de parâmetros da bacia hidrográfica na rede de drenagem e do relevo. A análise multivariada mostrou-se eficiente na discriminação dos solos, quando utilizaram-se parâmetros de relevo. CARVALHO *et. al* (1990).

FERRAUDO *et. al* (2004) utilizaram análise multivariada de agrupamento para avaliar utilizou-se da distância euclidiana para identificar, segundo variáveis (características do solo) pré-definidas, grupos com atributos de solos similares. E através do geoprocessamento foi utilizado no desenvolvimento dos mapas temáticos com o objetivo de mostrar a distribuição dos grupos de solos, baseando-se na análise de agrupamento. Este trabalho foi desenvolvido na Faculdade de Ciências Agrárias e Veterinárias, Campus de Jaboticabal, SP, onde utilizaram a carta de solos do Projeto RADAMBRASIL, Volume 32 (Rio de Janeiro-Vitória). O resultado deste estudo foi a construção de um dendrograma

resultante da análise de agrupamento que conteve dois grandes grupos com distância de corte no valor de 3,0. O grupo (G1) contendo os solos eutróficos e o grupo (G2) contendo os solos distróficos e álicos e com a distância de corte igual a 2,20 foram formados 4 grupos de solos: eutrófico, distrófico, distrófico e álico, álico. Quando estimada a distância euclidiana 1,6 observou-se 7 grupos com as seguintes características :eutrófico com textura argilosa e muito argilosa, distrófico com textura argilosa e muito argilosa, distrófico com textura média e argilosa, álico com textura argilosa e muito argilosa, distrófico com textura argilosa, álico com textura média e argilosa, álico com textura média com presença de arenosa. Portanto, os mapas com maior precisão foram os gerados com distância euclidiana de 2,20 e 1,60. As técnicas de geoprocessamento e análise de agrupamento foram adequadas para estudar a exploração de criação de mapas com atributos diagnósticos de solos. Então, a ferramenta da análise de multivariada como análise de agrupamento auxilia nos planejamentos e gerenciamentos regionais.

Os estudos cujo objetivo é identificar classes de plantas de acordo com classes de tolerância a um dado elemento tóxico, como por exemplo a tolerância de alumínio (Al), a análise multivariada poderá ser muito útil. Assim, por exemplo, DANTAS *et. al.* (2001) avaliaram a tolerância ao alumínio de 18 porta-enxertos somacionais e 3 variedades de macieira, obtidos por seleção *in vitro*. Para tanto, empregaram a análise de componentes principais e agrupamentos pelo método de *Ward* separando os clones em três classes de tolerância especificadas, a saber: não tolerante ou bem sensível, moderadamente tolerante e tolerante.

Com o objetivo de desenvolver e avaliar um método para discriminar solos a partir de suas respostas espectrais, em uma área do Sudoeste do Estado de São Paulo, utilizando um sensor em laboratório e modelos estatísticos, NANNI *et al.* (2004) utilizaram a estatística multivariada como ferramenta discriminante. O estudo envolveu um total de 370 amostras da porção superficial e da porção subsuperficial, coletadas em uma área de 185 hectares, sendo uma amostragem. A estatística multivariada permitiu individualizar e distinguir classes de diferentes solos, com acertos acima de 80%. O acerto global foi de 90,71% quando se utilizaram todas as classes para a geração dos modelos e de 93,44% quando se utilizaram as dez classes com maior número de indivíduos. Portanto, a análise discriminante mostrou-se eficiente com taxa média de acerto acima de 91%, ou seja, com erro global de apenas 8,8%. Para um subconjunto de 20% das amostras obteve-se um erro global de 33,9%, e, portanto, a análise multivariada demonstrou uma redução na qualidade.

2.2 ANÁLISE ESTATÍSTICA MULTIVARIADA

2.2.1 Introdução

BARROSO (2003), informa que a tecnologia disponível que analisa dados com mais de uma variável é conhecida como Análise de Estatística Multivariada. Com isso, os computadores podem analisar grandes quantidades de dados complexos. Portanto, com o aumento da expansão do conhecimento, essas técnicas estatísticas são utilizadas nas indústrias, centros de pesquisas e universidades. Citam-se alguns exemplos da Análise de Estatística Multivariada. Considere a Deinter (divisão territorial de polícias) do Estado de São Paulo que reuniu dados referentes às taxas de delitos por 10000 habitantes no ano de 2002, nas cidades de São José do Rio Preto, Ribeirão Preto, Bauru, Campinas, Sorocaba, São Paulo, São José dos Campos e Santos. Os delitos eram: homicídio doloso, furto, roubo; roubo e furto de veículos. Para cada município obteve-se uma taxa relativa aos delitos. Admita que se deseja dividir em 4 grupos de regiões que sejam homogêneas quanto à incidência de homicídios dolosos e furtos. Qual técnica multivariada é possível aplicar?

Em outro exemplo, na Universidade de Lavras foi desenvolvido um estudo sobre melões híbridos visando frutos mais produtivos e saborosos. As unidades amostrais são conjuntos de meloeiros, para as quais foram tomadas medidas das seguintes variáveis: o número total de melões por hectare; o peso médio dos melões (kg); a produção (kg); o número médio de melões por planta; o índice de formato (diâmetro transversal por diâmetro longitudinal) e teor de açúcar em graus brix. Se o objetivo é obter combinações interpretáveis e entender a estrutura de correlação das variáveis, qual técnica multivariada se pode usar?

Em outro exemplo, tem-se um pesquisador que realizou um estudo com variedades de feijão para avaliar as relações existentes entre 3 componentes primários versus 4 componentes secundários da produção de grãos de feijão. Os caracteres avaliados foram:

- Componentes primários: número de vagens/planta; número de grãos/vagem e o peso médio dos grãos.
- Componentes secundários: área foliar; número de folhas/planta; altura da planta; peso total da palha.

2.2.2 Conceitos

Para JOHNSON & WICHERN (1998), o objetivo da investigação científica com método multivariado tem os propósitos:

- Redução dos dados ou simplificação das estruturas: este fenômeno estudado é representado como uma simplificação das possibilidades de variáveis de informação.
- Classificação e Agrupamento: são grupos semelhantes de objetos ou criação de variáveis baseadas em vez de características medidas. Alternadamente pode classificar objetos dentro de um grupo bem definido.
- Entre outras.

Investigação de dependência entre variáveis: a relação natural entre variáveis é de interesse. São todas as variáveis mutuamente independentes ou são uma ou mais variáveis dependentes uma das outras?

- Previsão: relação entre variáveis determinadas com o propósito de prever valores de uma ou mais variáveis básicas observadas de outras variáveis.
- Construção de Hipóteses e Testes: a hipótese é formulada em termos de parâmetros da população multivariada. Esta permite ser uma suposição válida ou reforça uma convicção anterior.

Segundo LIMA (2002) os métodos estatísticos multivariados analisam inúmeras variáveis ao mesmo tempo. A Análise Multivariada é uma metodologia de grande potencial de aplicação, pois é possível interpretar diversas variáveis aleatórias ao mesmo tempo. As técnicas de Análise Multivariada ocupam diversas áreas do conhecimento e essas técnicas estão disponíveis em inúmeros *softwares* de acesso ao usuário.

Para FERREIRA (1996), a organização dos dados é representada de várias formas como gráficos, tabelas, etc. Os dados de uma pesquisa multivariada são relacionados $p \geq 1$ variáveis. A representação destes dados é feita matricialmente e cada elemento é dado por x_{jk} , que corresponde ao valor particular da j -ésima unidade amostral na k -ésima variável mensurada. As medidas das p variáveis em n unidades amostrais ou experimentais podem ser descritas em forma de uma matriz X com n linhas e p colunas da seguinte forma:

$$\mathbf{X} = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1k} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2k} & \cdots & X_{2p} \\ \vdots & \vdots & & \vdots & & \vdots \\ X_{j1} & X_{j2} & & X_{jk} & & X_{jp} \\ \vdots & \vdots & & \vdots & & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{nk} & \cdots & X_{np} \end{bmatrix}$$

2.2.3 Estatísticas Descritivas

Para FERREIRA (1996), na extração de informações visuais de um grande conjunto de dados tem-se um sério obstáculo, como para obter certos números, média amostral, desvio padrão, etc conhecidos como estatísticas descritivas. A estatística descritiva fornece informações dessas medidas que não são possíveis visualizar e que medem posição, variação e associação linear são enfatizadas a seguir:

Em LIMA (2002) uma medida estatística central é a média amostral $\bar{\mathbf{X}}$ que é uma estimação do vetor médio $\underline{\boldsymbol{\mu}}$, dado pela fórmula:

$$\bar{\mathbf{X}} = [\bar{X}_1, \bar{X}_2, \dots, \bar{X}_p] \text{ com } \bar{X}_j = \frac{\sum_{i=1}^n X_{ij}}{n} \text{ para } j=1, 2, \dots, p \quad (2.1)$$

A matriz de covariância do vetor $\bar{\mathbf{X}} = [X_1, X_2, \dots, X_p]$ é dado por:

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp}^2 \end{bmatrix}$$

onde σ^2 é a variância das variável aleatória X_i e σ_{is} é a covariância entre as variáveis X_i e X_j .

Para estimar a matriz de covariância populacional, Σ , utiliza-se a matriz de covariância amostral, S , que é dada por:

$$S = \begin{pmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & \cdots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \cdots & s_{pp} \end{pmatrix}$$

Onde cada s_{ij} é calculado pela fórmula:

$$s_{ij} = \frac{1}{n} \sum_{i=1}^n (X_{ij} - \bar{X}_j)^2 \text{ é a variância amostral da variável aleatória } X_j. \quad (2.2)$$

$$s_{ik} = \frac{\sum_{i=1}^n (X_{ij} - \bar{X}_j)(X_{ik} - \bar{X}_k)}{n-1} \text{ é a covariância amostral entre as variáveis } X_j \text{ e } X_k \quad (2.3)$$

A matriz de correlação do vetor \underline{X} é dada por $\rho = \begin{pmatrix} \rho_{11} & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{21} & \rho_{22} & \cdots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p1} & \rho_{p2} & \cdots & \rho_{pp} \end{pmatrix}$.

O estimador da matriz de correlação populacional S é dada pela matriz de correlação amostral R :

$$R = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{21} & 1 & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & 1 \end{bmatrix}$$

Onde cada elemento da matriz r_{jk} é calculado pelas seguintes fórmulas:

$$r_{jK} = \frac{S_{jK}}{S_j S_K} \quad (2.4)$$

onde S_{jK} é a covariância amostral entre as variáveis X_j e X_K e S_j e S_K para os desvios padrões das variáveis citadas.

2.2.4 Densidade Normal Multivariada

Para FERREIRA (1996), na Análise Multivariada a densidade normal com uma ou mais variáveis desempenha um papel muito importante. Na utilização de muitas técnicas de inferência de estatística multivariada parte-se do pressuposto de que os dados venham de uma distribuição normal.

Em JOHNSON & WICHERN (1998), a densidade normal univariada pode ser generalizada para a densidade normal multivariada com $p \geq 2$ dimensões. A distribuição normal univariada com média μ e variância σ^2 , tem função densidade de probabilidade é dada por :

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad x \in \mathbb{R}, \mu \in \mathbb{R} \text{ e } \sigma \in \mathbb{R}^+ \quad (2.5)$$

O termo $\frac{x-\mu}{\sigma}$ pode ser escrito como:

$$\left(\frac{x-\mu}{\sigma}\right)^2 = (x-\mu)(\sigma^2)^{-1}(x-\mu) \quad (2.6)$$

que é o expoente da função densidade normal univariada e mede o quadrado da distância de x para μ em unidade desvio padrão. O vetor das observações \underline{x} com dimensão p pode ter a expressão (2.6) generalizada por:

$$(\underline{\mathbf{x}} - \underline{\boldsymbol{\mu}})' \boldsymbol{\Sigma} (\underline{\mathbf{x}} - \underline{\boldsymbol{\mu}}) \quad (2.7)$$

Então, a função densidade de probabilidade (f.d.p.) multivariada do vetor $\underline{\mathbf{X}}$ é:

$$f(\underline{\mathbf{X}}) = \frac{1}{(2\pi)^{p/2} |\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{(\underline{\mathbf{X}} - \underline{\boldsymbol{\mu}})' \boldsymbol{\Sigma}^{-1} (\underline{\mathbf{X}} - \underline{\boldsymbol{\mu}})}{2}\right) \quad (2.8)$$

A densidade normal p-dimensional pode ser indicada como $N_p(\underline{\boldsymbol{\mu}}, \boldsymbol{\Sigma})$, com analogia a densidade normal univariada.

2.2.4.1 Verificando a normalidade de uma distribuição

Segundo JOHNSON & WICHERN (1998) a generalização da família da densidade normal na forma de “sino” para várias dimensões existem técnicas na análise multivariada. Quando os dados estão realmente perto da normal multivariada, a densidade normal aproxima-se da verdadeira distribuição populacional. Outra vantagem da distribuição normal multivariada é que existe tratamento matemático para obter resultados. Com a teoria do limite central várias distribuições amostrais de estatística multivariada podem aproximar-se da normalidade.

Pode-se avaliar a normalidade conjunta de dados de distribuição com 2, 3, ..., p dimensão. Na prática é suficiente e usual investigar as distribuições univariada e bivariada. As observações podem ser generalizadas para uma distribuição normal multivariada e a distribuição bivariada pode ser normal e que o contorno da constante da densidade pode ser uma elipse. No gráfico *scatter plot* é possível exibir uma elipse quando os dados tendem a uma normalidade.

O contorno da constante da densidade para p-dimensional da distribuição normal são elipses definidas por $\underline{\mathbf{X}}$ indicada por $(\underline{\mathbf{X}} - \underline{\boldsymbol{\mu}})' \boldsymbol{\Sigma}^{-1} (\underline{\mathbf{X}} - \underline{\boldsymbol{\mu}}) = c^2$. As elipses são centradas na $\underline{\boldsymbol{\mu}}$ (média) e tem eixos $\pm c\sqrt{\lambda_i} \mathbf{e}_i$ onde $\boldsymbol{\Sigma} \mathbf{e}_i = \lambda_i \mathbf{e}_i$ para $i = 1, 2, \dots, p$ onde $\boldsymbol{\Sigma}$ tem que ser uma

matriz definida positiva com o autovetor \underline{e} e os autovalores λ_i . O conjunto da distribuição bivariada é dado por $(\underline{\mathbf{X}} - \underline{\boldsymbol{\mu}})' \boldsymbol{\Sigma}^{-1} (\underline{\mathbf{X}} - \underline{\boldsymbol{\mu}}) \leq \chi_2^2(0,5)$ e tem probabilidade de 0.5, ou seja 50% da observação das amostras são dadas pela elipse $(\underline{\mathbf{X}} - \bar{\mathbf{X}})' \mathbf{S}^{-1} (\underline{\mathbf{X}} - \bar{\mathbf{X}}) \leq \chi_2^2(0,5)$, onde $\underline{\boldsymbol{\mu}}$ estima $\bar{\mathbf{X}}$ e $\boldsymbol{\Sigma}^{-1}$ estima \mathbf{S}^{-1} , caso contrário a normalidade não é aceita. Calculando as razões dos pontos de contorno e subjetivamente comparando com a teoria das probabilidades usuais.

Um método formal de julgar a normalidade de um conjunto de dados é baseada no quadrado da distância generalizada:

$$d_j^2 = (\underline{\mathbf{X}}_j - \bar{\mathbf{X}})' \mathbf{S}^{-1} (\underline{\mathbf{X}}_j - \bar{\mathbf{X}}), j=1, 2, \dots, n \quad (2.9)$$

onde $\underline{\mathbf{X}}_1, \underline{\mathbf{X}}_2, \dots, \underline{\mathbf{X}}_n$ são as observações das amostras. O procedimento descrito não é limitado somente para o caso bivariado, mas é válido para $p \geq 2$. Quando a população aproxima-se de uma normal multivariada e n e $n-p > 25$ ou 30, cada uma das distâncias $d_1^2, d_2^2, \dots, d_n^2$ comporta-se como uma variável aleatória qui-quadrado (χ^2). Para verificar estes resultados é através da lei dos grandes números:

Seja Y_1, Y_2, \dots, Y_n são observações independentes para a população com média $E(Y_i) = \mu$, então $Y = \frac{Y_1 + Y_2 + \dots + Y_n}{n}$ converge em probabilidade para μ e n para o infinito. Onde $\varepsilon > 0$, $\mathbf{P}[-\varepsilon < Y - \mu < \varepsilon]$ e a probabilidade aproxima-se da unidade quando n tende para o infinito.

Prova:

Como consequência a lei dos grandes números que é dado pelos vetores $\underline{\mathbf{X}}_i$ que converge em probabilidade para μ_i , $i=1, 2, \dots, p$, $\underline{\mathbf{X}}$ converge em probabilidade para $\boldsymbol{\mu}$. Também cada covariância amostral S_{iK} converge em probabilidade para σ_{iK} , $i, K=1, 2, \dots, p$ e $\hat{\mathbf{S}}$ (ou $\hat{\boldsymbol{\Sigma}} = \mathbf{S}_n$) converge em probabilidade para $\boldsymbol{\Sigma}$. Em consequência pode ser indicado por:

$$\begin{aligned}
(n-1)s_{iK} &= \sum_{j=1}^n (X_{ji} - \bar{X}_i)(X_{iK} - \bar{X}_K) \\
&= \sum_{j=1}^n (X_{ji} - \mu_i + \mu_i + \bar{X}_i)(X_{iK} - \mu_K + \mu_K - \bar{X}_K) \\
&= \sum_{j=1}^n (X_{ji} - \mu_i)(X_{iK} - \mu_K) + n(\bar{X}_i - \mu_i)(\bar{X}_K - \mu_K)
\end{aligned}$$

Seja $Y_j = (X_{ji} - \mu_i)(X_{iK} - \mu_K)$ com $E(Y_i) = \sigma_{iK}$ em que o primeiro termo s_{iK} converge σ_{iK} e o segundo converge para zero, isto é a lei dos grandes números.

Para analisar a normalidade dos dados mesmo que as distâncias não são independentes ou exatamente uma distribuição qui-quadrado, elas são úteis quando estão no gráfico. O resultado traçado no gráfico é chamado qui-quadrado ou gama, porque a distribuição qui-quadrado é um caso especial da generalização da distribuição gama. A construção do gráfico pela distribuição qui-quadrado:

1. Ordenar os quadrados das distâncias em ordem crescente.
2. O gráfico de pares de pontos $\left(q_{c,p} \left(\frac{j-0,5}{n} \right), d_j^2 \right)$ onde $q_{c,p} \left(\frac{j-0,5}{n} \right)$, está $100 \left(\frac{j-0,5}{n} \right)$ quartis abaixo da distribuição qui-quadrado com p graus de liberdade.

Os *quantiles* são especificados em proporção e os *percentis* são termos em porcentagem. Os quartis são relações acima dos percentis da distribuição qui-quadrado, particularmente

$$q_{c,p} \left(\frac{j-0,5}{n} \right) = \chi_p^2 \left(\frac{n-j+0,5}{n} \right).$$

Os pontos que estão dispersos no gráfico quando se traça uma linha reta aproximada pode acontecer que todos os pontos não pertencem à reta. Então uma curva sistemática sugere a falta de normalidade. Um ou dois pontos acima da linha indica-se uma grande distância. Os pontos que estão fora da linha reta, chamados de *outliers*, merecem uma atenção especial.

2.2.5 Inferência acerca do Vetor de Médias $\underline{\mu} = E(\underline{X})$.

2.2.5.1 Comparações de vetores médias de duas populações

Segundo JOHNSON & WICHERN (1998) o teste T^2 de Hotelling é feito para testar se dois vetores de médias são iguais, seguindo a mesma analogia do procedimento da estatística univariada. O teste T^2 compara a resposta média da população 1 com a da população 2 com tamanhos das amostras n_1 e n_2 . Nestas amostras calculam-se estatísticas que estimam parâmetros populacionais $\underline{\mu}_i$ e Σ_i .

Quando a diferença entre as médias das populações 1 e 2 for nula, significa que não existem efeitos dos tratamentos. Isto é o mesmo que testar a hipótese dos vetores de médias sejam iguais ($H_0 : \mu_1 = \mu_2$). Os pressupostos para aplicar o teste são: que ambas as populações de tamanhos n_1 e n_2 sejam normais multivariados e que as matrizes de covariância amostral sejam iguais.

Para testar a hipótese $H_0 = \underline{\mu}_1 = \underline{\mu}_2 = \underline{\delta}_0$, considera-se que:

$$E(\underline{\bar{X}}_1 - \underline{\bar{X}}_2) = E(\underline{\bar{X}}_1) - E(\underline{\bar{X}}_2) = \underline{\mu}_1 - \underline{\mu}_2 = \underline{0} \text{ e } V(\underline{\bar{X}}_1 - \underline{\bar{X}}_2) = V(\underline{\bar{X}}_1) + V(\underline{\bar{X}}_2) = \left(\frac{1}{n_1} + \frac{1}{n_2} \right) S_p$$

Onde S_p é a matriz de covariância amostral conjunta, dada por: $S_p = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2}{n_1 + n_2 - 2}$,

que estima a matriz covariância populacional Σ . A estatística do teste é:

$$T^2 = \left[\underline{\bar{X}}_1 - \underline{\bar{X}}_2 - \underline{\delta}_0 \right]^t \left[\left(\frac{1}{n_1} + \frac{1}{n_2} \right) S_p \right]^{-1} \left[\underline{\bar{X}}_1 - \underline{\bar{X}}_2 - \underline{\delta}_0 \right] \sim \left(\frac{n_1 + n_2 - 2}{n_1 + n_2 - p - 1} \right) p F_{p, n_1 + n_2 - p - 1}$$

onde $F_{p, n_1 + n_2 - p - 1}$ é a distribuição F ou Snedecor com p e $(n_1 + n_2 - p - 1)$ graus de liberdade.

2.2.5.2 Comparação entre vetores médios de várias populações

Em JOHNSON & WICHERN (1998), os tratamentos possuem covariância Σ fazendo um paralelo com o modelo univariado. A técnica Análise da Variância, usada para comparar e vetores de médias g populações (grupos) é baseada no modelo de observações multivariada:

$$\underline{\mathbf{X}}_{ij} = \underline{\boldsymbol{\mu}} + \underline{\boldsymbol{\tau}}_i + \underline{\boldsymbol{\varepsilon}}_{ij}, j = 1, 2, \dots, n_i \text{ e } i = 1, 2, \dots, g \quad (2.9)$$

Onde ε_{ij} são variáveis aleatórias independentes $N_p(\underline{\mathbf{0}}, \Sigma)$ em $\hat{\alpha} = \sigma^2 I$. O parâmetro $\underline{\boldsymbol{\mu}}$ é o vetor uma média global e $\underline{\boldsymbol{\sigma}}_i$ e $\underline{\boldsymbol{\tau}}_{-1}$ representa o efeito do tratamento i (grupo) com

$\sum_{i=1}^g n_i \tau_i = 0$. De acordo com o modelo (2.9), cada componente do vetor de observação

$\underline{\mathbf{X}}_{ij}$ satisfaz o modelo univariado $X_{ij} = \mu + \tau_i + \varepsilon_{ij}$. Cada erro dos componentes do vetor $\underline{\mathbf{X}}_{ij}$ são não correlacionados e a matriz de covariância Σ é a mesma para todas os grupos.

O teste da hipótese nula:

$$H_0 = \underline{\boldsymbol{\mu}}_1 = \underline{\boldsymbol{\mu}}_2 = \dots = \underline{\boldsymbol{\mu}}_g = \underline{\boldsymbol{\mu}} \text{ equivalente a } H_0 = \underline{\boldsymbol{\tau}}_1 = \underline{\boldsymbol{\tau}}_2 = \dots = \underline{\boldsymbol{\tau}}_g = \underline{\mathbf{0}}$$

É feito usando-se a distribuição do lambda de Wilks, detalhado adiante, através da tabela 1 da manova que segue:

TABELA 1 - ANÁLISE DA VARIÂNCIA MULTIVARIADA

Fonte de variação	Graus de liberdade	Matriz da soma de quadrados e produtos
Tratamento	$g-1$	$B = \sum_{i=1}^g n_i \left(\bar{\mathbf{X}}_i - \bar{\mathbf{X}} \right) \left(\bar{\mathbf{X}}_i - \bar{\mathbf{X}} \right)^t$
Resíduo	$\sum_{i=1}^g n_i - g$	$E = \sum_{i=1}^g \sum_{j=1}^{n_i} \left(\mathbf{X}_{ij} - \bar{\mathbf{X}}_i \right) \left(\mathbf{X}_{ij} - \bar{\mathbf{X}}_i \right)^t$
Total corrigido	$\sum_{i=1}^g n_i - 1$	$B + E = \sum_{i=1}^g \sum_{j=1}^{n_i} \left(\mathbf{X}_{ij} - \bar{\mathbf{X}} \right) \left(\mathbf{X}_{ij} - \bar{\mathbf{X}} \right)^t$

Em JOHNSON & WHICHERN (1998), o teste da igualdade dos vetores de médias, $H_0 = \underline{\boldsymbol{\mu}}_1 = \underline{\boldsymbol{\mu}}_2 = \dots = \underline{\boldsymbol{\mu}}_g$ envolve variâncias generalizadas. Rejeita-se H_0 se a razão das variâncias generalizadas dado pelo Lâmbda de Wilks, $\Lambda = \frac{|E|}{|B + E|}$, é maior que o valor do escore da distribuição exata (tabela 2).

A distribuição de Λ é escrito na tabela 2 Bartlett mostrou que se H_0 é verdadeira e $n = \sum_{i=1}^g n_i$ é grande, então :

$$-\left(n - 1 - \frac{p + g}{2} \right) \ln \Lambda = -\left(n - 1 - \frac{p + g}{2} \right) \ln \left(\frac{|E|}{|B + E|} \right) \sim \chi_{p(g-1)}^2 \quad (3.0)$$

Assim, esta estatística tem, aproximadamente distribuição qui-quadrado com $p(g-1)$ graus de liberdade. Em conseqüência, para $n = \sum_{i=1}^g n_i$ grande, rejeita-se H_0 com nível de significância

α se

$$-\left(n-1-\frac{p+g}{2}\right)\ln\left(\frac{|E|}{|B+E|}\right) > \chi_{p(g-1)}^2 \quad (3.1)$$

onde $\chi_{p(g-1)}^2(\alpha)$ é um escore correspondente a área de $1-\alpha$ da distribuição qui-quadrado com $p(g-1)$ graus de liberdade.

TABELA 2 - DISTRIBUIÇÃO DE LÂMBDA DE WILKS

Número de variáveis	Número de grupos	Distribuição Exata para dados normais multivariados
$p = 1$	$g \geq 2$	$\left(\frac{\sum_{i=1}^g n_i - g}{g-1}\right) \left(\frac{1-\Lambda}{\Lambda}\right) \sim F_{g-1, n-g}$
$p = 2$	$g \geq 2$	$\left(\frac{\sum_{i=1}^g n_i - g - 1}{g-1}\right) \left(\frac{1-\sqrt{\Lambda}}{\Lambda}\right) \sim F_{2(g-1), 2(n-g-1)}$
$p \geq 1$	$g = 2$	$\left(\frac{\sum_{i=1}^g n_i - p - 1}{p}\right) \left(\frac{1-\Lambda}{\Lambda}\right) \sim F_{p, n-p-1}$
$p \geq 1$	$g = 3$	$\left(\frac{\sum_{i=1}^g n_i - p - 2}{p}\right) \left(\frac{1-\sqrt{\Lambda}}{\sqrt{\Lambda}}\right) \sim F_{2p, 2(n-g-2)}$

2.2.6 Análise de Componentes Principais

2.2.6.1 Introdução

Análise de Componentes Principais é uma técnica estatística que transforma um conjunto de p variáveis em um conjunto com número menor (k) de variáveis aleatórias não-correlacionadas. Essas variáveis explicam uma parcela substancial das informações do conjunto original. As p variáveis Y_1, Y_2, \dots, Y_p , são denominadas componentes principais, de modo que Y_1 é aquela que explica a maior parcela da variabilidade total dos dados, Y_2 explica a segunda maior parcela e assim por diante. BARROSO (2003).

Os objetivos da Análise de Componentes Principais são:

- i) a redução da dimensão dos dados originais;
- ii) facilitar a interpretação das análises realizadas. Com a variabilidade dos dados podem ser explicadas por um número menor de componentes. A análise de componentes principais é uma técnica intermediária, portanto não é um método final como uma conclusão. Ela é aplicada em análise de regressão múltipla em casos de colinearidade ou de multicolinearidade. Também pode ser aplicada em análise de agrupamento que são utilizados com estimadores de fatores nas técnicas multivariadas chamadas de análise fatoriais.

BARROSO (2003) diz que, esta análise resume o padrão de correlação entre as variáveis e às vezes é possível chegar a algumas variáveis não correlacionadas entre elas, levando assim a um agrupamento. As componentes principais são combinações lineares das variáveis originais.

Geometricamente, essas combinações lineares representam a seleção de novos eixos coordenados, os quais são obtidos por rotação do sistema de eixos original, representado por X_1, X_2, \dots, X_p . Os novos eixos representam as direções máximas de variabilidade. FERREIRA (1996)

Em BARROSO (2003), a Análise de Componentes Principais dependerá apenas da matriz covariância (Σ) ou da matriz correlação (ρ) de X_1, X_2, \dots, X_p , ou seja do vetor \underline{X} .

2.2.6.2 Obtenção das componentes principais

Em JOHNSON & WHICHERN (1998), seja o vetor aleatório $\underline{\mathbf{X}}' = [X_1, X_2, \dots, X_p]$ com médias $\underline{\boldsymbol{\mu}}$ e matriz covariância Σ e autovalores $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$. Considerando as combinações lineares:

$$\begin{aligned} Y_1 &= \underline{\mathbf{c}}_1' \underline{\mathbf{X}} = c_{11}X_1 + c_{21}X_2 + \dots + c_{p1}X_p \\ Y_2 &= \underline{\mathbf{c}}_2' \underline{\mathbf{X}} = c_{12}X_1 + c_{22}X_2 + \dots + c_{p2}X_p \\ &\vdots \\ Y_p &= \underline{\mathbf{c}}_p' \underline{\mathbf{X}} = c_{p1}X_1 + c_{p2}X_2 + \dots + c_{pp}X_p \end{aligned} \quad (3.2)$$

$$\text{ou } \underline{\mathbf{Y}} = \mathbf{C}' \underline{\mathbf{X}} \text{ onde } \underline{\mathbf{Y}} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_p \end{bmatrix} \text{ e } \mathbf{C} = \begin{bmatrix} c_{11} & c_{12} & \dots & c_{1p} \\ c_{21} & c_{22} & \dots & c_{2p} \\ \vdots & \vdots & \dots & \vdots \\ c_{p1} & c_{p2} & \dots & c_{pp} \end{bmatrix} \text{ com :}$$

$$E(Y_j) = E(\underline{\mathbf{c}}_j' \underline{\mathbf{X}}) = \underline{\mathbf{c}}_j' E(\underline{\mathbf{X}}) = \underline{\mathbf{c}}_j' \underline{\boldsymbol{\mu}} \quad (3.3)$$

$$V(Y_j) = V(\underline{\mathbf{c}}_j' \underline{\mathbf{X}}) = \underline{\mathbf{c}}_j' V(\underline{\mathbf{X}}) \underline{\mathbf{c}}_j = \underline{\mathbf{c}}_j' \Sigma \underline{\mathbf{c}}_j \quad (3.4)$$

$$\text{Cov}(Y_i, Y_j) = V(\underline{\mathbf{c}}_i' \underline{\mathbf{X}}, \underline{\mathbf{c}}_j' \underline{\mathbf{X}}) = \underline{\mathbf{c}}_i' \Sigma \underline{\mathbf{c}}_j \quad (3.5)$$

para $i \neq j = 1, 2, \dots, p$. E a solução normalizada $\underline{\mathbf{c}}_i' \underline{\mathbf{c}}_i = \sum_{i=1}^p c_{ij}^2 = 1$

As componentes principais são combinações lineares não correlacionadas Y_1, Y_2, \dots, Y_p , em que a primeira componente principal é uma combinação linear $\underline{\mathbf{c}}_1' \underline{\mathbf{X}}$ que maximiza $V(\underline{\mathbf{c}}_1' \underline{\mathbf{X}})$ sujeita a restrição $\underline{\mathbf{c}}_1' \underline{\mathbf{c}}_1 = 1$, a segunda componente principal é a

combinação linear $\underline{c}_2' \underline{X}$ que maximiza $V(\underline{c}_2' \underline{X})$ sujeita a restrição $\underline{c}_2' \underline{c}_2 = 1$ e $\text{Cov}(\underline{c}_1' \underline{X}, \underline{c}_2' \underline{X}) = 0$ até j -ésima componente principal como combinação linear $\underline{c}_j' \underline{X}$ que maximiza $V(\underline{c}_j' \underline{X})$ sujeita as restrições $\underline{c}_j' \underline{c}_j = 1$ e $\text{Cov}(\underline{c}_i' \underline{X}, \underline{c}_j' \underline{X}) = 0$ para todo $i < j$.

2.2.6.3 Propriedades das componentes principais

Seja Σ a matriz da covariância associada com o vetor aleatório $\underline{X}' = [X_1, X_2, \dots, X_p]$ e com seus pares de autovalores-autovetores $(\lambda_1, \underline{e}_1), (\lambda_2, \underline{e}_2), \dots, (\lambda_p, \underline{e}_p)$, onde $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$. A j -ésima componente principal é dada por:

$Y_j = \underline{e}_j' \underline{X} = e_{1j}X_1 + e_{2j}X_2 + \dots + e_{pj}X_p$ para $j = 1, 2, \dots, p$ onde $V(Y_j) = \underline{e}_j' \Sigma \underline{e}_j = \lambda_j$ e $\text{Cov}(Y_i, Y_j) = \underline{e}_i' \Sigma \underline{e}_j = 0$ para $i \neq j$. Os autovalores λ_i são iguais e a escolha dos coeficientes dos vetores \underline{e}_i e Y_i não únicos.

Considere $\underline{X}' = [X_1, X_2, \dots, X_p]$ com matriz de covariância Σ com seus pares de autovalores e autovetores $(\lambda_1, \underline{e}_1), (\lambda_2, \underline{e}_2), \dots, (\lambda_p, \underline{e}_p)$ onde $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$. Sendo $Y_1 = \underline{e}_1' \underline{X}, Y_2 = \underline{e}_2' \underline{X}, \dots, Y_p = \underline{e}_p' \underline{X}$ são as componentes principais, então:

$$\sigma_{11} + \sigma_{22} + \dots + \sigma_{pp} = \sum_{i=1}^p \text{Var}(X_i) = \lambda_1 + \lambda_2 + \dots + \lambda_p = \sum_{i=1}^p \text{Var}(Y_i)$$

Se $Y_1 = \underline{e}_1' \underline{X}, Y_2 = \underline{e}_2' \underline{X}, \dots, Y_p = \underline{e}_p' \underline{X}$ são componentes principais obtidas da matriz covariância Σ , então:

$$\rho_{Y_j, X_i} = \frac{e_{iK} \sqrt{\lambda_i}}{\sqrt{\sigma_{KK}}} \quad \text{para } i, j = 1, 2, \dots, p \quad (3.6)$$

São correlações dos coeficientes entre as componentes principal Y_j e a variável X_i , onde

$(\lambda_1, \underline{e}_1), (\lambda_2, \underline{e}_2), \dots, (\lambda_p, \underline{e}_p)$ são autovalores-autovetores de Σ .

A proporção da variância total explicada à j -ésima componente principal é :

$$\frac{\lambda_j}{\lambda_1 + \lambda_2 + \dots + \lambda_p}, \quad j = 1, 2, \dots, p \quad (3.7)$$

Cada autovetor $\mathbf{e}'_j = [e_{1j}, e_{2j}, \dots, e_{pj}]$ pode auxiliar na interpretação da componente principal Y_j . A magnitude de e_{ij} mede a importância da i -ésima variável X_i para i -ésima componente principal Y_j . Na realidade, e_{ij} , é proporcional ao coeficiente de correlação entre Y_j e X_i .

2.2.6.4 Interpretação Das Componentes Principais

A interpretação das componentes principais é feita com base nas relações entre variáveis originais e as componentes principais e nos coeficientes dados pelas combinações lineares que levam às componentes principais. As correlações são medidas de cada contribuição individual de cada variável e não da contribuição multivariada das outras variáveis, mas os coeficientes são medidas das contribuições multivariadas. Geralmente utilizam-se as primeiras K componentes principais para uma análise, mas quanto menor for a parte desprezada, melhor o ajuste. Quando se tem 80% até 90% do total das variâncias da população, para um p (variáveis) grande, pode-se atribuir 1, 2 ou 3 componentes. Estas componentes podem representar a p variáveis originais com pouca perda de informação.

2.2.7 Análise Fatorial

2.2.7.1 Introdução

Segundo BARROSO (2003) a Análise Fatorial é uma técnica estatística que tem por objetivo descrever a estrutura de dependência de um conjunto de variáveis através de fatores, que são variáveis e que supostamente, medem aspectos comuns.

$$\mathbf{L}_{(pxm)} = \begin{bmatrix} l_{11} & l_{12} & \cdots & l_{1m} \\ l_{21} & l_{22} & \cdots & l_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ l_{p1} & l_{p2} & \cdots & l_{pm} \end{bmatrix} \quad \mathbf{F}_{mx1} = \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_m \end{bmatrix} \quad \mathbf{\varepsilon}_{px1} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_p \end{bmatrix} \quad \text{onde:}$$

o coeficiente l_{ij} é chamado de carga da i -ésima variável (X_i) do j -ésimo fator (F_j), e a matriz L é a matriz dos pesos ou cargas. A i -ésima erro do fator ε_i é associado com o i -ésimo vetor (\underline{X}_i). Os p desvio de $X_1 - \mu_1, X_2 - \mu_2, \dots, X_p - \mu_p$ são expressos em termos de $p+m$ variáveis aleatórias de $F_1, F_2, \dots, F_m, \varepsilon_1, \varepsilon_2, \dots, \varepsilon_p$ quando são variáveis não observáveis. O modelo fatorial de (3.8) distingue do modelo de regressão multivariada em que as variáveis são independentes ou seja a posição do \mathbf{F} pode ser observada.

Com a existência de uma grande quantidade de variáveis não observáveis ficaria impossível verificar a direção das variáveis observadas no modelo fatorial. Assim, os vetores das variáveis aleatórias de \mathbf{F} e $\mathbf{\varepsilon}$, no modelo (3.8) implica uma relação da covariância que pode ser verificada.

Assume que:

$$1) E(\mathbf{F}) = \mathbf{0}_{px1} \quad \text{Cov}(\mathbf{F}) = E(\mathbf{F}\mathbf{F}') = \mathbf{I}_{mxm} \quad E(\mathbf{\varepsilon}) = \mathbf{0}_{(px1)}$$

$$2) \text{Cov}(\mathbf{\varepsilon}) = E(\mathbf{\varepsilon}\mathbf{\varepsilon}') = \mathbf{\Psi}_{pxp} = \begin{bmatrix} \psi_1 & 0 & \cdots & 0 \\ \vdots & \psi_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \psi_p \end{bmatrix} \quad (4.0)$$

3) \mathbf{F} e $\mathbf{\varepsilon}$ são independentes, ou seja, $\text{Cov}(\mathbf{\varepsilon}, \mathbf{F}) = \mathbf{0}_{pxm}$. Essas hipóteses e as relações (3.8) constituem o modelo fatorial ortogonal.

O modelo fatorial ortogonal implica na estrutura da covariância:

$$1) \Sigma = \text{cov}(\mathbf{X}) = \mathbf{L}\mathbf{L}' + \mathbf{\Psi} \quad \text{ou} \quad \text{Var}(X_i) = l_{i1}^2 + l_{i2}^2 + \dots + l_{im}^2 + \psi_i$$

$$\text{Cov}(X_i, X_k) = l_{i1}l_{k1} + \dots + l_{im}l_{km}$$

$$2) \text{Cov}(\underline{\mathbf{X}}, \underline{\mathbf{F}}) = \mathbf{L} \text{ ou } \text{Cov}(X_i, F_j) = l_{ij}$$

A porção da variância da i -ésima variável que contribui para m fatores comuns é chamada de i -ésima comunalidades. E a porção da variância $\text{Var}(X_i) = \sigma_{ii}$ devido ao fator específico que é chamado com frequência variância específica (Ψ_i). Denota-se que a i -ésima comunalidade dada por h_i^2 que é indicado por:

$$h_i^2 = l_{i1}^2 + l_{i2}^2 + \dots + l_{im}^2 \quad (4.1)$$

A i -ésima comunalidade é a soma dos quadrados dos carregamentos da i -ésima variável sobre os m fatores comuns. O coeficiente da variância $\text{Var}(X_i)$ não pode passar de 1 em valor absoluto, caso contrário não tem solução própria.

A exigência que $m \leq p$ dado por m fatores e p variáveis implica que a estrutura dos dados (fatores comuns) não é mais complicado que os fatores observados, caso contrário não tem vantagem de usar Análise Fatorial.

2.2.7.3 Método de obtenção de fatores

Segundo FERREIRA (1996), existem dois métodos de estimação de parâmetros do modelo de fatores; o método das componentes principais. Neste método aplicado, a solução pode sofrer rotações, com o objetivo de simplificar as interpretações dos fatores.

2.2.7.3.1 Método das componentes principais

Segundo JOHNSON & WHICHERN (1998), a Análise Fatorial pela técnica da componente principal da matriz covariância amostral (S) é especificado em termos de pares de autovalores e autovetores $(\hat{\lambda}_1, \hat{e}_1), (\hat{\lambda}_2, \hat{e}_2), \dots, (\hat{\lambda}_p, \hat{e}_p)$ onde $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_p$. Seja $m < p$ o número de fatores comuns. A matriz dos pesos estimados (\hat{l}_{ij}) é dada por:

$$\hat{L} = \left[\sqrt{\hat{\lambda}_1} \hat{e}_1 : \sqrt{\hat{\lambda}_2} \hat{e}_2 : \dots : \sqrt{\hat{\lambda}_m} \hat{e}_m \right] \quad (4.2)$$

ou pode ser escrito na forma de: $\hat{L} = \hat{C} \hat{D}_\lambda^{1/2}$, onde :

$$\hat{C} = \begin{pmatrix} \hat{e}_{11} & \hat{e}_{12} & \dots & \hat{e}_{1p} \\ \hat{e}_{21} & \hat{e}_{22} & \dots & \hat{e}_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{e}_{p1} & \hat{e}_{p2} & \dots & \hat{e}_{pp} \end{pmatrix} \quad \hat{D}_\lambda^{1/2} = \begin{pmatrix} \sqrt{\hat{\lambda}_1} & & & \\ & \sqrt{\hat{\lambda}_{21}} & & \\ & & \ddots & \\ & & & \sqrt{\hat{\lambda}_p} \end{pmatrix}$$

\hat{C} = autovetores obtidos da matriz correlação.

$\hat{D}_\lambda^{1/2}$ = autovalores

Para aplicar o conjunto de dados $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n$ primeiro centralizar as observações subtraindo-se da média \underline{x} . As observações centralizadas são indicadas como:

$$\underline{x}_j - \bar{\underline{x}} = \begin{pmatrix} x_{ij} - \bar{x}_1 \\ x_{ij} - \bar{x}_2 \\ \vdots \\ x_{pj} - \bar{x}_p \end{pmatrix} \quad \text{para } j = 1, 2, \dots, n \quad (4.3)$$

É uma matriz covariância amostral (S) das observações originais. Quando as variáveis têm escalas diferentes, é usual trabalhar com as variáveis padronizadas:

$$z_i = \begin{bmatrix} \frac{x_{j1} - \bar{x}_1}{\sqrt{s_{11}}} \\ \frac{x_{j2} - \bar{x}_2}{\sqrt{s_{22}}} \\ \vdots \\ \frac{x_{jp} - \bar{x}_p}{\sqrt{s_{pp}}} \end{bmatrix} \quad \text{para } j = 1, 2, \dots, n$$

Como consequência a matriz covariância amostral S é a matriz correlação amostral R .

As variáveis específicas estimadas são dadas pelos elementos diagonais da matriz $\Psi = S - \hat{L}\hat{L}'$, assim:

$$\hat{\Psi} = \begin{bmatrix} \hat{\psi}_1 & 0 & \dots & 0 \\ \vdots & \hat{\psi}_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \hat{\psi}_p \end{bmatrix} \quad \text{com } \hat{\psi}_{ii} = s_{ii} - \sum_{j=1}^m \hat{l}_{ij}^2 \quad (4.4)$$

As comunalidades são estimadas como:

$$\hat{h}_i^2 = \hat{l}_{i1}^2 + \hat{l}_{i2}^2 + \dots + \hat{l}_{im}^2 \quad (4.5)$$

A matriz residual é dada por:

$$S - (\hat{L}\hat{L}' + \hat{\psi}) \quad (4.6)$$

Resulta em uma aproximação da matriz covariância S da solução das componentes principais. Os elementos da diagonal é zero e os outros elementos da matriz são pequenos, onde subjetivamente os m fatores do modelo são apropriados.

Segundo BARROSO (2003), outra maneira de avaliar o modelo é que na matriz dos resíduos tenha valores pequenos perto de 15%, indicando assim um ajuste razoável e mediano.

2.2.7.3.2 Rotação de Fatores

Em JOHNSON & WHICHERN (1998), a rotação de fatores permite obter estrutura para os pesos tal que cada variável tenha um peso alto em único fator e pesos baixos ou moderados nos demais fatores. Nem sempre é possível obter uma estrutura mais simples. Pode-se conciliar o método analítico e o método gráfico para determinar uma rotação ortogonal de uma estrutura simples. Para $m = 2$ os fatores comuns são considerados dois a dois ao mesmo tempo, e a estrutura simples pode ser determinada graficamente. Os eixos coordenados são rotacionados sob um ângulo θ . Os novos pesos rotacionados $\hat{\lambda}_{ij}^*$ são determinados pela relação:

$$\hat{L}^* = \hat{L}T \quad (4.7)$$

Onde:

$$T = \begin{pmatrix} \cos\theta & \text{sen}\theta \\ -\text{sen}\theta & \cos\theta \end{pmatrix} \text{ (sentido horário)} \quad T = \begin{pmatrix} \cos\theta & -\text{sen}\theta \\ \text{sen}\theta & \cos\theta \end{pmatrix} \text{ (sentido anti - horário)}$$

Para dimensões de ordem 2 podem ser analisadas graficamente e identificados os fatores comuns. Quando $m > 2$, a análise é possível apenas com programas computacionais para efetuar a rotação.

2.2.7.3.2.1 Rotação varimax

Uma medida analítica de uma estrutura simples foi sugerida por Kaiser conhecida como rotação varimax (ou normal varimax). Define $\hat{l}_{ij}^* = \frac{\hat{l}_{ij}}{\hat{h}_i}$ para serem coeficientes finais rotacionados escalonados pela raiz quadrada das comunalidades. O procedimento normal varimax seleciona a transformação ortogonal T que torna

$$V = \frac{1}{p} \sum_{j=1}^m \left[\sum_{i=1}^p \hat{l}_{ij}^{*4} - \left(\frac{\sum_{i=1}^p \hat{l}_{ij}^{*2}}{p} \right)^2 \right] \quad (4.8)$$

o maior possível. A magnitude de rotação dos coeficientes \hat{l}_{ij}^* tem efeito nas variáveis das comunalidades relativas com pesos pequenos de uma estrutura simples. Depois que a transformação T é determinada, os carregamentos de \hat{l}_{ij}^* são multiplicados por \hat{h}_i e as comunalidades das variáveis originais são preservadas. Efetivamente, maximizando V que corresponde para um maior valor quadrado dos carregamentos de cada um dos fatores sendo este o maior possível. Espera-se encontrar nos grandes grupos e coeficientes insignificantes nas colunas da rotação dos carregamentos da matriz \hat{L}^* .

2.2.7.3.4 Escolha do número de fatores

Para BARROSO (2003), na análise fatorial a escolha do número de fatores tem grande importância para se fazer uma interpretação. Caso o pesquisador faça uma opção por um número reduzido de fatores terá dificuldades na análise dos dados, podendo com isso perder identificações importantes na estrutura, caso contrário com um número excessivo de fatores também terá problemas de interpretação. Existem alguns critérios para determinação do número de fatores tais como:

- Critério de Kaiser
- Critério da porcentagem da variância explicada

- Critério *scree test*
- Métodos inferenciais

O método da escolha descreve alguns critérios, sendo de caráter de indicação e não de hierarquia:

a) Critério de Kaiser

É um critério desenvolvido por Kaiser (1958) em caso de dados padronizados, o número de fatores deve ser igual ao número de autovalores maiores ou iguais a um. O valor 1 corresponde à variância de cada variável padronizada, pois esse critério descarta fatores que possuam grau com explicação inferior ao de uma variável isolada. Se a análise for feita em cima da matriz covariância, sugere-se em vez de 1, que se faça como ponto de corte a média das variâncias das variáveis analisadas, BARROSO (2003).

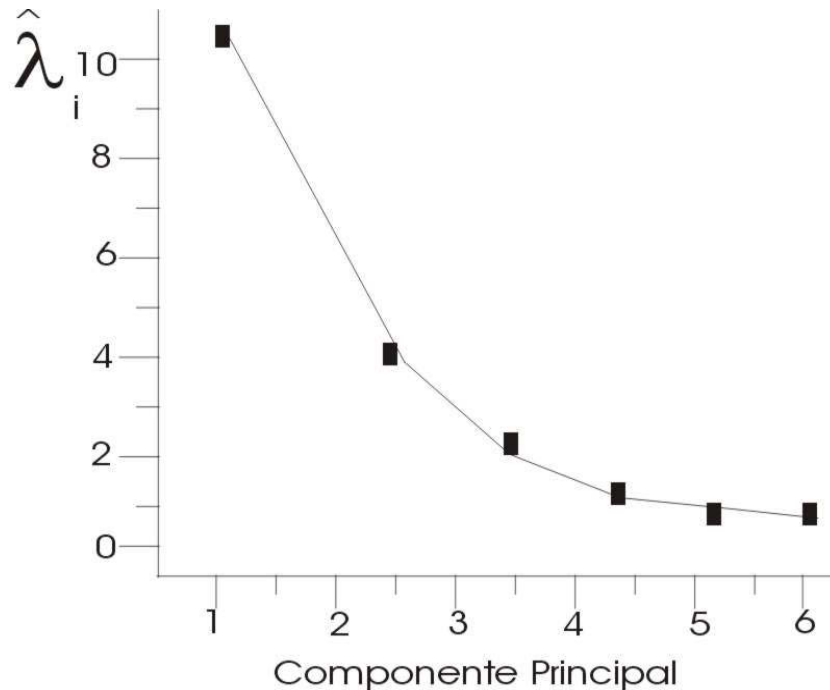
b) Critério da Porcentagem da Variância Explicada

Determina-se pelo conjunto de fatores comuns que explique uma certa porcentagem pré-definida da variabilidade global, pode-se adotar 70%, BARROSO (2003).

c) Critério *scree test*

Para FERREIRA (1996), *scree plot* é chamado de “gráfico de cotovelos”. Na figura 1 representam-se graficamente os autovalores e as componentes principais:

FIGURA 1 - DETERMINAÇÃO DO NÚMERO APROPRIADO DE COMPONENTES A SEREM RETIDOS



Para BARROSO (2003), é comum que a diferença de explicação entre os primeiros fatores de uma Análise fatorial seja grande e que tenda a diminuir com o aumento no número de fatores. Por este critério, o número ótimo de fatores é obtido quando a variação da explicação entre fatores consecutivos passa a ser pequena”.

Em FERREIRA (1996), os componentes no gráfico têm aproximadamente a mesma magnitude e são relativamente pequenos, com isso é suficiente resumir a variação amostral total.

d) Métodos Inferenciais

Em BARROSO (2003), este método não é adequado para dados que não são normais. É utilizado como um método indicativo, sendo que certas significâncias obtidas não podem ser interpretadas ao “pé da letra”. Destaca-se o método de Bartlett que verifica o modelo de análise fatorial estimado pelo método da máxima verossimilhança.

2.2.7.4 Algumas conclusões sobre análise fatorial

Na obtenção de uma Análise Fatorial com um mínimo de qualidade é necessário que o tamanho da amostra seja relativamente grande comparando com o número de variáveis envolvidas. Na literatura estatística exige-se para uma escolha da amostra em ordem de 20 vezes o número de variáveis envolvidas, outra sugere que seja no mínimo 5 vezes o número de variáveis ou que seja no mínimo 100 observações. HAIR (1998), enfatiza que ela não deve ser utilizada em amostras inferiores a 50 observações.

Na análise fatorial o sucesso dependerá do pesquisador, caso seja uma redução de variáveis que possa explicar uma boa parte considerada da variabilidade do conjunto original de variáveis. Há algumas propriedades que interessam a uma solução da análise fatorial:

- a) encontrar um número relativamente pequeno de fatores que possuam um alto grau de explicação da variabilidade original dos dados;
- b) encontrar fatores interpretáveis.

Tem algumas razões ainda que resultam no insucesso da análise fatorial como:

- i) tamanho insuficiente da amostra, ou seja, uma amostra pequena pode não conseguir refletir de maneira precisa a estrutura de interdependência dos dados;
- ii) variáveis com fraca interdependência: considere uma escala composta por itens, onde cada item mede um aspecto diferente do *constructo* de interesse, nesse caso é possível que uma análise fatorial não consiga identificar fatores com um grau razoável de interpretação.
- iii) a estrutura de dependência pode não ser homogênea em toda a amostra, considere como ilustração, itens de uma escala que se associam diferentemente (possuem estruturas de dependência diferentes) para homens e mulheres, nesse caso, uma análise fatorial aplicada apenas a um dos sexos pode ser bem sucedida, mas aplicada à amostra total não. Parece razoável que, no caso de insucesso e quando existirem razões teóricas para isso, se faça uma análise fatorial para cada subgrupo de interesse de uma amostra.

2.2.8 Análise Discriminante

2.2.8.1 Introdução

Análise Discriminante é uma técnica de análise multivariada frequentemente utilizada com o objetivo de diferenciar populações e ou classificar objetos em populações pré-definidas, BARROSO (2003).

A Análise Discriminante e de classificação é uma técnica multivariada que faz a separação de uma coleção de objetos distintos e que aloca novos objetos em grupos previamente definidos.

Para BARROSO (2003), os objetivos principais da análise discriminante são:

- Discriminação: tem o objetivo de encontrar funções das variáveis observadas (funções discriminantes) que são responsáveis ou que possam explicar as diferenças entre as g populações;
- Classificação ou alocação: tem o objetivo de determinar funções das variáveis observadas que sejam possíveis para classificar novos objetos em uma das g populações.

2.2.8.2 Análise discriminante para mais de duas populações

BARROSO (2003), classifica a população quando tem envolvimento sendo maior que dois. Seja g o número de populações, sendo dadas por $\tau_1, \tau_2, \dots, \tau_g$ e considere os vetores de médias das g populações denotadas por $\underline{\mu}_1, \underline{\mu}_2, \dots, \underline{\mu}_g$ e as matrizes de covariância por $\Sigma_1, \Sigma_2, \dots, \Sigma_g$. Esta classificação da população é tratada pelo Método de Fisher.

2.2.8.3 Método de Fisher

Segundo BARROSO (2003), na aplicação do Método de Fisher não é necessário que o vetor de variável aleatório \underline{x} possa vir de uma população com distribuição normal multivariada. Mas supõe-se que as matrizes de covariância sejam iguais

$\Sigma_1 = \Sigma_2 = \dots = \Sigma_g = \Sigma$. Seja $\underline{\mu}$ o vetor de médias das médias das g populações e B_0 a soma de produtos cruzados das g populações, isto é,

$$\underline{\mu} = \frac{1}{g} \sum_{i=1}^g \underline{\mu}_i \text{ e } B_0 = \sum_{i=1}^g (\underline{\mu}_i - \underline{\mu})(\underline{\mu}_i - \underline{\mu})^T \quad (4.9)$$

Quando os vetores de médias forem iguais, portanto não há diferença entre as populações e $B_0 = 0$. Seja uma combinação das variáveis $Y = \underline{1}^T \underline{x}$, a média e a variância de Y para i -ésima população são dadas por :

$$\mu_{iy} = E(Y / \tau_i) = \underline{1}^T E(\underline{x} / \tau_i) = \underline{1}^T \underline{\mu}_i \quad (5.0)$$

$$\text{Var}(Y) = \text{Var}(\underline{1}^T \underline{x}) = \underline{1}^T \text{Var}(\underline{x}) \underline{1} = \underline{1}^T \Sigma \underline{1} \quad (5.1)$$

A média em Y das médias das g populações, é dada por:

$$\bar{\mu}_y = \frac{1}{g} \sum_{i=1}^g \mu_{iy} = \frac{1}{g} \sum_{i=1}^g \underline{1}^T \underline{\mu}_i = \underline{1}^T \left(\frac{1}{g} \sum_{i=1}^g \underline{\mu}_i \right) = \underline{1}^T \bar{\underline{\mu}} \quad (5.2)$$

Onde $\bar{\mu}_y$ é a média global univariada e $\underline{1}^T \bar{\underline{\mu}}$ é a média global multivariada.

Para obter combinações lineares que discriminam as g populações, e que maximiza tem-se a seguinte razão:

$$\frac{\left(\text{soma das distâncias ao quadrado entre as médias} \right)}{\left(\text{em Y de cada população e a média global de Y} \right)} = \frac{\sum_{i=1}^g (\bar{\mu}_{iy} - \bar{\mu}_y)^2}{\sigma_y^2} = \frac{\sum_{i=1}^g (\mathbf{1}^T \underline{\mu}_i - \mathbf{1}^T \bar{\underline{\mu}})^2}{\mathbf{1}^T \underline{\Sigma} \mathbf{1}} =$$

$$= \frac{\mathbf{1}^T \left[\sum_{i=1}^g (\mathbf{1}^T \underline{\mu}_i - \underline{\mu})(\underline{\mu}_i - \bar{\underline{\mu}})^T \right] \mathbf{1}}{\mathbf{1}^T \underline{\Sigma} \mathbf{1}} = \frac{\mathbf{1}^T \mathbf{B}_0 \mathbf{1}}{\mathbf{1}^T \underline{\Sigma} \mathbf{1}} \quad (5.3)$$

Os coeficientes da primeira função discriminante são os elementos do autovetor padronizado de $\Sigma^{-1} \mathbf{B}_0$, associado ao maior autovalor da matriz. O vetor $\mathbf{1}$ que minimiza a razão sujeito à restrição $\text{Cov}(\mathbf{1}_1^T \underline{\mathbf{x}}, \mathbf{1}_2^T \underline{\mathbf{x}}) = 0$ é o autovetor padronizado ao segundo autovalor de $\Sigma^{-1} \mathbf{B}_0$ e a combinação linear que resulta é chamada de segunda função discriminante. Portanto, a k -ésima função discriminante é $\mathbf{1}_k^T \underline{\mathbf{x}}$, onde $\mathbf{1}_k$ corresponde k -ésimo autovetor padronizado sujeito à condição $\text{Cov}(\mathbf{1}_i^T \underline{\mathbf{x}}, \mathbf{1}_j^T \underline{\mathbf{x}}) = 0$, onde $i < k$. Pela decomposição espectral de Σ , temos que $\Sigma = \Gamma^T \Lambda \Gamma$, onde Λ é a matriz diagonal dos autovalores positivos. Seja $\underline{\mathbf{a}}$ o vetor $\Sigma^{1/2} \mathbf{1}$, então:

$$\underline{\mathbf{a}}^T \underline{\mathbf{a}} = \mathbf{1}^T \Sigma^{1/2} \Sigma^{1/2} \mathbf{1} = \mathbf{1}^T \underline{\Sigma} \mathbf{1} \quad (5.4)$$

é o denominador da razão que maximiza, e

$$\underline{\mathbf{a}}^T \Sigma^{-1/2} \mathbf{B}_0 \Sigma^{-1/2} \underline{\mathbf{a}} = \mathbf{1}^T \Sigma^{1/2} \Sigma^{-1/2} \mathbf{B}_0 \Sigma^{-1/2} \Sigma^{1/2} \mathbf{1} = \mathbf{1}^T \mathbf{B}_0 \mathbf{1} \quad (5.5)$$

é o numerador. Para maximizar (5.3) é mesmo que maximizar:

$$\frac{\underline{\alpha}^T \Sigma^{-1/2} B_0 \Sigma^{-1/2} \underline{\alpha}}{\underline{\alpha}^T \underline{\alpha}} \text{ para todo } \underline{\alpha} \quad (5.6)$$

O valor máximo de (5.6) é o maior autovalor de $\Sigma^{-1/2} B_0 \Sigma^{-1/2}$, e ocorre quando α é o autovalor padronizado. Nota-se que as matrizes $\Sigma^{-1} B_0$ são iguais a Σ^{-1} vezes os autovetores de $\Sigma^{-1/2} B_0 \Sigma^{-1/2}$.

O número de funções discriminantes de (s) é igual ao número de autovalores não nulos e é o máximo igual ao menor valor entre p, o número de variáveis observadas e g-1, o número de populações menos 1. Seja o vetor \underline{y} de dimensão s por 1 cujos elementos são funções discriminantes e $\underline{\bar{\mu}}_{iy}$ que corresponde o vetor de médias n a i-ésima população. Para

classificar tem uma regra que consiste em alocar x na população Γ_k se a distância ao quadrado entre \underline{y} e $\underline{\bar{\mu}}_{iy}$ for menor que a mesma distância entre \underline{y} e $\underline{\bar{\mu}}_{iy}$, para todo $i \neq k$, ou seja,

alocar \underline{x} em Γ_k se :

$$\sum_{j=1}^s (\underline{y}_j - \underline{\bar{\mu}}_k \underline{y}_j)^2 = \sum_{j=1}^s \left[\mathbf{I}_j^T \left(\underline{x} - \underline{\bar{\mu}}_k \right) \right]^2 \leq \sum_{j=1}^s \left[\mathbf{I}_{j=1}^T \left(\underline{x} - \underline{\bar{\mu}}_k \right) \right]^2 \quad (5.7)$$

para todo $i \neq k$, onde \mathbf{I}_j é i-ésimo autovetor padronizado de $\Sigma^{-1} B_0$. Os $\underline{\bar{\mu}}_i$ e Σ são desconhecidos, são substituídos por suas estimativas amostrais obtidas de $\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \dots, \bar{\mathbf{x}}_g$ e S_p ,

que é a combinação de S_1, S_2, \dots, S_p e são dados por:

$$\bar{\mathbf{x}}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}, \quad i = 1, 2, \dots, g \quad (5.8)$$

$$S_i = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} \left(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i \right) \left(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i \right)^T, \quad i = 1, 2, \dots, g \quad (5.9)$$

$$\bar{\mathbf{x}} = \frac{\sum_{i=1}^g n_i \bar{\mathbf{x}}_i}{\sum_{i=1}^g n_i} = \frac{\sum_{i=1}^g \sum_{j=1}^{n_i} \mathbf{x}_{ij}}{\sum_{i=1}^g n_i} \quad (6.0)$$

$$\hat{B}_0 = \sum (\bar{\mathbf{x}}_i - \bar{\mathbf{x}})(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})^T \quad (6.1)$$

$$W = \sum_{i=1}^g (n_i - 1) S_i = \sum_{i=1}^g \sum_{j=1}^{n_i} (\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)^T \quad (6.2)$$

$$S_p = \frac{1}{n_1 + n_2 + \dots + n_{g-G}} W \quad (6.3)$$

Que é estimador de Σ . Não é necessário utilizar todas as (s) funções discriminantes, somente as primeiras delas podem ser usadas para alocar os novos objetos nas g populações.

2.2.8.4 Problema geral de classificação

A função densidade de probabilidade $f_i(x)$ é associada á população Γ_i , onde $i = 1, 2, \dots, g$ onde:

- p_i : a probabilidade da observação pertencer à população Γ_i , $i = 1, 2, \dots, g$;
- $c(k/i)$: custo de classificação de um objeto de Γ_i em Γ_k (para $k = i$, $c(i/i) = 0$, $k, i = 1, 2, \dots, g$;

- R_k : conjunto dos x classificados em Γ_k ;
- $P(k/i)$ = probabilidade de se classificar um objeto em Γ_k quando na verdade ele é de Γ_i .
-

$$P(k/i) = \int_{R_k} f_i(x) dx \quad (6.4)$$

Para $k, i = 1, 2, \dots, g$.

Para cada custo esperado de erro de classificação (CEEC) x de Γ_1 em $\Gamma_2, \Gamma_3 \dots$ ou Γ_g , é:

$$\text{CEEC}(1) = P(2/1)c(2/1) + P(3/1)c(3/1) + \dots + P(g/1)c(g/1) = \sum_{k=2}^g P(k/1)c(k/1) \quad (6.5)$$

Quando o custo de erro de classificação ocorre com probabilidade p_1 , então o custo esperado de erro classificatório é dado por:

$$\text{CEEC} = p_1 \text{CEEC}(1) + p_2 \text{CEEC}(2) + \dots + p_g \text{CEEC}(g) =$$

$$p_1 \left(\sum_{k=2}^g (P(k/1)c(k/1)) \right) + \dots + p_g \left(\sum_{k=1}^{g-1} (P(k/g)c(k/g)) \right) = \sum_{i=1}^g p_i \left(\sum_{k=1, k \neq i}^g (P(k/i)c(k/i)) \right) \quad (6.6)$$

Para a regra de classificação determina-se R_1, R_2, \dots, R_n que torne mínimo o custo do erro encontrado de classificação. É um procedimento que aloca x na população $\Gamma_k, k = 1, 2, \dots, g$ para o qual:

$$\sum_{i=1, i \neq k}^g (p_i f_i(x) c(k/i)) \quad (6.7)$$

seja menor. Caso ocorra um empate, x pode ser classificado em qualquer uma das populações. Quando todos os custos $c(k/i)$, $k, i = 1, 2, \dots, g$ são iguais, aloca-se x na população Γ_k em que :

$$\sum_{i=1, i \neq k}^g p_i f_i(x) \quad (6.8)$$

também seja menor, e que o termo excluído $p_k f_k(x)$ maior. A regra para classificar o custo do erro encontrado de classificação mínimo com custos iguais por falhas na classificação é: alocar x em Γ_k se $p_k f_k(x) > p_i f_i(x)$ para todo $i \neq k$, ou alocar x em Γ_k se $\ln(p_k f_k(x)) > \ln(p_i f_i(x))$ para todo $i \neq k$.

2.2.8.5 Classificação para populações normais

Conforme BARROSO(2003), os vetores de variáveis aleatórias \underline{x} de todas as populações têm distribuição normal multivariada com parâmetros $\underline{\mu}_i$ e Σ_i , $i = 1, 2, \dots, g$, ou seja:

$$f_i(x) = \frac{1}{(2\pi)^{p/2} |\Sigma_i|^{1/2}} \exp\left\{-\frac{1}{2}(\underline{x} - \underline{\mu}_i)^T \Sigma_i^{-1} (\underline{x} - \underline{\mu}_i)\right\}, i = 1, 2, \dots, g \quad (6.9)$$

e tem-se ainda $c(i/i)=0$ e $c(k/i)=1$, $k \neq i$, $k, i = 1, 2, \dots, g$ tem-se a seguinte regra para alocar x em Γ_k se:

$$\ln p_k f_k(x) = \ln p_k - \frac{p}{2} \ln(2\pi) - \frac{1}{2} \ln |\Sigma_k| - \frac{1}{2} (\underline{x} - \underline{\mu}_k)^T \Sigma_k^{-1} (\underline{x} - \underline{\mu}_k) = \max_i \ln p_i f_i(x) \quad (7.0)$$

A constante $\frac{p}{2} \ln(2\pi)$ pode ser ignorada, pois é a mesma para todas as populações.

Define-se o escore quadrático de classificação $Q_i(x)$ para toda população r_i como sendo:

$$Q_i(x) = \frac{-1}{2} \ln |S_i| - \frac{1}{2} (\underline{x} - \underline{\bar{x}}_i)^T S_i^{-1} (\underline{x} - \underline{\bar{x}}_i) + \ln p_i \quad (7.1)$$

Para várias populações normais, a regra de classificação consiste em alocar \underline{x} em r_k se $Q_k(x) = \max_i Q_i(x)$ para $i = 1, 2, \dots, g$. Se as matrizes de covariância das populações são iguais, os termos que dependem de Σ_i e não de $\underline{\mu}_i$ e são constantes para as g populações podem ser ignorados. O escore de classificação é linear e é dado por:

$$\underline{l}_i(x) = \underline{\bar{x}}_i^T S_p^{-1} x - \frac{1}{2} \underline{\bar{x}}_i^T S_p^{-1} \underline{\bar{x}}_i + \ln p_i \quad (7.2)$$

E a regra de classificação consiste em alocar x em r_k se $\underline{l}_k(x) = \max_i \underline{l}_i(x)$ para $i = 1, 2, \dots, g$.

2.2.8.6 Avaliação da função de classificação

Conforme BARROSO (2003), para avaliar a classificação da população conhecida, é necessário verificar se os elementos da amostra foram alocados corretamente ou não. Para contornar as possibilidades de erros de alocação, pode-se usar um procedimento de validação cruzada (*cross-validation*), sugerido por LACHEMBRUCH & MICKEY (1968). Estes procedimentos consistem em:

- dividir a amostra em pequenos grupos;
- retirar o primeiro grupo da amostra;
- determinar os grupos restantes à classificação para serem alocados;
- devolver à amostra o primeiro grupo alocado;
- retirar o segundo grupo;

- seguir assim até que o último grupo seja retirado.

A taxa estimativa de erro é calculada pela “matriz de confusão”, dada por:

		Classificado em				total
		τ_1	τ_2	\cdots	τ_g	
População verdadeira	τ_1	n_{11}	n_{12}	\cdots	n_{1g}	n_1
	τ_2	n_{21}	n_{22}	\cdots	n_{2g}	n_2
	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
	τ_g	n_{g1}	n_{g2}	\cdots	n_{gg}	n_g
Total	\hat{n}_1	\hat{n}_2	\cdots	\hat{n}_g	N	

Onde:

n_{ij} = número de observações de τ_i classificados em τ_j ;

\hat{n}_i = número de observações classificadas em τ_i .

2.2.9 Análise de Agrupamentos

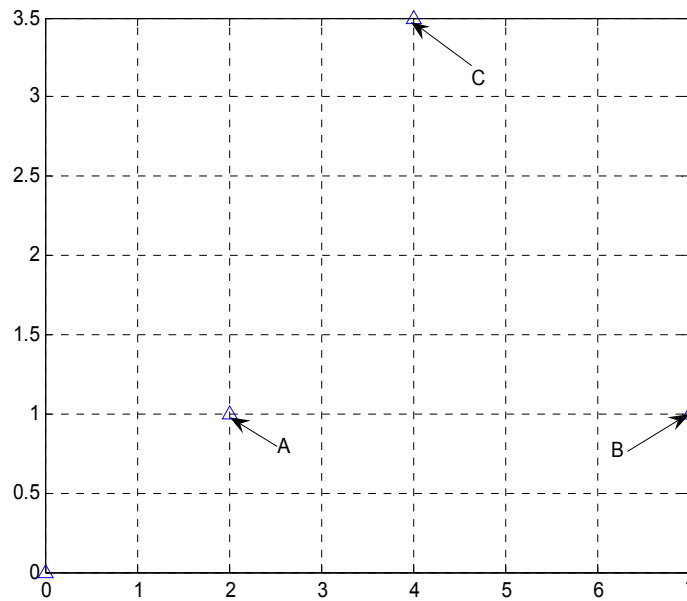
2.2.9.1 Introdução

Conforme BARROSO (2003), análise de agrupamentos é o nome dado a um conjunto de técnicas utilizadas na identificação de padrões de comportamentos em bancos de dados através da formação de grupos homogêneos de casos.

Para FERREIRA (1996), a análise de agrupamento não considera o número de grupos, é realizada com base na similaridade e dissimilaridade. Esta análise tem o objetivo de agrupar objetos semelhantes conforme suas características ou variáveis. Outro procedimento consiste em verificar se um indivíduo com as coordenadas A(2;1) é mais parecido com B(7;1) do que com C(4;3,5). Na figura 2, tem uma situação em A que é mais parecido com B do que com C.

Para fazer tal inferência estatística usou o conceito de distância euclidiana. Quando o número de variáveis é muito pequeno, é possível fazer uma inspeção visual para responder tal decisão.

FIGURA 2 – DISPERSÃO ENTRE TRÊS INDIVÍDUOS MENSURADOS COM RELAÇÃO A DUAS VARIÁVEIS QUANTITATIVAS CONTÍNUAS



2.2.9.2 Medidas de similaridade e dissimilaridade

Segundo BARROSO (2003), as medidas de similaridade e dissimilaridade tem um papel importante quando se agrupam através de algoritmos. Define-se critério para avaliação dos grupos se eles estão próximos ou não. Existem dois tipos de medidas:

- medida de similaridade: quanto maior o valor, maior a semelhança entre os objetos;
- medida de dissimilaridade: quanto maior o valor, mais diferentes são os objetos.

Em FERREIRA (1996), há uma necessidade de especificar um coeficiente que indique uma proximidade entre os indivíduos. Para todos os casos verifica-se a natureza das variáveis

se são discretas, contínuas ou binárias e a escala de medida se é nominal, ordinal, real ou razão.

Para BARROSO (2003), para definir o número de grupos é necessário que se tenha algum conhecimento sobre os dados como características ou pode se basear no resultado da análise. Quanto à formação de grupos pode-se definir o tipo de algoritmo utilizado na sua identificação. Na validação do agrupamento as variáveis possuem comportamento diferenciado nos diversos grupos, e para que possa aplicar técnicas inferenciais para compará-las. No final do processo para interpretar os grupos podem-se caracterizar os grupos formados usando estatísticas descritivas.

2.2.9.2.1 Distâncias

Conforme FERREIRA (1996), a maioria das técnicas multivariadas baseia-se no conceito de distâncias. Algumas distâncias são dadas pelas fórmulas a seguir:

2.2.9.2.1.1 Distância Euclidiana

Sejam dois vetores no espaço p-dimensional $\underline{\mathbf{X}}_1^T = [X_{11}, X_{12} \dots X_{1p}]$ e $\underline{\mathbf{X}}_2^T = [X_{21}, X_{22} \dots X_{2p}]$ observações entre dois objetos ou indivíduos, então a distância Euclidiana pode ser dada por:

$$d(\underline{\mathbf{X}}_1, \underline{\mathbf{X}}_2) = \sqrt{(X_{11} - X_{21})^2 + (X_{12} - X_{22})^2 + \dots + (X_{1p} - X_{2p})^2} \quad (7.3)$$

2.2.9.2.1.2 Distância Mahalanobis

Esta distância é dada pela fórmula:

$$d(\underline{\mathbf{X}}_1, \overline{\underline{\mathbf{X}}}_2) = \sqrt{(\underline{\mathbf{X}}_1 - \overline{\underline{\mathbf{X}}}_2)^T \mathbf{S}^{-1} (\underline{\mathbf{X}}_1 - \overline{\underline{\mathbf{X}}}_2)} \quad (7.4)$$

em que \mathbf{S}^{-1} é matriz inversa variância e covariância amostral.

2.2.9.2.1.3 Distância de Minkowski

Esta distância depende de funções modulares e é dado pela fórmula:

$$d(\underline{\mathbf{X}}_1, \underline{\mathbf{X}}_2) = \left[\sum_{i=1}^p |X_{1i} - X_{2i}|^m \right]^{1/m} \quad (7.5)$$

Quando $m = 1$ a equação (7.5) é conhecida por “métrica do quarteirão” ou métrica de *city-block* e para $m = 2$ é a mesma que distância Euclidiana, mas com variações nos valores de m podem causar trocas nos pesos dados com pequenas e grandes diferenças.

2.2.9.3 Agrupamentos

Em análise estatística multivariada segundo FERREIRA (1996), tem algumas técnicas de agrupamentos que são denominadas hierárquico e agrupamentos não hierárquico.

2.2.9.3.1 Agrupamentos hierárquicos

Em FERREIRA (1996), nestes agrupamentos hierárquicos são feitas infinitas divisões. Os objetos mais similares são agrupados, e o processo é repetido decrescendo a similaridade e os subgrupos são divididos até formar um único grupo.

Segundo BARROSO (2003), para formação destes grupos são utilizados alguns métodos hierárquicos de agrupamento:

- Método do Vizinho Mais Próximo: usa a menor distância entre um elemento de \underline{X}_1 e um elemento de \underline{X}_2 , ou seja, $d[\underline{X}_1, \underline{X}_2] = \min d_{ik}$ para $i \in \underline{X}_1$ e $k \in \underline{X}_2$.
- Método Do Vizinho Mais Longe: é definido pela maior distância entre um elemento de \underline{X}_1 e um elemento de \underline{X}_2 , ou pode-se escrever $d[\underline{X}_1, \underline{X}_2] = \max d_{ik}$ para $i \in \underline{X}_1$ e $k \in \underline{X}_2$.
- Método das Médias das Distâncias: calcula-se a média das distâncias entre os elementos de \underline{X}_1 e os de \underline{X}_2 , $d[\underline{X}_1, \underline{X}_2] = \frac{\sum \sum d_{ik}}{|\underline{X}_1| |\underline{X}_2|}$ para $i \in \underline{X}_1$ e $k \in \underline{X}_2$.
- Método do Centróide: define a coordenada de cada grupo como sendo a média das coordenadas de seus objetos. Uma vez obtida essa coordenada, denominada centróide, a distância entre os grupos é obtida através do cálculo das distâncias entre as centróides.
- Método de *Ward*: a alocação de um elemento a um grupo é feita de modo a minimizar uma medida de homogeneidade interna.

2.2.9.3.2 Correlação cofenética

Para BARROSO (2003), é uma medida de validação utilizada nos métodos de agrupamentos hierárquicos. Comparam-se as distâncias observadas entre os objetos e as distâncias previstas a partir de um processo de agrupamento. Para um bom agrupamento as distâncias previstas podem respeitar a ordem determinada pelas distâncias observadas, ou seja, quando duas observações estão próximas, espera-se que a distância prevista entre elas seja pequena. Em uma avaliação da ocorrência deste comportamento, define-se correlação cofenética como sendo a correlação entre as distâncias efetivamente observadas e as previstas. Por exemplo, quando houver uma correlação cofenética em torno de 0,95, indica-se um agrupamento de boa qualidade.

3 MATERIAL E MÉTODOS

3.1 CARACTERIZAÇÃO DA AMOSTRA E DAS VARIÁVEIS

O período de execução deste experimento agrônômico conforme Anexo I foi compreendido entre julho de 2003 e abril de 2004. O local do experimento foi a propriedade do Sr. José Antonio Bueno, localizada no município de Água Doce, SC, região denominada “Campos de Palmas” conforme figura 13. A área total destinada ao experimento foi aproximadamente 6,93 hectares dividida em 6 potreiros.

FIGURA 3 – LOCAÇÃO¹ DO EXPERIMENTO AGRONÔMICO



Fonte:

<http://www.brasounds.hpg.ig.com.br/foto/sul.jpg>

Para fins deste estudo foram compostas 11 matrizes de dados, onde as 5 primeiras populações conforme Anexos II até VI são resultados de análises de solos em períodos de 28 dias. As análises de rotinas de solos foram coletadas em três profundidades: 0-5 cm, 5-10cm e 10-20cm e os atributos químicos analisados foram: potencial de hidrogênio, matéria orgânica,

¹ As cidades de Palmas e Água Doce foram adaptadas no mapa da figura original a fim de facilitar a observação.

alumínio, hidrogênio mais alumínio (acidez total), cálcio, fósforo, potássio e saturação por bases, nitrato e amônia. Também foram avaliadas as seguintes variáveis referentes às populações 6 até 11 conforme anexos VII a XII:

- 1) variáveis derivadas de animais: ganho médio diário, produção animal e carga animal;
- 2) variáveis medidas na pastagem: taxa de lotação, massa de forragem, taxa de acúmulo.

Usando técnicas de Estatística Multivariada verifica-se que os dados amostrais referentes às populações de 1 até 11 são provenientes de uma distribuição normal multivariada. E também se utiliza o teste T^2 de Hotteling com a finalidade de verificar a igualdade das médias e a partir daí indicar quais os elementos que fazem a diferença significativa. Na Análise Fatorial é possível descrever fatores que explicam a variabilidade dos dados, correlação entre as variáveis e matriz dos resíduos. Nas Análises de Agrupamento e Fatorial é possível também comparar os grupos homogêneos quanto à correlação das variáveis, distância média e os pesos atribuídos nos fatores. Com Análise Discriminante verifica-se a alocação dos grupos no espaço discriminante dos resultados de solos e taxa aparente de erro.

Os dados amostrais das populações de 1 até 11 foram padronizados para as Análises Fatorial e Agrupamento.

4 RESULTADOS E ANÁLISE

4.1 DISTRIBUIÇÃO NORMAL MULTIVARIADA

Para verificar se os resultados das análises de solos das populações 1 até 5 obedecem a uma distribuição normal, utilizam-se os métodos gráficos de probabilidade, que julga a normalidade dos dados multivariados baseando-se no quadrado da distância generalizada conforme as figuras 4, 5, 6, 7 e 8. Estas verificações foram feitas no *software* MATLAB 7.0.

FIGURA 4 – VERIFICAÇÃO DA NORMALIDADE DA POPULAÇÃO 1

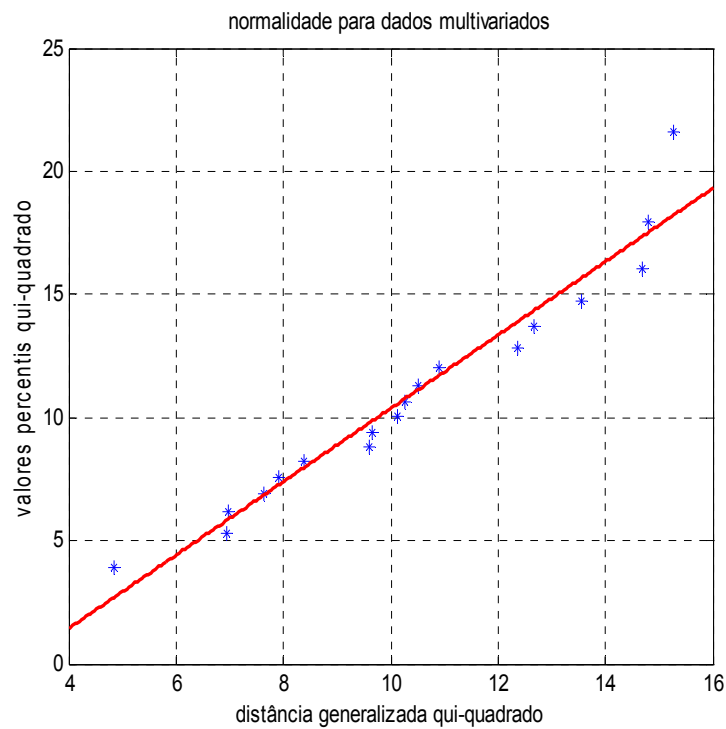


FIGURA 5 – VERIFICAÇÃO DA NORMALIDADE DA POPULAÇÃO 2

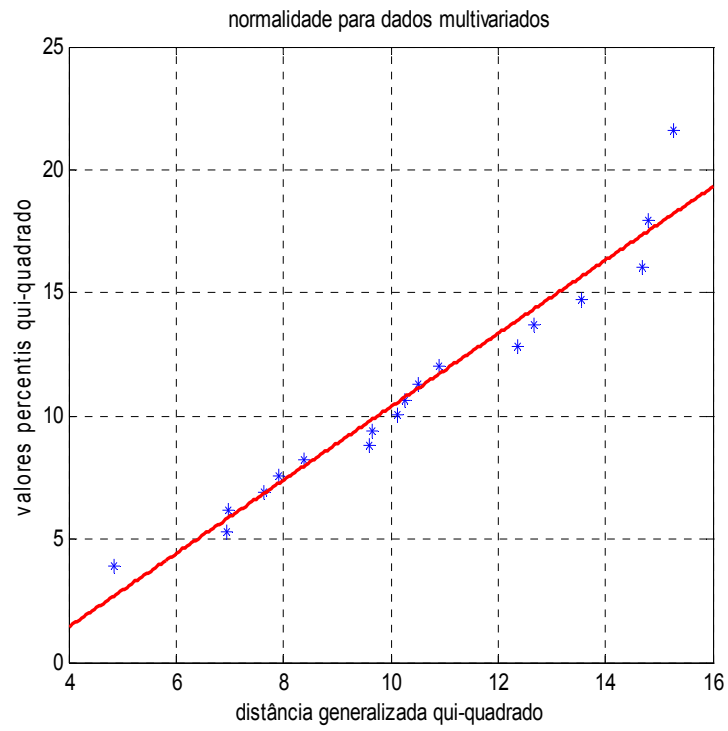


FIGURA 6 – VERIFICAÇÃO DA NORMALIDADE DA POPULAÇÃO 3

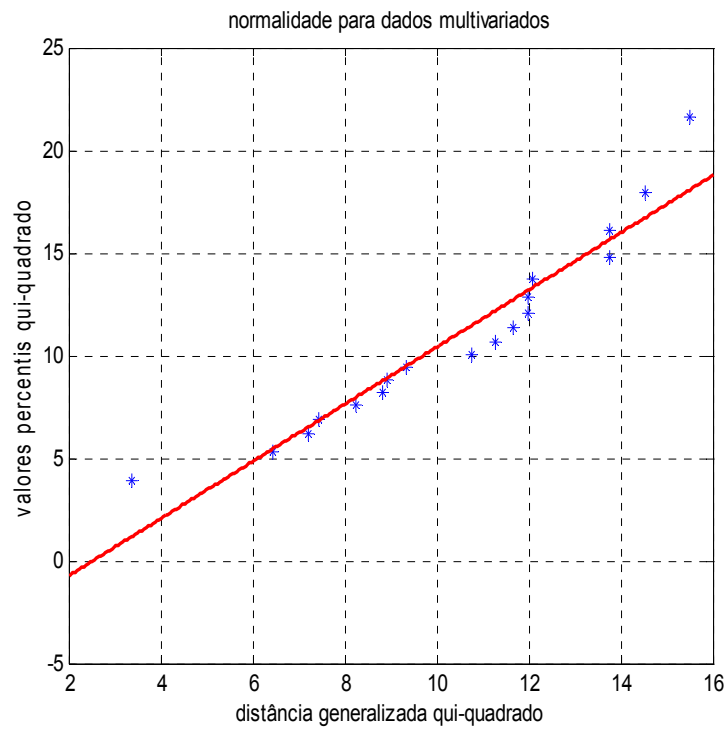


FIGURA 7 – VERIFICAÇÃO DA NORMALIDADE DA POPULAÇÃO 4

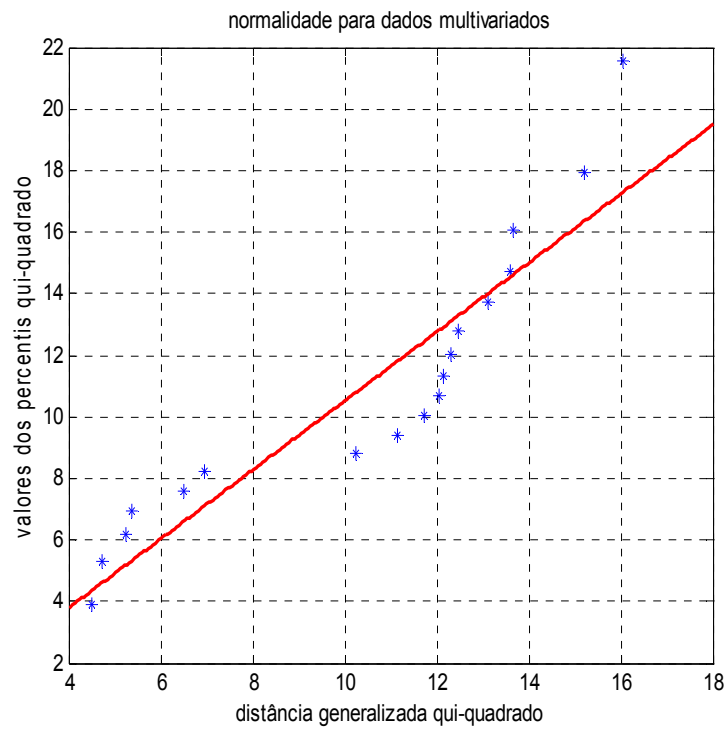
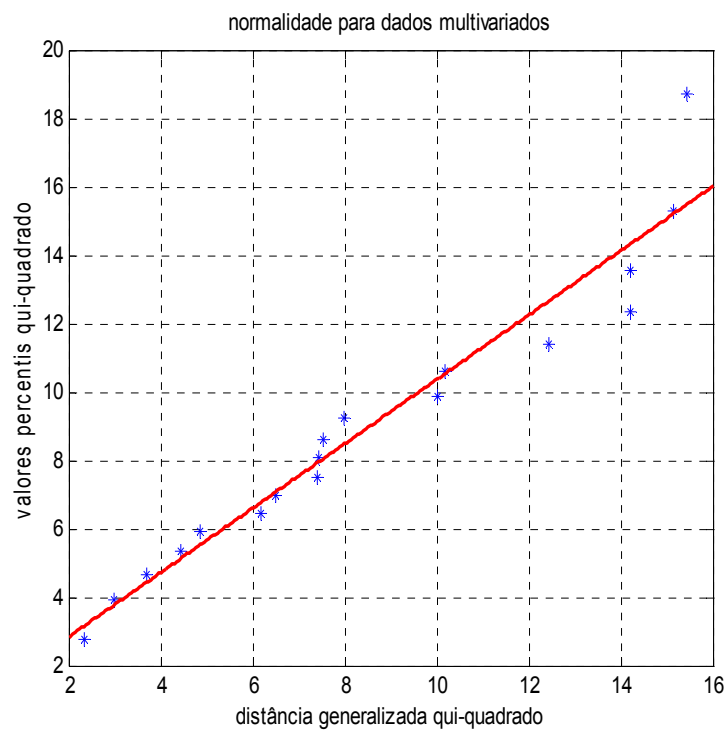


FIGURA 8 – VERIFICAÇÃO DA NORMALIDADE DA POPULAÇÃO 5



As populações de resultados de análises de solos dadas nas figuras de 4 a 8 aproximam-se de uma adequada e útil da verdadeira distribuição normal multivariada. As populações de 6 até 11 para verificar a normalidade das variáveis derivadas de animais e variáveis medidas na pastagem estão nas figuras seguintes:

FIGURA 9 – VERIFICAÇÃO DA NORMALIDADE DA POPULAÇÃO 6

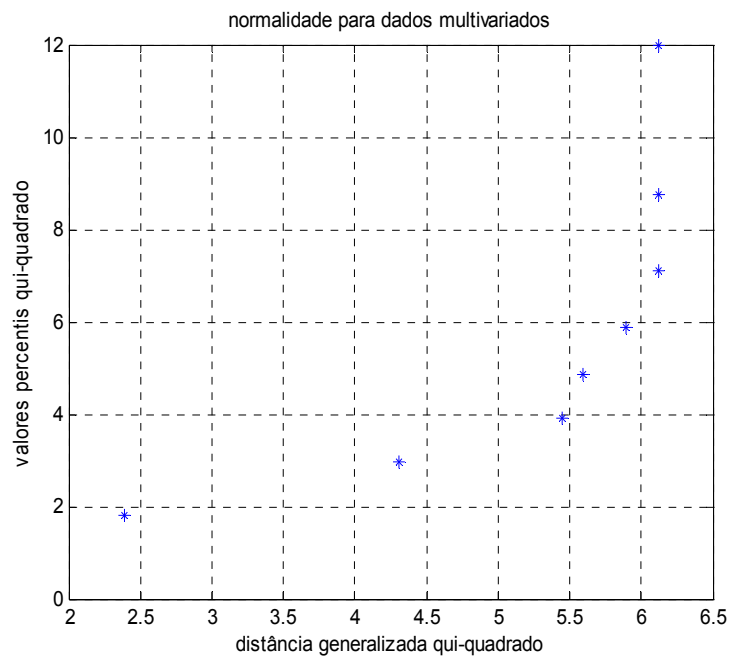


FIGURA 10 – VERIFICAÇÃO DA NORMALIDADE DA POPULAÇÃO 7

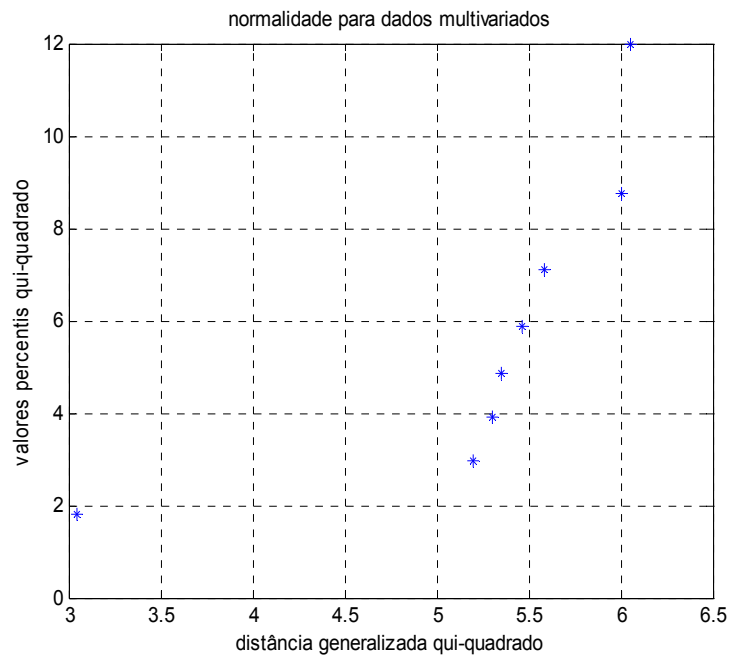


FIGURA 11 – VERIFICAÇÃO DA NORMALIDADE DA POPULAÇÃO 8

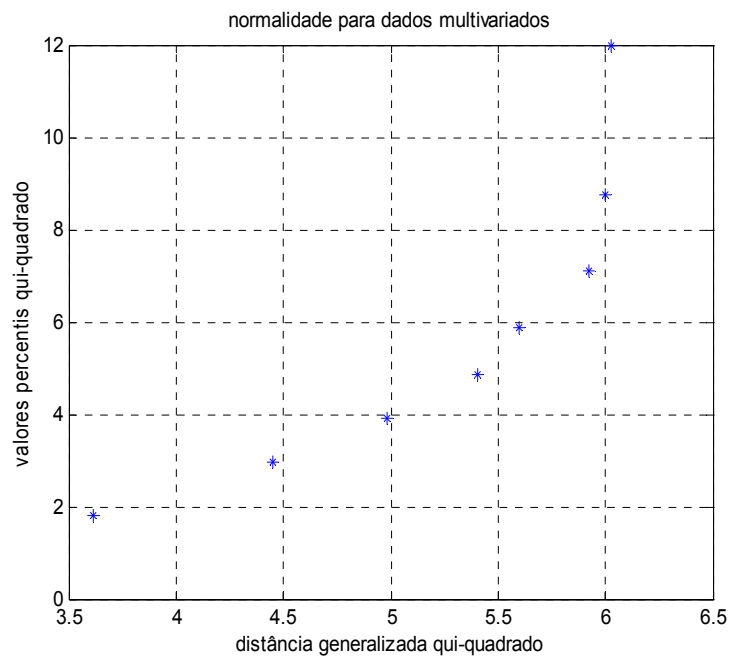


FIGURA 12 – VERIFICAÇÃO DA NORMALIDADE DA POPULAÇÃO 9

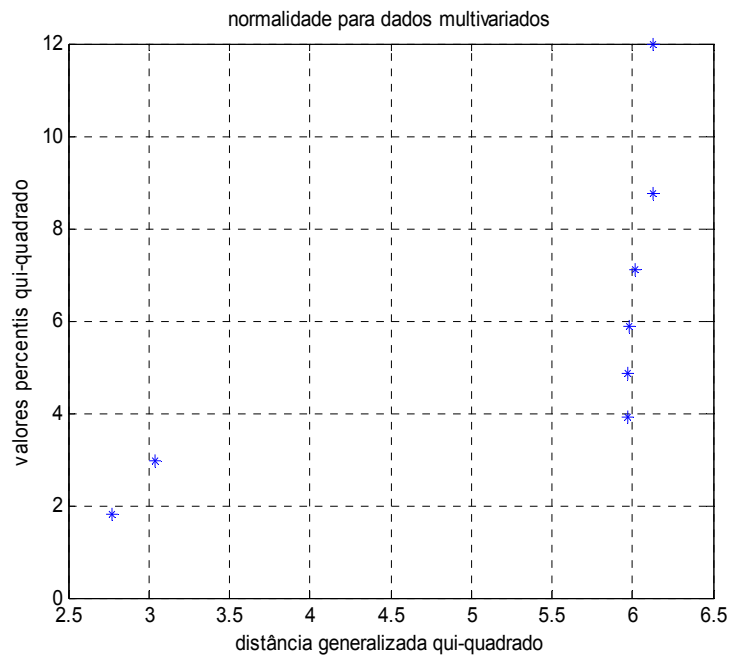


FIGURA 13 – VERIFICAÇÃO DA NORMALIDADE DA POPULAÇÃO 10

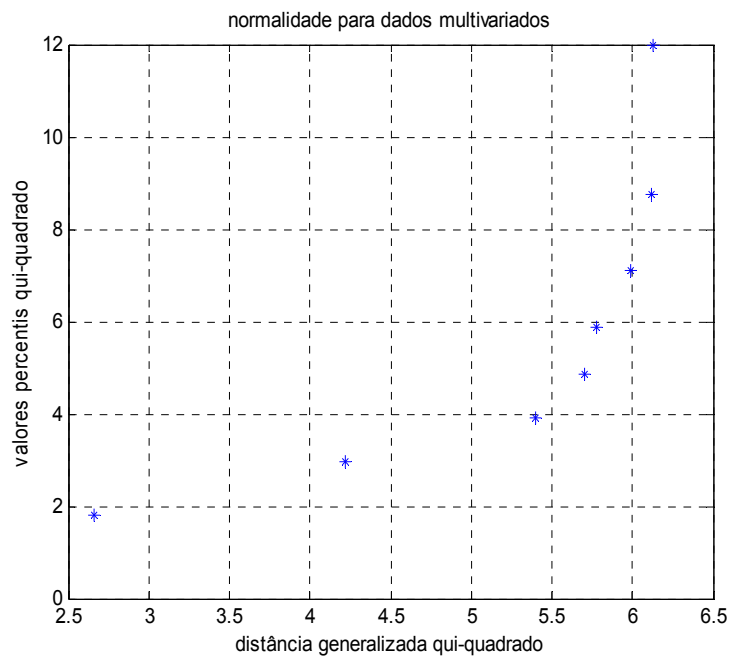
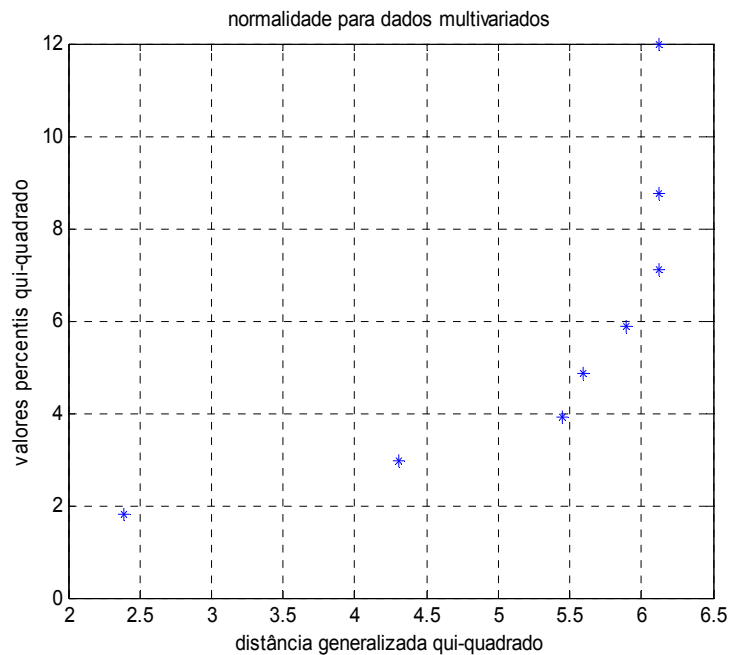


FIGURA 14 – VERIFICAÇÃO DA NORMALIDADE DA POPULAÇÃO 11



As populações 6 até 11 dadas nas figuras 9 até 14 não obedecem a uma distribuição normal multivariada.

4.2 COMPARAÇÃO ENTRE VETORES DE MÉDIAS DE VÁRIAS POPULAÇÕES

A comparação entre vetores de médias de várias populações foi programada no *software* MATLAB 7.0. A análise de variância multivariada (MANOVA) investigou que os vetores de médias das populações 1 até 4 possuem diferenças significativas com intervalo de confiança de 95%. A estatística do teste (qui-quadrado = 275.662) é maior que qui-quadrado teórico = $\text{qui}2(1-\alpha, p*(g-1)) = \text{qui}2(0,95; 11(4-1)) = 47.3999$, portanto rejeita-se que H_0 (que todas as médias são iguais) e se aceita que pelo menos uma das médias é diferente das demais.

Na tabela 3 estão os intervalos de confiança das populações 1 e 2 , 1 e 3 , 1 e 4:

TABELA 3 - INTERVALOS DE CONFIANÇA PARA OS EFEITOS DOS TRATAMENTOS DAS POPULAÇÕES 1 E 2; 1 E 3, 1 E 4.

Variáveis	Intervalos de confiança das Pop. 1 e Pop.2		Intervalos de confiança das Pop. 1 e Pop. 3		Intervalos de confiança das Pop. 1 Pop. 4	
	pH	-0,3457	0,2457	-0,1569	0,4345	-0,0235
MO	6,1620	31,7020	5,7898	31,3980	3,7798	29,3880
Al	-2,2739	0,8873	-3,0394	0,1218	-3,2006	-0,0394
H+Al	-1,7139	5,8273	-1,9722	5,5690	-3,8022	3,7390
Ca	-0,4579	3,8845	-0,4396	3,9028	-0,1679	4,1745
Mg	-0,0321	2,0575	0,2979	2,3875	0,3174	2,4070
K	0,0623	0,3009	-0,0077	0,2309	-0,0149	0,2237
P	-4,4624	12,6580	-7,6757	9,4447	-8,0385	9,0819
V	-1,5934	1,8070	-1,5735	1,8269	-1,5413	1,8591
NH ₄	62,0520	89,4746	25,5887	53,0113	36,5554	63,9780
NO ₃	-9,0567	14,0025	-5,7935	17,2747	-13,2591	9,8091

O intervalo de 95% de confiança das populações 1 até 4 onde não estão incluídos o zero na diferença entre as médias apresentam diferenças significativas são :

- matéria orgânica (MO) nas populações 1 e 2; 1 e 3, 1 e 4;
- magnésio (Mg) nas populações 1 e 3, 1 e 4;
- potássio (K) nas populações 1 e 2;
- amônia (NH₄) nas populações 1 e 2, 1 e 3, 1 e 4.

Os intervalos de confiança das populações 2 e 3, 2 e 4, 3 e 4 é dada na tabela 4:

TABELA 4 - INTERVALOS DE CONFIANÇA PARA OS EFEITOS DOS TRATAMENTOS DAS POPULAÇÕES 2 E 3, 2 E 4, 3 E 4.

Variáveis	Intervalos de confiança das Pop. 2 e Pop.3		Intervalos de confiança das Pop. 2 e Pop. 4		Intervalos de confiança das Pop. 3 e Pop. 4	
	pH	-0,1069	0,4845	0,0265	0,6179	-0,1623
MO	-13,1763	12,4319	-15,1863	10,4219	-14,8141	10,7941
Al	-2,3461	0,8151	-2,5073	0,6539	-1,7418	1,4194
H+Al	-4,0289	3,5123	-5,8589	1,6823	-5,6006	1,9406
Ca	-2,1529	2,1895	-1,8812	2,4612	-1,8995	2,4429
Mg	-0,7148	1,3748	-0,6953	1,3943	1,0253	1,0643
K	-0,1893	0,0493	-0,1965	0,0421	-0,1265	1,0643
P	-11,7735	5,3469	-12,1363	4,9841	-8,9230	8,1974
V	-1,6803	1,7201	-1,6481	1,7523	-1,6680	1,7324
NH ₄	-50,1746	-22,7520	-39,2079	-11,7853	-2,7446	24,6780
NO ₃	-8,2619	14,8063	-15,7275	7,3407	-18,9997	4,0685

Os intervalos 95% de confiança dados na tabela 4 a variável potencial de hidrogênio (pH) tem diferença significativa nas populações 2 e 4, a variável magnésio (Mg) apresenta diferença significativa nas populações 3 e 4 e a amônia (NH₄) nas populações 2 e 3, 2 e 4, pois não estão incluídos o zero na diferença de suas médias.

4.3 ANÁLISE FATORIAL E ANÁLISE DE AGRUPAMENTOS

Na tabela 5 apresentam-se os fatores da população 1 através da análise fatorial utilizando dois métodos: análise de componentes principais (ACP) e a rotação varimax:

TABELA 5 - FATORES DA ROTAÇÃO VARIMAX E O MÉTODO DAS COMPONENTES PRINCIPAIS DA POPULAÇÃO 1

Variáveis	Método das componentes		Rotação Varimax		Comunalidades
	Y ₁	Y ₂	F ₁	F ₂	
pH	0,9066	-0,0470	0,8890	-0,1840	0,8242
MO	0,1552	0,8694	0,2853	-0,8357	0,8850
Al	-0,9033	0,2627	-0,8530	-0,3967	0,7897
H+Al	-0,2755	0,8449	-0,1441	0,8769	0,9817
Ca	0,9739	0,1821	0,9903	0,0322	0,8908
Mg	0,9372	0,1118	0,9433	-0,0317	0,8697
K	0,8290	0,4272	0,8842	0,2964	0,9232
P	0,9570	0,0856	0,9589	-0,0606	0,9756
V	0,9865	0,0480	0,9824	-0,1023	0,6598
NH ₄	0,5064	-0,6351	0,4042	-0,7046	0,3742
NO ₃	0,6017	-0,1101	0,5780	-0,2001	
autovalores	6,7596	2,1940	6,6544	2,2991	
Proporção Acumulada da Variância	61,45%	81,40%	60,50%	81,40%	

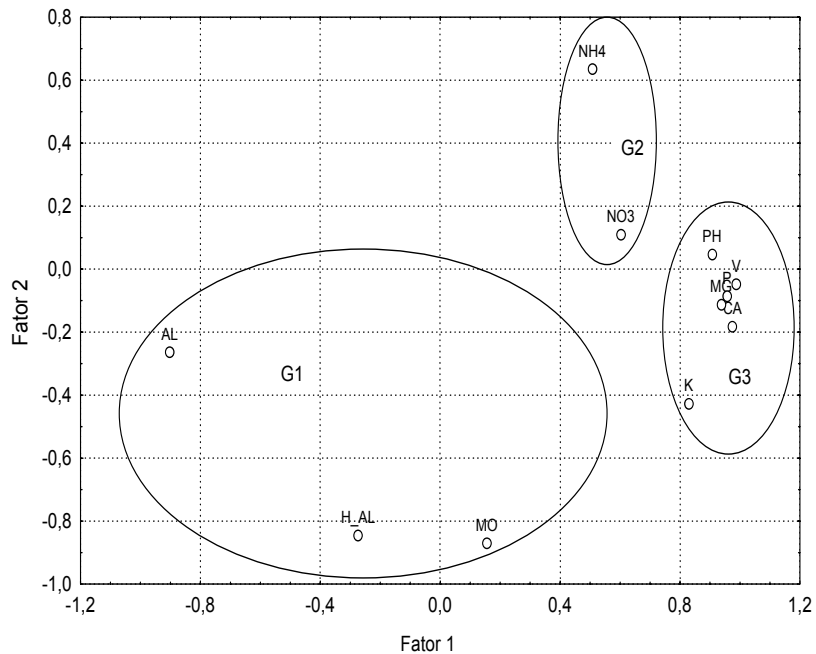
Conforme os dados obtidos na tabela 5 o modelo ajustado levou em consideração dois fatores na rotação varimax e duas componentes principais Y₁ e Y₂. No conjunto do modelo ajustado, explicaram 81,40% das variâncias das variáveis com autovalores maiores que 1.

O percentual das comunalidades de cada variável que é explicado pelos fatores em conjunto mostra que as variáveis que melhor foram explicadas pelos fatores (apresentaram maior comunalidade) foi H+Al que teve 98,17% e P(fósforo) 97,56% de suas variações captadas pelos fatores, e a de menor valor percentual foi NO₃ (nitrito), apresentando 37,42% de sua variação explicada pelos fatores estudados. As demais variáveis apresentaram comunalidades entre 65% e 92%, o que mostra que suas variações foram, de modo geral, captadas de forma satisfatória pelos fatores ajustados.

Cada variável dentro do seu fator correspondente possui uma alta correlação entre elas e entre os fatores as variáveis possuem uma correlação mais fraca. Esta correlação varia de -1 até 1, sendo que perto de zero não existe correlação, portanto quanto mais perto de 1 ou -1 melhor é a correlação. Na tabela 6 é dada a matriz de correlação da população 1.

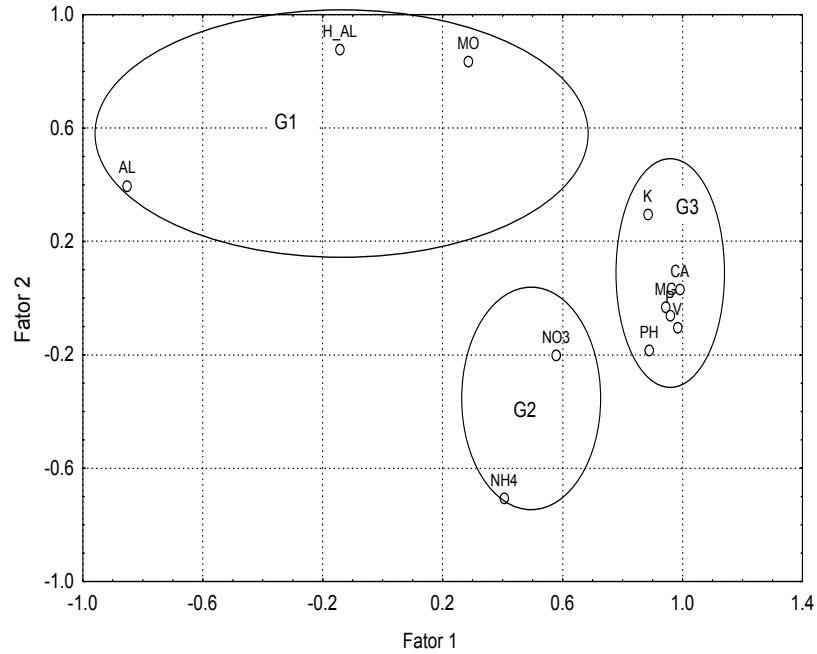
Os fatores calculados pelo método das componentes principais dispostas na tabela 6 da população 1 foram plotados no gráfico da figura 15 a seguir:

FIGURA 15 – ORDENAÇÃO DAS VARIÁVEIS PELO MÉTODO DAS COMPONENTES PRINCIPAIS DA POPULAÇÃO 1



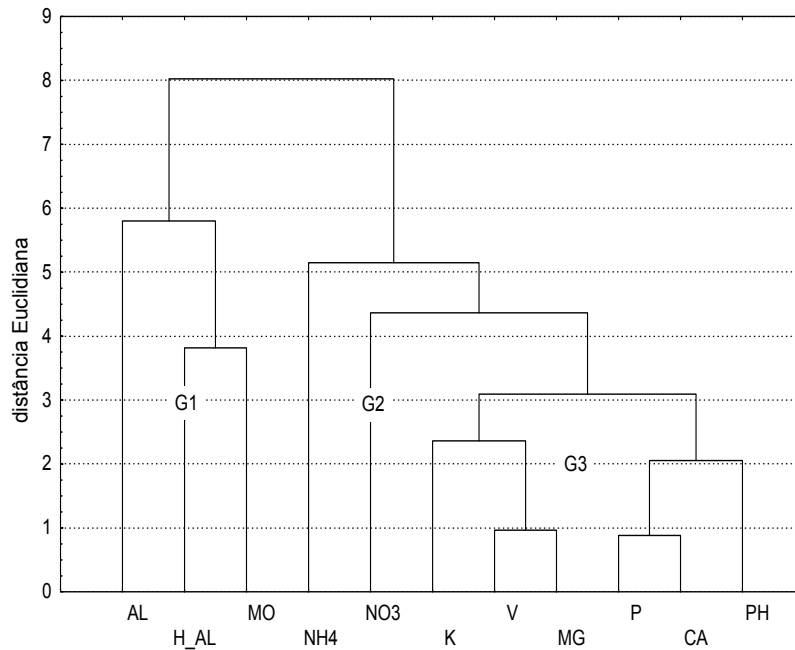
Conforme a tabela 15 os fatores da rotação varimax da população estão no gráfico da figura 16.

FIGURA 16 – ORDENAÇÃO DAS VARIÁVEIS PELA ROTAÇÃO VARIMAX DA POPULAÇÃO 1



Observando-se as figuras 15 e 16 permaneceram os mesmos grupos nas duas rotações. Os fatores obtidos pelo método das componentes principais dos atributos químicos, são agrupados pela análise fatorial da população 1 através da figura 17.

FIGURA 17 - DENDROGRAMA DA POPULAÇÃO 1



A observação conjunta da tabela 6 com as figuras 15 e 16 permite que a análise de componentes principais confirme a análise de agrupamentos, sendo que as variáveis dos resultados das análises de solos foram sensíveis para discriminá-los.

Foram agrupados de 3 grupos distintos constituídos pela semelhança (pode ser a correlação) e pela distância euclidiana dos atributos químicos. O Grupo G1 reuniu as variáveis Al, H+Al e MO com uma distância média 6 e o grupo G2 reuniu as variáveis NH₄ e NO₃ com uma distância média de 5 e o último grupo G3 tem uma distância média de 3 e estão alocadas neste grupo as variáveis K, V, Mg, P, Ca, pH.

TABELA 6 – MATRIZ DE CORRELAÇÃO DA POPULAÇÃO 1

VARIAVEIS	pH	MO	Al	H+Al	Ca	Mg	K	P	V	NH ₄	NO ₃
Ph	1										
MO	0,0716	1									
Al	-0,8134	0,0102	1								
H+Al	-0,2951	0,5717	0,5558	1							
Ca	0,8759	0,2833	-0,8239	-0,1189	1						
Mg	0,7843	0,2191	-0,8479	-0,1456	0,9132	1					
K	0,7187	0,3788	-0,5493	0,2183	0,8915	0,8359	1				
P	0,8806	0,2164	-0,8629	-0,2165	0,9771	0,8623	0,7997	1			
V	0,8663	0,1781	-0,8937	-0,2422	0,9701	0,9726	0,8372	0,9452	1		
NH ₄	0,4289	-0,4743	-0,5382	-0,4713	0,3525	0,4248	0,2207	0,3825	0,4472	1	
NO ₃	0,4922	0,0957	-0,4677	-0,2487	0,5465	0,4398	0,4545	0,4855	0,5212	0,4219	1

As variáveis que possuem uma correlação significativa é maior ou igual que 70%, e abaixo de 70% têm uma correlação fraca.

No tabela 7 seguinte tem-se a matriz dos resíduos da população 1 :

TABELA 7 – MATRIZ DOS RESÍDUOS DA POPULAÇÃO 1

	Ph	MO	Al	H+Al	Ca	Mg	K	P	V	NH ₄	NO ₃
pH	0										
MO	-0,0282	0									
Al	0,0180	-0,0780	0								
H+Al	-0,0057	-0,1200	0,0851	0							
Ca	0,0014	-0,0261	0,0080	-0,0045	0						
Mg	-0,0601	-0,1212	-0,0306	0,0181	-0,0199	0					
K	-0,0127	-0,0065	0,0873	0,0858	0,0064	0,0112	0				
P	0,0170	-0,0168	-0,0209	-0,0252	0,0294	-0,0442	-0,0302	0			
V	-0,0285	-0,0168	-0,0152	-0,0110	0,0005	0,0427	-0,0012	-0,0030	0		
NH ₄	-0,0600	-0,0007	0,0861	0,2048	-0,0250	0,0213	0,0721	0,0478	-0,0219	0	
NO ₃	-0,0585	0,0980	0,1048	0,0101	-0,0195	-0,1118	0,0027	-0,0810	-0,0671	0,0473	0

Na matriz residual da população 1 os resultados estão abaixo de 15%, pode-se considerar um bom ajuste do modelo.

Na população 2 estão relacionados na tabela 8 os fatores das análises de componentes principais e a rotação varimax:

TABELA 8 - FATORES DA ROTAÇÃO VARIMAX E O MÉTODO DAS COMPONENTES PRINCIPAIS DA POPULAÇÃO 2

Variáveis	Método das componentes		Rotação Varimax		comunalidades
	Y ₁	Y ₂	F ₁	F ₂	
pH	0,9673	-0,0210	0,9504	0,1812	0,9361
MO	0,8350	0,2594	0,8665	-0,1173	0,7645
Al	-0,9806	0,0375	-0,9608	-0,1997	0,9631
H+Al	-0,9491	0,0047	-0,9351	-0,1621	0,9008
Ca	0,9580	0,0634	0,9552	0,0964	0,9218
Mg	0,7679	-0,2668	0,7129	0,3905	0,6608
K	0,9039	0,1256	0,9123	0,0262	0,8329
P	0,8011	0,1006	0,8066	0,0337	0,6518
V	-0,2407	0,9440	-0,0808	-0,9709	0,9491
NH ₄	0,8852	0,0192	0,8761	0,1279	0,7839
NO ₃	0,8565	0,00614	0,8456	0,1360	0,7336
autovalores	8,04	1,06	7,84	1,25	
Proporção Acumulada da Variância	73,06%	82,71%	71,34%	82,71%	

Conforme a tabela 8 o modelo ajustado pelos métodos das componentes principais e na rotação varimax levou em consideração dois fatores, que em conjunto explicam 82,71% das variáveis incluídas. Sendo as comunalidades o percentual de cada variável que é explicado pelos fatores no conjunto, mostra que a variável que melhor explicada pelos fatores é o Al com 96,31% de sua variação que foi captada pelos fatores, e as de menores explicações são P com 65,81% e o Mg com 65,18%, e as demais apresentaram comunalidades entre 73% em 95%, de um modo geral mostra que suas variações foram captadas de forma satisfatória

pelos fatores ajustados. Quanto à correlação das variáveis dentro do seu próprio fator as variáveis tem uma alta correlação e entre os fatores uma correlação fraca. As correlações das variáveis estão na matriz na tabela 9.

Nas figuras 18 e 19 estão os fatores da população 2 pelo método das componentes principais e a rotação varimax:

FIGURA 18 – ORDENAÇÃO DAS VARIÁVEIS PELO MÉTODO DAS COMPONENTES PRINCIPAIS DA POPULAÇÃO 2

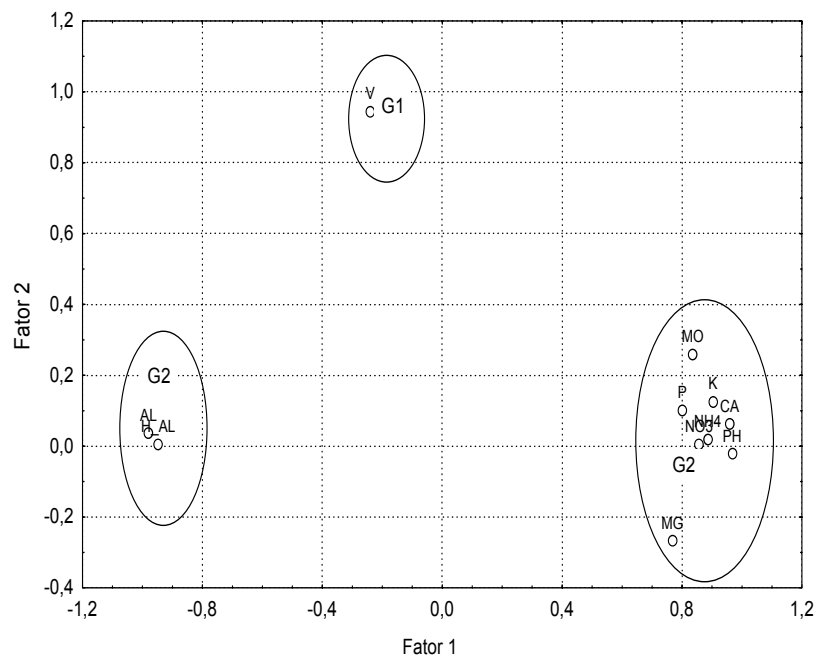
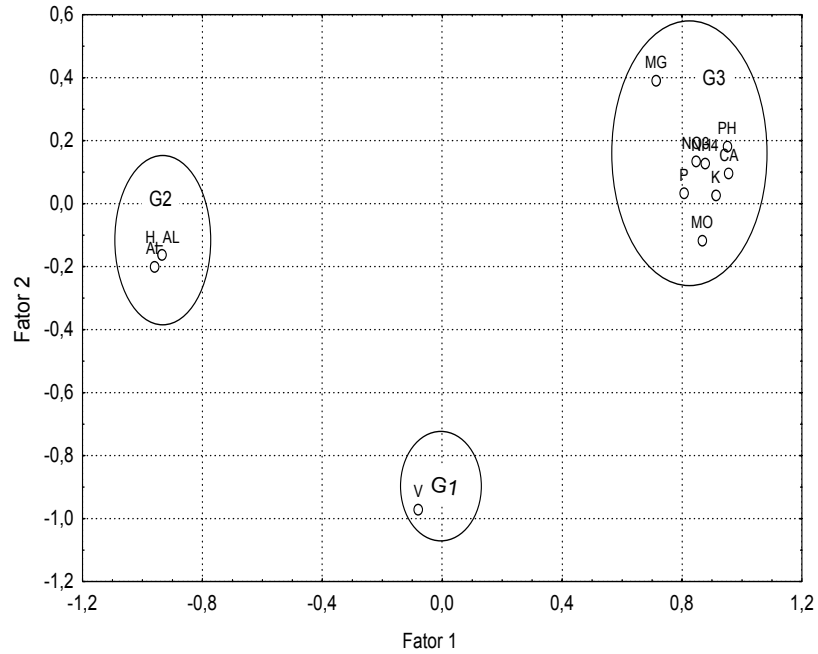
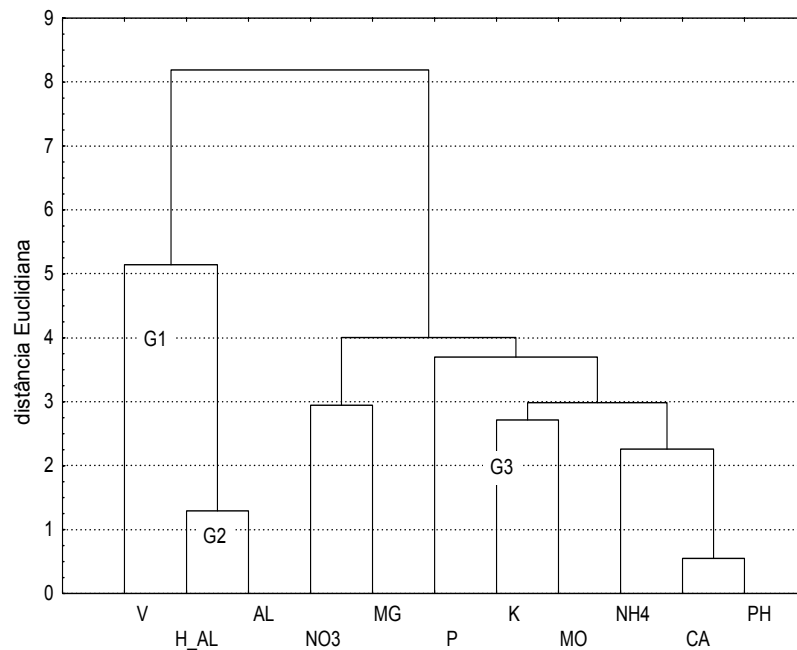


FIGURA 19 – ORDENAÇÃO DAS VARIÁVEIS PELA ROTAÇÃO VARIMAX DA POPULAÇÃO 2



Os agrupamentos das variáveis da população 2 estão na figura 20:

FIGURA 20 - DENDROGRAMA DA POPULAÇÃO 2



Observando as figuras 18, 19 e 20 a análise de componentes principais confirma a análise de agrupamentos, sendo que as variáveis são agrupadas conforme suas correlações e suas distâncias médias. A saturação por base (V) tem sua distância média 5,1, portanto está no grupo G1, e a distância média do grupo G2 é H+Al e Al , com um valor de 1,4, e a distância média do grupo G3 é 4 onde estão agrupados as demais variáveis.

No tabela 9 estão as correlações das variáveis da população 2:

TABELA 9 – MATRIZ DE CORRELAÇÃO DA POPULAÇÃO 2

Variáveis	pH	MO	Al	H+Al	Ca	Mg	K	P	V	NH ₄	NO ₃
pH	1										
MO	0,7734	1									
Al	-0,9732	-0,8092	1,0000								
H+Al	-0,9243	-0,7582	0,9508	1							
Ca	0,9911	0,7761	-0,9664	-0,9148	1						
Mg	0,6650	0,5644	-0,7461	-0,7129	0,6622	1					
K	0,8497	0,7832	-0,8257	-0,8473	0,8531	0,6628	1				
P	0,7767	0,5975	-0,7851	-0,7299	0,7609	0,5287	0,7506	1			
V	-0,2676	-0,0103	0,2687	0,2220	-0,1736	-0,3300	-0,1211	-0,1237	1		
NH ₄	0,8660	0,7722	-0,8672	-0,7897	0,8500	0,5715	0,7378	0,6252	-0,2187	1	
NO ₃	0,7566	0,6597	-0,7876	-0,8077	0,7412	0,7447	0,7720	0,6157	-0,1523	0,7854	1

As correlações iguais ou acima de 70% são significativas e as variáveis abaixo de 70% tem uma correlação fraca

Na tabela 10 seguinte tem-se a matriz dos resíduos da população 2:

TABELA 10 – MATRIZ DOS RESÍDUOS DA POPULAÇÃO 2

Variáveis	pH	MO	Al	H+Al	Ca	Mg	K	P	V	NH ₄	NO ₃
pH	0										
MO	-0,0289	0									
Al	-0,0238	-0,0001	0								
H+Al	-0,0062	0,0331	0,0199	0							
Ca	0,0658	-0,0403	-0,0293	-0,0059	0						
Mg	-0,0834	-0,0076	0,0169	0,0171	-0,0565	0					
K	-0,0221	-0,0042	0,0560	0,0100	-0,0208	0,0022	0				
P	0,0039	-0,0975	-0,0033	0,0299	-0,0128	-0,0595	0,0138	0			
V	-0,0148	-0,0542	-0,0028	-0,0109	-0,0028	0,1067	-0,0220	-0,0258	0		
NH ₄	0,0101	0,0281	0,0001	0,0503	0,0008	-0,1030	-0,0648	-0,0858	-0,0237	0	
NO ₃	-0,0717	-0,0571	0,0520	0,0052	-0,0797	0,0887	-0,0029	-0,0710	0,0481	0,0272	0

Com os dados da população 2 a matriz dos resíduos na grande maioria de seus valores menores que 15%, portanto pode-se considerar um bom ajuste do modelo.

Para analisar a população 3, na tabela 11 são dados os fatores obtidos pelo método das componentes principais e a rotação varimax:

TABELA 11 - FATORES DA ROTAÇÃO VARIMAX E O MÉTODO DAS COMPONENTES PRINCIPAIS DA POPULAÇÃO 3

Variáveis	Método das componentes		Rotação Varimax		comunalidades
	Y ₁	Y ₂	F ₁	F ₂	
pH	-0,8943	0,3671	0,9486	0,1863	0,9346
MO	-0,6823	-0,6066	0,2353	0,8821	0,8335
Al	0,9501	-0,0429	-0,8166	-0,4876	0,9045
H+Al	0,9132	-0,3208	-0,9389	-0,2354	0,9369
Ca	-0,9834	0,0453	0,8457	0,5040	0,9691
Mg	-0,8132	0,4672	0,9360	0,0581	0,8795
K	-0,8734	-0,4434	0,4847	0,8512	0,9594
P	-0,7978	-0,3601	0,4674	0,7401	0,7662
V	-0,9587	0,2601	0,9434	0,3111	0,9868
NH ₄	-0,6542	-0,1802	0,4467	0,5108	0,4605
NO ₃	-0,2049	-0,6318	-0,1770	0,6402	0,4412
autovalores	7,42	1,65	5,67	3,4031	
Proporção Acumulada da Variância	67,44%	82,47%	51,54%	83,48%	

De acordo com a tabela 11 a análise fatorial da população 3 o modelo foi ajustado por 2 fatores tanto no método das componentes principais e na rotação varimax com autovalores maiores que 1. No conjunto explicaram 82,47% no método das componentes principais e na rotação varimax 83,48% das variâncias das variáveis incluídas no modelo. A variável que mais é explicada pelos fatores de acordo com suas comunalidades é V (saturação por bases) com 98,68% e as variáveis que são menos explicadas pelos fatores são amônia e nitrato com 46,05% e 44,12%, e as demais variáveis estão entre 76% e 97%, o que mostra que suas

variações, de um modo geral são captadas de forma satisfatória pelos fatores ajustados. E que cada fator relacionado no quadro 6 tem as suas variáveis uma correlação alta e entre eles uma correlação fraca. Nas figuras 21 e 22 estão ordenados os fatores e na figura 23 foi feita a análise de agrupamento.

FIGURA 21- ORDENAÇÃO DAS VARIÁVEIS PELO MÉTODO DAS COMPONENTES PRINCIPAIS DA POPULAÇÃO 3

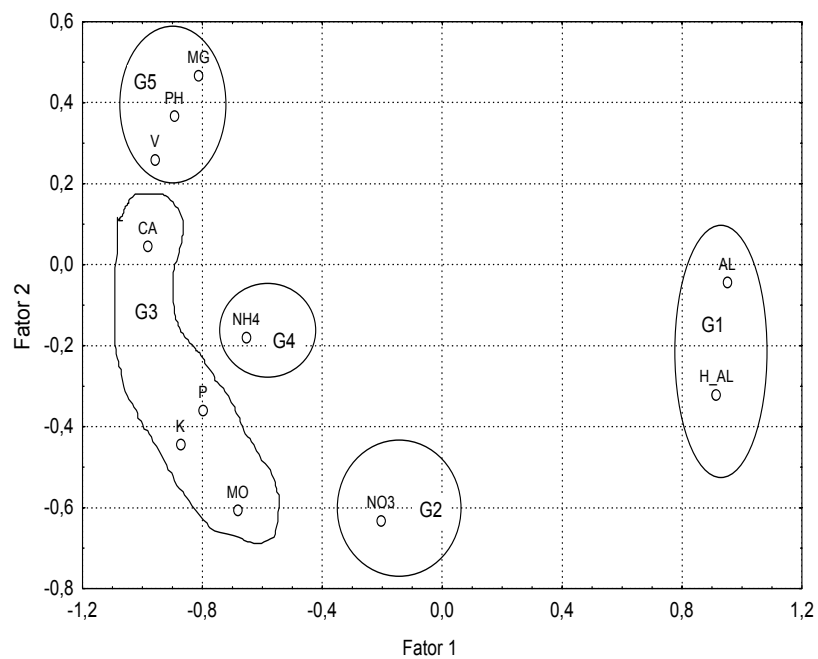


FIGURA 22 – ORDENAÇÃO DAS VARIÁVEIS PELA ROTAÇÃO VARIMAX DA POPULAÇÃO 3

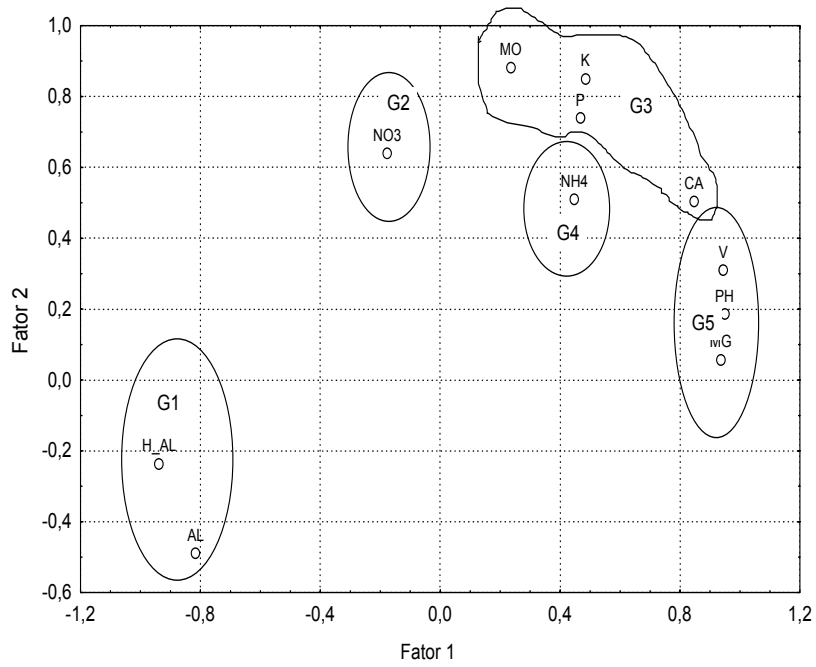
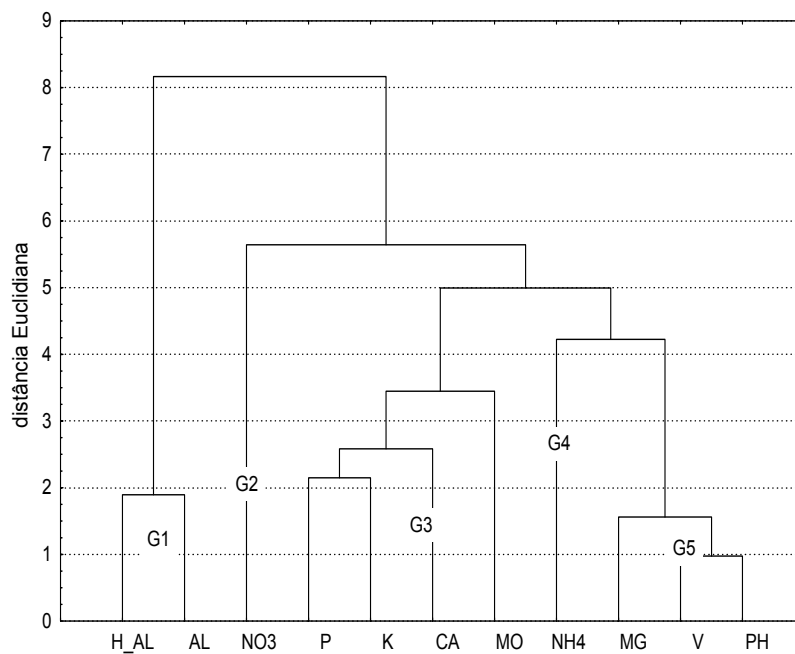


FIGURA 23 – DENDROGRAMA DA POPULAÇÃO 3



Comparando-se as figuras 21, 22 e 23 confirma-se a análise das componentes principais e análise de agrupamento. Para cada grupo relacionado na figura 23 eles são homogêneos com características de correlação e distância média. As distâncias médias para os grupo G1 é de 1,9,o grupo G2 é de 5,8, o grupo G3 3,6 ,o grupo G4 é 5,0 e o grupo G5 é 1,8.

Na tabela 12 a matriz das correlações da população 3:

TABELA 12 – MATRIZ DE CORRELAÇÃO DA POPULAÇÃO 3

Variáveis	pH	MO	Al	H+Al	Ca	Mg	K	P	V	NH ₄	NO ₃
Ph	1										
MO	0,3683	1									
Al	-0,8216	-0,6127	1								
H+Al	-0,9102	-0,4257	0,8944	1							
Ca	0,8840	0,6507	-0,9612	-0,9114	1						
Mg	0,9559	0,2663	-0,7356	-0,8602	0,7899	1					
K	0,6017	0,8440	-0,8218	-0,6460	0,8298	0,4818	1				
P	0,5375	0,7597	-0,7818	-0,6542	0,8037	0,3602	0,8642	1			
V	0,9720	0,4925	-0,9100	-0,9541	0,9578	0,9284	0,7067	0,6542	1		
NH ₄	0,5003	0,5054	-0,5331	-0,4818	0,5750	0,4748	0,7130	0,4184	0,5445	1	
NO ₃	0,0838	0,3758	-0,1120	0,0203	0,1440	0,0632	0,3923	0,2081	0,1002	0,1504	1

Na tabela 12 as variáveis que possuem uma correlação igual ou acima de 70% é significativa e as demais variáveis que possuem abaixo de 70% tem uma correlação fraca.

Na tabela 13 está a matriz dos resíduos da população 3:

TABELA 13 – MATRIZ DOS RESÍDUOS DA POPULAÇÃO 3

Variáveis	pH	MO	Al	H+Al	Ca	Mg	K	P	V	NH ₄	NO ₃
pH	0										
MO	-0,0193	0									
Al	0,0438	0,0096	0								
H+Al	0,0242	0,0029	0,0129	0							
Ca	-0,0121	0,0072	-0,0249	0,0012	0						
Mg	0,0572	-0,0051	0,0570	0,0323	-0,0309	0					
K	-0,0167	-0,0209	-0,0110	0,0094	-0,0090	-0,0213	0				
P	-0,0437	-0,0031	-0,0392	-0,0412	0,0355	-0,1203	0,0077	0			
V	0,0191	-0,0039	0,0121	0,0049	0,0032	0,0273	-0,0154	-0,0170	0		
NH ₄	-0,0186	-0,0504	0,0808	0,0579	-0,0603	0,0270	0,0616	-0,1685	-0,0359	0	
NO ₃	0,1324	-0,1472	0,0556	0,0048	-0,0289	0,1918	-0,0668	-0,1829	0,0680	-0,0975	0

A matriz de resíduos da população 3 , conforme tabela 13 tem os seus resíduos abaixo de 15%, portanto teve um bom ajuste do modelo.

Com os dados da população 4 foram obtidos os pesos pela rotação varimax e pelo método das componentes principais, conforme a tabela 14:

TABELA 14 - FATORES DA ROTAÇÃO VARIMAX E O MÉTODO DAS COMPONENTES PRINCIPAIS DA POPULAÇÃO 4

Variáveis	Método das componentes		Rotação Varimax		comunalidades
	Y ₁	Y ₂	F ₁	F ₂	
pH	0,9584	0,2536	0,9592	0,2506	0,9829
MO	0,8220	-0,2714	0,5822	0,6406	0,7493
Al	-0,9561	0,1174	-0,7747	-0,5724	0,9279
H+Al	-0,9748	-0,0446	-0,8707	-0,4406	0,9521
Ca	0,9544	-0,0426	0,8100	0,5065	0,9127
Mg	0,8763	0,3388	0,9296	0,1360	0,8827
K	0,5627	-0,7461	0,1230	0,9264	0,8733
P	0,6806	-0,5482	0,3229	0,8121	0,7637
V	0,9855	0,1106	0,9125	0,3884	0,9835
NH ₄	0,8862	0,1530	0,8469	0,3027	0,8088
NO ₃	0,8537	0,3994	0,9398	0,0722	0,8884
Autovalores	8,4026	1,3227	6,6900	3,0353	
Proporção Acumulada da Variância	76,39%	88,41%	60,82%	88,41%	

Com o instrumento da análise fatorial, conforme tabela 14 o modelo ajustado levou em consideração dois fatores tanto no método das componentes principais e como na rotação varimax. Neste modelo foi apresentado autovalor maior que 1 (um) que, em conjunto, explicaram 88,41% das variâncias das variáveis incluídas no modelo no método das componentes principais e na rotação varimax. O percentual das comunalidades que obtiveram maior valor foi pH e V, com 98,29% e 98,35% e a de menor explicação foi MO com 74,39% de sua variação explicada pelos fatores. As demais variáveis apresentaram comunalidades

entre 76% e 96%, de modo que suas variações foram captadas de forma satisfatória pelos fatores ajustados. Na tabela 15 está a matriz das correlações das variáveis da população 4.

Nas figuras 24 e 25 estão ordenados as variáveis pelo método das componentes principais e a rotação varimax. Na figura 26 as variáveis estão agrupadas pela análise de agrupamento.

FIGURA 24 – ORDENAÇÃO DAS VARIÁVEIS PELO MÉTODO DAS COMPONENTES PRINCIPAIS DA POPULAÇÃO 4

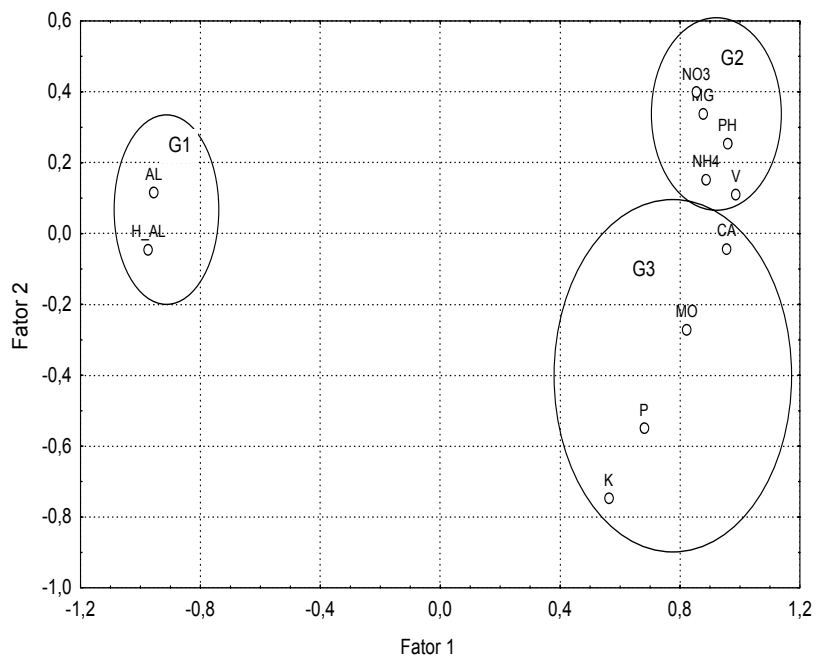


FIGURA 25 – ORDENAÇÃO DAS VARIÁVEIS PELA ROTAÇÃO VARIMAX DA POPULAÇÃO 4

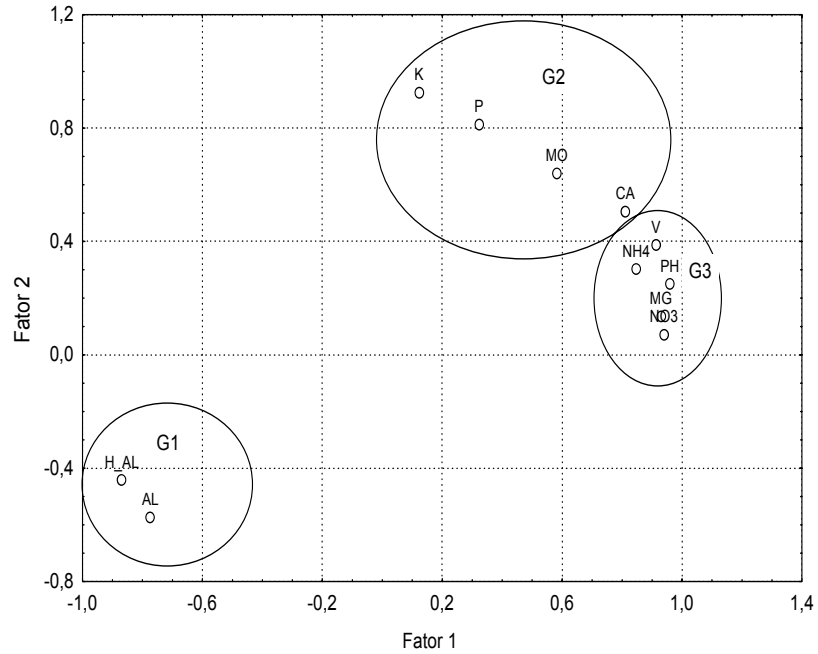
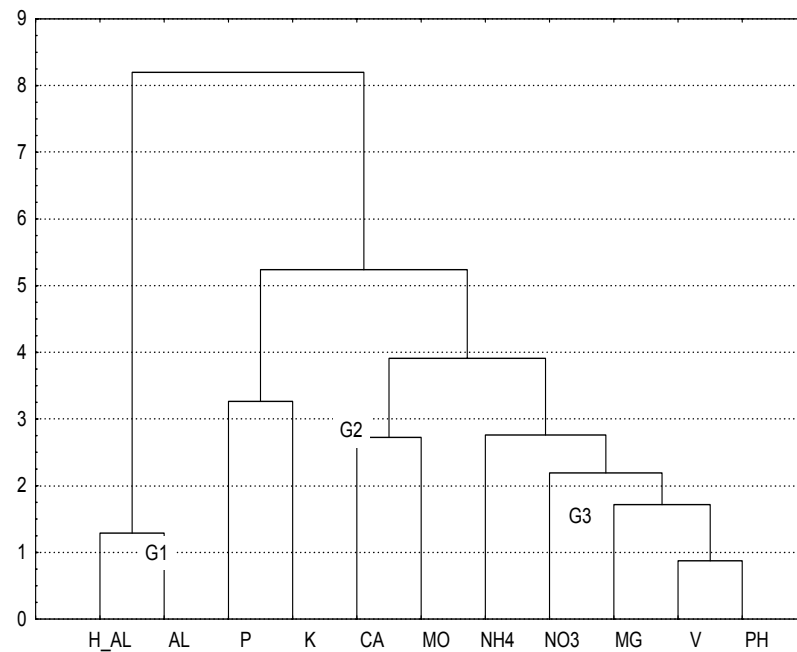


FIGURA 26 – DENDROGRAMA DA POPULAÇÃO 4



Comparando-se as figuras 24 e 25 a análise das componentes principais confirma a análise de agrupamentos mesmo com a rotação dos fatores os grupos permaneceram os mesmos. Dos grupos formados: G1 teve uma distância média de 1,5, o grupo G2 teve uma distância de 3,3 e o grupo G3 teve uma distância média de 4,0. Estes grupos foram constituídos pela similaridade (correlação) e dissimilaridade (distância média) das 11 variáveis estudadas.

Na tabela 15 tem-se a matriz das correlações da população 4:

TABELA 15 –MATRIZ DE CORRELAÇÃO DA POPULAÇÃO 4

Variáveis	pH	MO	Al	H+Al	Ca	Mg	K	P	V	NH ₄	NO ₃
pH	1										
MO	0,7107	1									
Al	-0,8801	-0,7807	1								
H+Al	-0,9521	-0,7559	0,9510	1							
Ca	0,9106	0,7818	-0,9342	-0,9340	1						
Mg	0,9135	0,6410	-0,7493	-0,8488	0,7916	1					
K	0,3640	0,5977	-0,6147	-0,5348	0,5455	0,2366	1				
P	0,4981	0,6511	-0,6898	-0,5962	0,6576	0,4679	0,6865	1			
V	0,9774	0,7719	-0,9324	-0,9771	0,9576	0,9133	0,4691	0,6010	1		
NH ₄	0,8810	0,6799	-0,8369	-0,8673	0,7817	0,7759	0,4452	0,4448	0,8573	1	
NO ₃	0,9180	0,5502	-0,7511	-0,8189	0,7745	0,8735	0,1927	0,4384	0,8587	0,8054	1

Através da tabela 15 as variáveis com correlação acima de 70% tem correlação significativa e abaixo de 70% possuem uma correlação fraca.

A seguir tem-se a matriz dos resíduos na tabelas 16:

TABELA 16 – MATRIZ DOS RESÍDUOS DA POPULAÇÃO 4

Variáveis	pH	MO	Al	H+Al	Ca	Mg	K	P	V	NH ₄	NO ₃
Ph	0										
MO	-0,0083	0									
Al	0,0065	0,0371	0								
H+Al	-0,0066	0,0332	0,0243	0							
Ca	0,0067	-0,0143	-0,0167	-0,0057	0						
Mg	-0,0123	0,0127	0,0488	0,0205	-0,0302	0					
K	0,0139	-0,0674	0,0109	-0,0195	-0,0234	-0,0040	0				
P	-0,0152	-0,0571	0,0253	0,0428	-0,0153	0,0573	-0,1050	0			
V	0,0049	-0,0081	-0,0032	-0,0115	0,0217	0,0123	-0,003	-0,0091	0		
NH ₄	-0,0072	-0,0071	-0,0075	0,0034	-0,0576	-0,053	0,0606	-0,0745	-0,0330	0	
NO ₃	-0,0015	-0,0431	0,0183	0,0311	-0,0233	-0,0100	0,0102	0,0763	-0,0269	-0,0123	0

Conforme a tabela 16 os resíduos estão abaixo de 15%, considera-se um bom ajuste do modelo.

Na tabela 17 apresentam-se os fatores da análise das componentes principais e a rotação varimax:

TABELA 17 - FATORES PELO MÉTODO DAS COMPONENTES PRINCIPAIS DA POPULAÇÃO 5

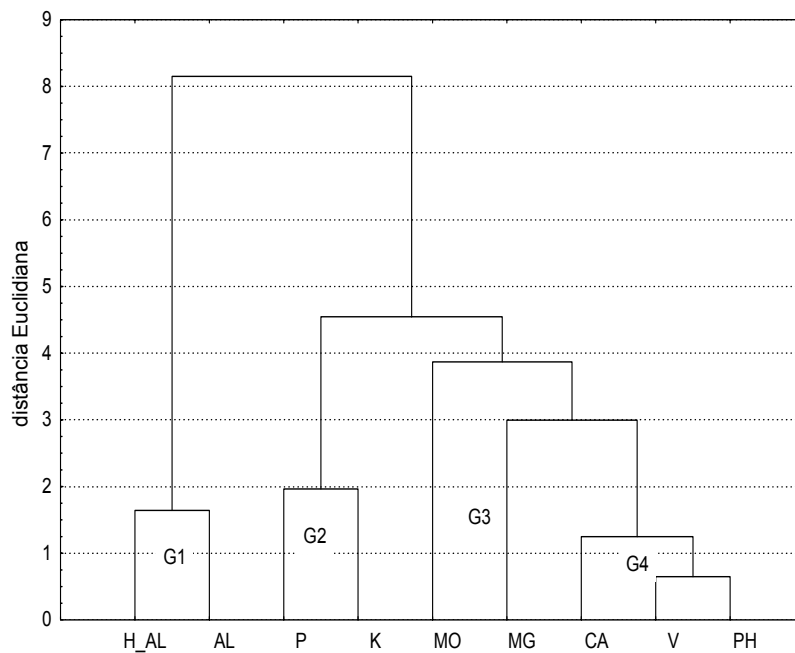
	Método das Componentes Principais	
Variáveis	Y_1	Comunalidades
pH	0,9733	0,9474
MO	0,6972	0,4861
Al	-0,9600	0,9215
H+Al	-0,9249	0,8554
Ca	0,9760	0,9526
Mg	0,8430	0,7107
K	0,7694	0,5919
P	0,8125	0,6602
V	0,9877	0,9756
NH ₄	0	0
NO ₃	0	0
autovalores	7,10	
Proporção Acumulada da Variância	78,91%	

De acordo com a tabela 17 o modelo ajustado levou em consideração apenas 1 fator que apresentou autovalor maior que 1 e que explica 78,91% das variâncias das variáveis incluídas no modelo. Sendo as comunalidades o percentual de cada variável que é explicado pelos fatores em conjunto, mostra que a variável que melhor foi explicada pelos fatores foi a saturação por base (V) com 97,56% de sua variação captada pelos fatores, e a de menor explicação foi a matéria orgânica (MO) com 48,61% de sua a variação explicada pelos fatores

estudados. As demais variáveis apresentaram comunalidades entre 59% e 96%, sendo que suas variações foram, de um modo geral, captadas de forma satisfatória pelos fatores ajustados.

Na figura 27 as variáveis da população 5 foram separadas pela análise de agrupamento.

FIGURA 27 - DENDROGRAMA DA POPULAÇÃO 5



Na análise de agrupamentos da população 5 conforme figura 27 obteve-se três grupos. O grupo G1 tem uma distância média de 1,8, o grupo G2 tem uma distância de 2,0, o grupo G3 tem uma distância média de 3,9 e o grupo G4 tem uma distância média de 1,3. Eles foram agrupados conforme sua distância média e suas correlações. Na tabela 18 é apresentada a matriz das correlações da população 5.

TABELA 18 – MATRIZ DE CORRELAÇÃO DA POPULAÇÃO 5

	pH	MO	Al	H+Al	Ca	Mg	K	P	V
pH	1								
MO	0,6211	1							
Al	-0,9364	-0,7113	1						
H+Al	-0,9442	-0,5178	0,9208	1					
Ca	0,9541	0,7105	-0,9533	-0,9077	1				
Mg	0,8485	0,5593	-0,7736	-0,7761	0,7360	1			
K	0,6862	0,3921	-0,6455	-0,5936	0,7502	0,5317	1		
P	0,7181	0,4265	-0,7091	-0,6800	0,7604	0,6007	0,8865	1	
V	0,9876	0,6830	-0,9545	-0,9414	0,9731	0,8625	0,6919	0,7298	1

Através da tabela 18 as variáveis com correlação acima de 70% tem correlação significativa e abaixo de 70% possuem uma correlação fraca.

Na tabela 19 tem-se a matriz dos resíduos:

TABELA 19 – MATRIZ DOS RESÍDUOS DA POPULAÇÃO 5

	pH	MO	Al	H+Al	Ca	Mg	K	P	V
Ph	0								
MO	-0,0575	0							
Al	-0,0021	-0,0420	0						
H+Al	-0,0440	0,1270	0,0329	0					
Ca	0,0041	0,0299	-0,0163	-0,0050	0				
Mg	0,0280	-0,0285	0,0356	0,0036	-0,0869	0			
K	-0,0626	-0,1444	0,0930	0,1180	-0,0007	-0,1169	0		
P	-0,0728	-0,1400	0,0709	0,0714	-0,0326	-0,0843	0,2614	0	
V	0,0262	-0,0056	-0,0063	-0,0278	0,0090	0,0298	-0,0680	-0,0727	0

Conforme quadro 14 os resíduos estão abaixo de 15%, considera-se um bom ajuste do modelo.

Na tabela 20 é apresentado o fator pelo método das componentes principais e pela rotação varimax da população 6 que é representada pelas variáveis derivadas dos animais no qual o objeto de estudo é o ganho médio diário.

TABELA 20 - FATORES DA ROTAÇÃO VARIMAX E O MÉTODO DAS COMPONENTES PRINCIPAIS DA POPULAÇÃO 6

		Método das componentes		Rotação Varimax		
Potreiros	variáveis	Y ₁	Y ₂	F ₁	F ₂	comunalidades
1	Massa alta	0,5979	-0,6453	0,2421	0,8457	0,7739
2	Massa baixa	0,9313	0,2536	0,9454	0,1942	0,9315
3	Massa alta	0,9269	-0,0602	0,7999	0,4723	0,8628
4	Massa baixa	0,9091	0,3955	0,9898	0,0576	0,9829
5	Massa alta	0,4825	-0,7875	0,0749	0,9205	0,8530
6	Massa baixa	0,7809	0,2892	0,8273	0,0945	0,6934
Autovalores		3,75	1,34	3,26	1,84	
Proporção Acumulada da Variância		62,55%	84,96%	54,37%	84,95%	

A população 6 é designada pelas variáveis derivadas dos animais (ganho médio diário) que obteve dois fatores nos dois métodos dados da tabela 20 com seus autovalores maiores que 1. O modelo ajustado em conjunto no método das componentes principais explicou 84,96% e na rotação varimax 84,95%. A variável que melhor foi explicada pelos fatores foi o potreiro 4 (massa baixa) com 98,29% de sua variação captada pelos fatores e a menor explicação foi o potreiro 6 (massa baixa) 69,34% de sua variação explicada pelos

fatores estudados. As demais variáveis apresentaram comunalidades entre 77% e 94%, o que mostra que suas variações foram de modo geral, captadas de forma satisfatória pelos fatores ajustados.

Nas figuras 28 e 29 estão ordenados no gráfico os fatores da análise de componente principal e a rotação varimax, e na figura 30 estão separados os grupos através da análise de agrupamento.

FIGURA 28 – ORDENAÇÃO DAS VARIÁVEIS PELO MÉTODO DAS COMPONENTES PRINCIPAIS DA POPULAÇÃO 6

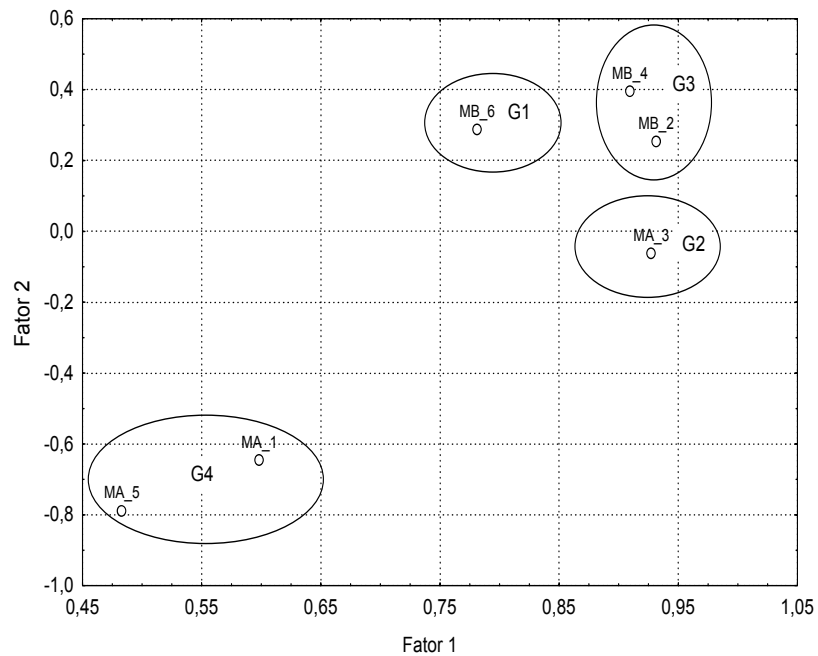


FIGURA 29 – ORDENAÇÃO DAS VARIÁVEIS PELA ROTAÇÃO VARIMAX DA POPULAÇÃO 6

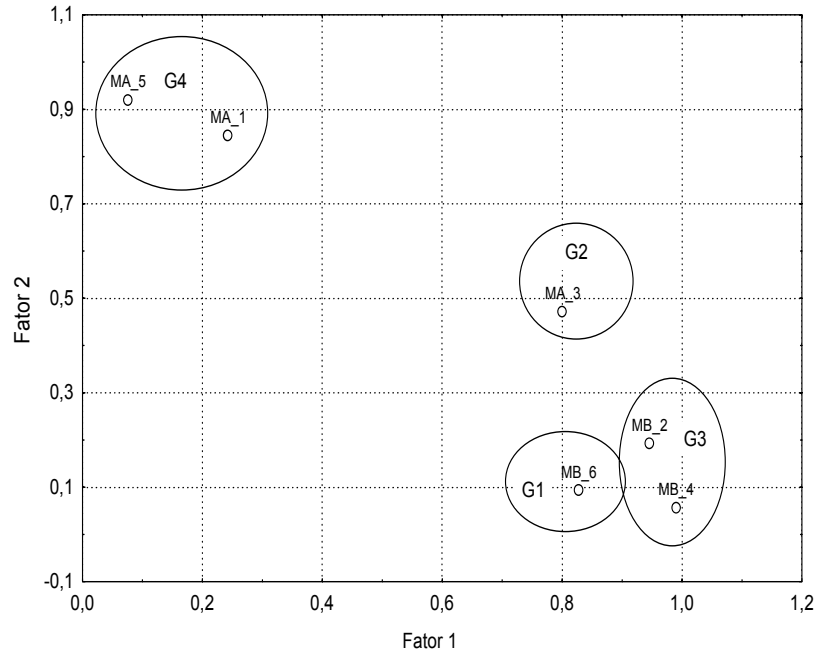
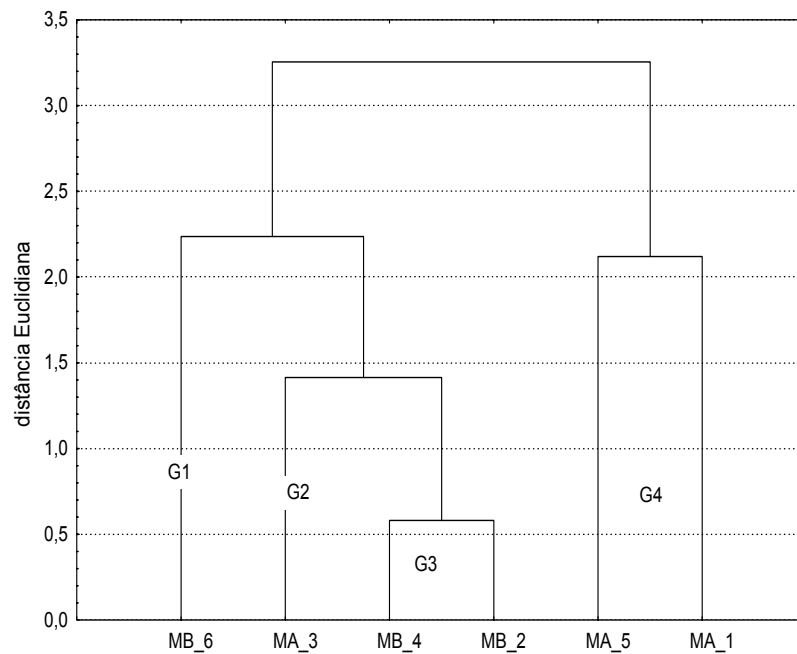


FIGURA 30 – DENDROGRAMA DA POPULAÇÃO 6



A análise de componentes principais confirma a análise de agrupamento sendo que os grupos formados são homogêneos segundo as suas similaridades (correlações) e suas dissimilaridades (distância média). Os grupos formados conforme suas distâncias médias são

o grupo G1 com sua distância média de 2,3 é o potreiro 6 com massa baixa dado pela variável MB_6 e o grupo G2 com uma distância 1,4 é o potreiro 3 com massa alta dado pela variável (MA_3). Os grupos foram constituídos pelos grupos G3 com distância média de 0,6 com os poteiros 4 e 2 com massa baixa e é dado pelas variáveis MB_4 e MB_2 e o grupo G4 tem uma distância média 2,2 para os poteiros 5 e 6 com massa alta dado pelas variáveis MA_5 e MA_1.

Na tabela 21 tem as correlações das variáveis derivadas dos animais (ganho médio diário).

TABELA 21 – MATRIZ DE CORRELAÇÃO DA POPULAÇÃO 6

	MA_1	MB_2	MA_3	MB_4	MA_5	MB_6
MA_1	1					
MB_2	0,4058	1				
MA_3	0,5140	0,8334	1			
MB_4	0,2868	0,9719	0,8429	1		
MA_5	0,6255	0,2394	0,5110	0,1171	1	
MB_6	0,2933	0,7038	0,5832	0,7535	0,1935	1

As correlações dadas na matriz da tabela 15 que estão acima de 70% têm uma correlação significativa e as que estão abaixo de 70% são correlações fracas.

As matrizes dos resíduos estão na tabela 22:

TABELA 22 - MATRIZ DOS RESÍDUOS DA POPULAÇÃO 6

	MA_1	MB_2	MA_3	MB_4	MA_5	MB_6
MA_1	0					
MB_2	0,0126	0				
MA_3	-0,0790	-0,0146	0			
MB_4	-0,0015	0,0250	0,0241	0		
MA_5	-0,1711	-0,0102	0,0164	-0,0100	0	
MB_6	0,0131	-0,0968	-0,1232	-0,0709	0,0445	0

Os resíduos da população 6 estão na maioria abaixo de 15% o que contribui para um bom ajuste do modelo.

Na tabela 23 são apresentados os fatores da população 7 variáveis derivadas dos animais (produção animal) através da rotação varimax e pelo método das componentes principais.

TABELA 23 - FATORES DA ROTAÇÃO VARIMAX E O MÉTODO DAS COMPONENTES PRINCIPAIS DA POPULAÇÃO 7

		Método das componentes		Rotação Varimax		
Potreiros	variáveis	Y ₁	Y ₂	F ₁	F ₂	comunalidades
1	Massa alta	0,7816	-0,4827	0,2940	0,8704	0,8440
2	Massa baixa	0,8448	0,3766	0,8908	0,2489	0,8555
3	Massa alta	0,9053	-0,1778	0,5837	0,7144	0,8511
4	Massa baixa	0,8520	0,4795	0,9620	0,1743	0,9559
5	Massa alta	0,6593	-0,7197	0,0486	0,9748	0,9527
6	Massa baixa	0,7057	0,4054	0,8021	0,1380	0,6624
Autovalores		3,80	1,32	2,79	2,33	
Proporção Acumulada da Variância		63,38%	85,36%	46,53%	85,36%	

A população 7 é designada pelas variáveis derivadas de animais (produção animal) que obteve dois fatores nos dois métodos dados na tabela 23 com seus autovalores maiores que 1. O modelo ajustado em conjunto no método das componentes principais explicou 85,36% e na rotação varimax 85,36%. As variáveis que melhor foram explicadas pelos fatores foi o potreiro 4 (massa baixa) com 95,59% e o potreiro 5 (massa alta) com 95,59% de suas variações captadas pelos fatores e a menor explicação foi o potreiro 6 (massa baixa) 66,24%

de sua variação explicada pelos fatores estudados. As demais variáveis apresentaram comunalidades entre 84% e 86%, o que mostra que suas variações foram de modo geral, captadas de forma satisfatória pelos fatores ajustados.

Nas figuras 23, 24 estão ordenados os fatores pelo método das componentes principais e a rotação varimax respectivamente, e na figura 25 tem-se o dendrograma da população 7.

FIGURA 23 –ORDENAÇÃO DAS 11 VARIÁVEIS PELO MÉTODO DAS COMPONENTES PRINCIPAIS DA POPULAÇÃO 7

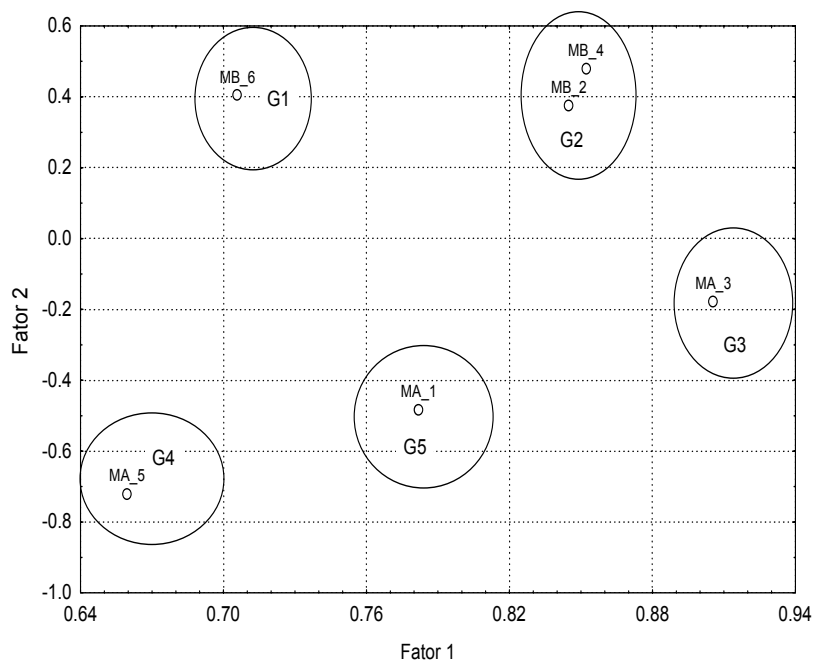


FIGURA 24 – ORDENAÇÃO DAS 11 VARIÁVEIS PELA ROTAÇÃO VARIMAX DA POPULAÇÃO 7

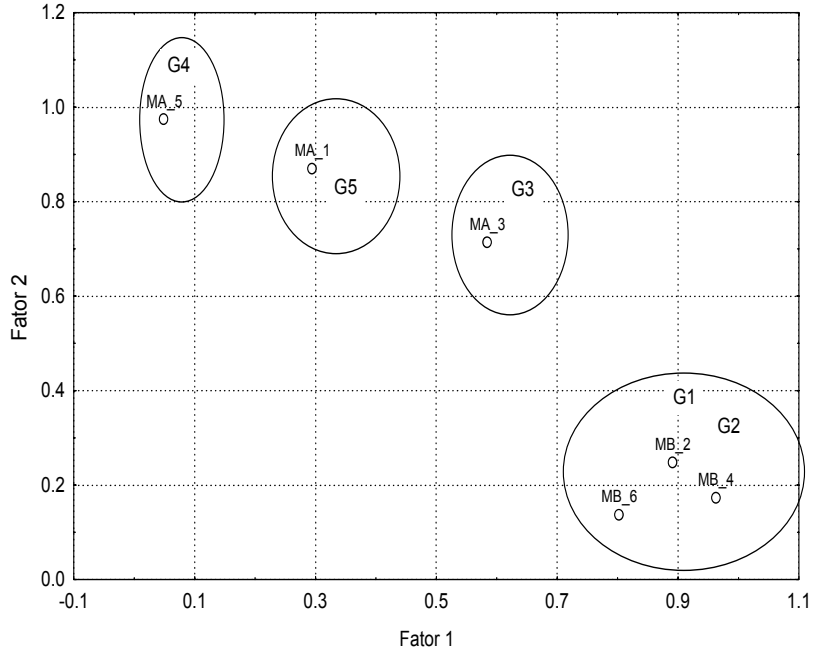
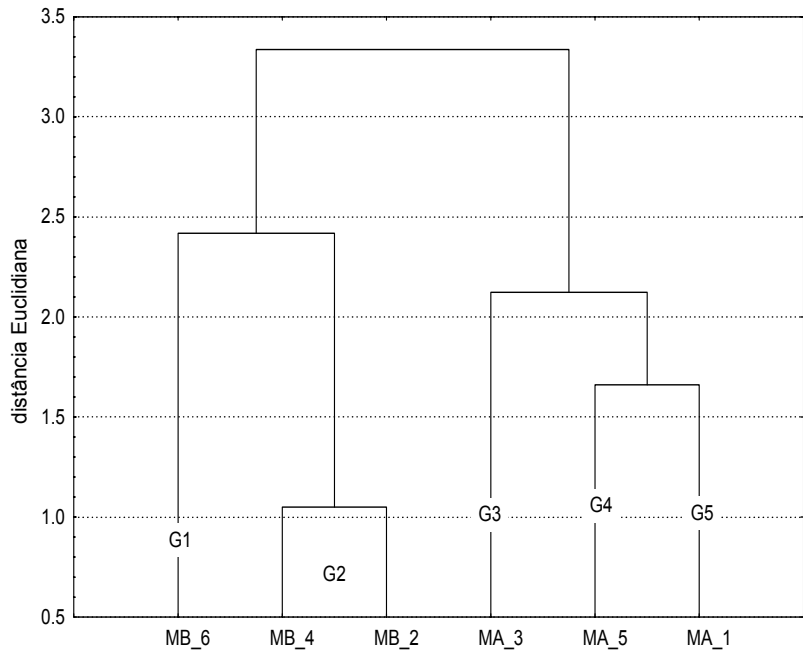


FIGURA 25 – DENDROGRAMA DA POPULAÇÃO 7



A método de componentes principais confirma a análise de agrupamento sendo que os grupos formados são homogêneos segundo as suas similaridades (correlações) e suas dissimilaridades (distância média). Os grupos formados conforme suas distâncias médias são o grupo G1 com sua distância média de 2,3 é o potreiro 6 com massa baixa dado pela variável MB_6 e o grupo G2 com uma distância 1,1 é o potreiro 4 com massa baixa dado pela variável (MB_4) e o potreiro 2 com massa baixa dado pela variável (MB_2). A distância média do grupo G3 é 2,3 dado pelo potreiro 3 com massa alta (MA_3) e o grupo G4 e G5 tem uma distância média 1,7 para os potreiros 5 e 1 dado pelas variáveis MA_5 e MA_1.

Na rotação varimax dado na figura 24 os grupos G3, G4 e G5 continuaram separados e formaram os grupos G1 e G2. Nas tabelas a seguir tem-se a matriz das correlações e matriz dos resíduos da população 7:

TABELA 24 –MATRIZ DE CORRELAÇÃO DA POPULAÇÃO 7

	MA_1	MB_2	MA_3	MB_4	MA_5	MB_6
MA_1	1					
MB_2	0,4885	1				
MA_3	0,6780	0,6670	1			
MB_4	0,4172	0,9212	0,7220	1		
MA_5	0,8029	0,2825	0,7308	0,2060	1	
MB_6	0,3960	0,5822	0,4914	0,7003	0,2047	1

As correlações da população 7 relativas produção animal igual ou acima de 70% possui correlação significativa dado na tabela 24.

TABELA 25 - MATRIZ DOS RESÍDUOS DA POPULAÇÃO 7

	MA_1	MB_2	MA_3	MB_4	MA_5	MB_6
MA_1	0					
MB_2	0,0100	0				
MA_3	-0,1154	-0,0308	0			
MB_4	-0,0173	0,0209	0,0359	0		
MA_5	-0,0598	-0,0034	0,0060	-0,0106	0	
MB_6	0,0401	-0,1667	-0,0753	-0,0953	0,0312	0

A matriz dos resíduos da população 7 possui na sua maioria resíduos abaixo de 15%, contribuindo assim para um bom ajuste do modelo.

Na tabela 26 são os fatores da população 8 variáveis medidas na pastagem (massa de forragem) através da rotação varimax e pelo método das componentes principais.

TABELA 26 - FATORES DA ROTAÇÃO VARIMAX E O MÉTODO DAS COMPONENTES PRINCIPAIS DA POPULAÇÃO 8

		Método das componentes		Rotação Varimax		
Potreiros	Variáveis	Y ₁	Y ₂	F ₁	F ₂	comunalidades
1	Massa alta	0,8618	0,1772	0,8708	-0,1261	0,7741
2	Massa baixa	0,9380	0,1986	0,9480	-0,1430	0,9192
3	Massa alta	0,9275	0,0311	0,9277	0,0236	0,8611
4	Massa baixa	0,8904	-0,1545	0,8798	0,2067	0,8167
5	Massa alta	0,0908	-0,9895	0,0323	0,9932	0,9874
6	Massa baixa	0,9841	-0,1427	0,9740	0,2005	0,9889
Autovalores		4,25	1,09	4,24	1,10	
Proporção Acumulada da Variância		70,87%	89,12%	70,69%	89,13%	

A população 8 é designada pelas variáveis medidas na pastagem (massa de forragem) que obteve dois fatores nos dois métodos dados na tabela 26 com seus autovalores maiores que 1. O modelo ajustado em conjunto no método das componentes principais explicou 89,12% e na rotação varimax 89,13%. As variáveis que melhor foram explicadas pelos fatores foi o potreiro 5 (massa alta) com 98,74% e o potreiro 6 (massa baixa) com 98,89% de suas variações captadas pelos fatores e a menor explicação foi o potreiro 1 (massa alta) 77,41% de sua variação explicada pelos fatores estudados. As demais variáveis apresentaram

comunalidades entre 81% e 92%, o que mostra que suas variações foram de modo geral, captadas de forma satisfatória pelos fatores ajustados.

Nas figuras 26 e 27 estão ordenados os fatores pelo método das componentes principais e a rotação varimax e na figura 28 tem-se o dendrograma da população 8.

FIGURA 26 –ORDENAÇÃO DAS 6 VARIÁVEIS PELO MÉTODO DAS COMPONENTES PRINCIPAIS DA POPULAÇÃO 8

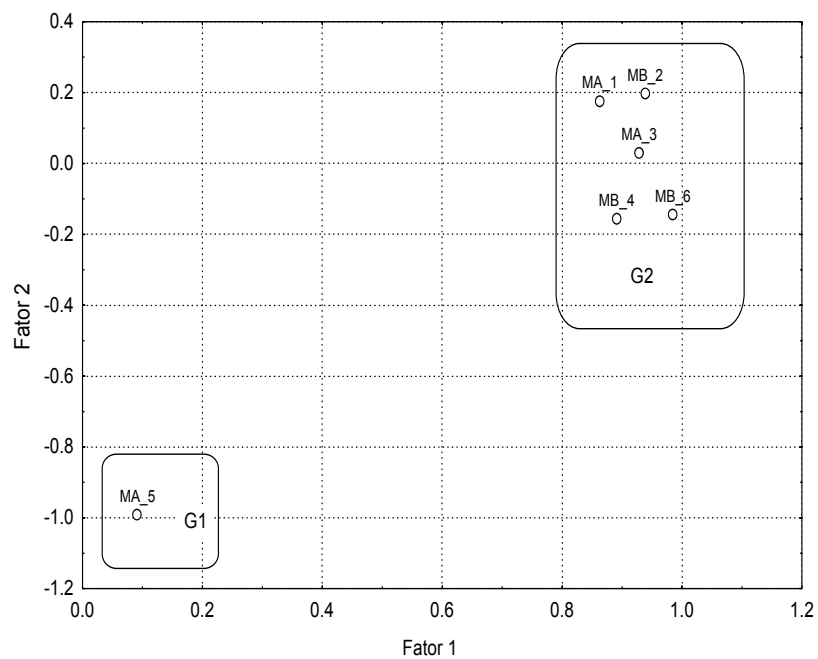


FIGURA 27 – ORDENAÇÃO DAS 11 VARIÁVEIS PELA ROTAÇÃO VARIMAX DA POPULAÇÃO 8

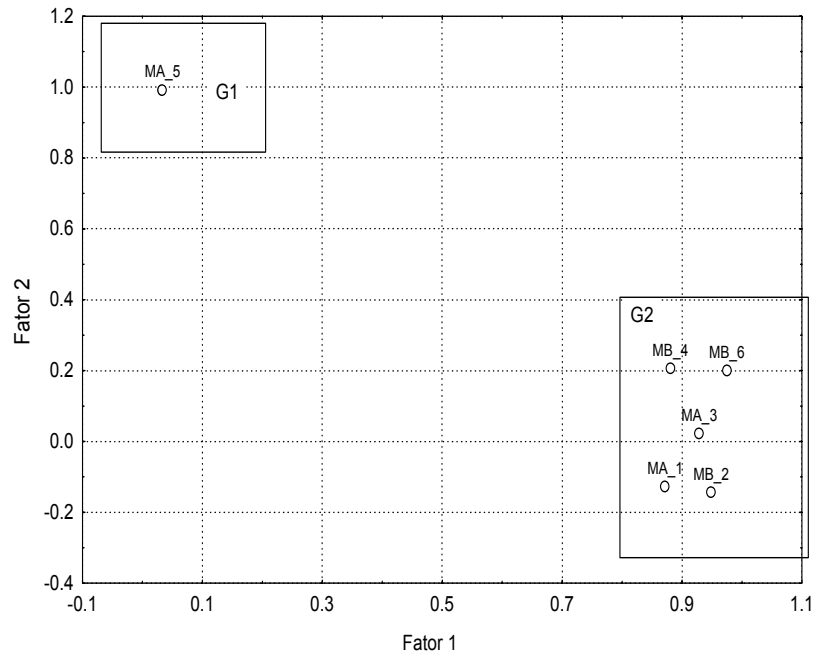
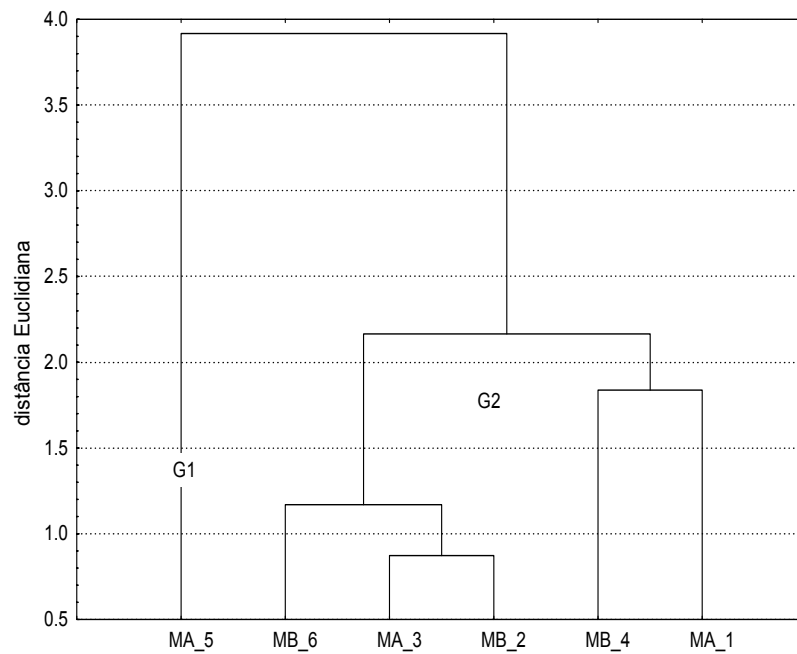


FIGURA 28 – DENDROGRAMA DA POPULAÇÃO 8



O método de componentes principais foi comparado com a análise de agrupamento sendo que os grupos formados são homogêneos segundo as suas similaridades (correlações) e suas dissimilaridades (distância média). Os grupos formados conforme suas distâncias médias são o grupo G1 com sua distância média de 3,9 é o potreiro 5 com massa alta dado pela variável MA_5 e o grupo G2 com uma distância média de 2,2 são os potreiros 6, 3, 2, 4 e 1.

Na rotação varimax os grupos permaneceram os mesmos. Nas tabelas seguintes são apresentadas as matrizes das correlações e dos resíduos:

TABELA 27- MATRIZ DE CORRELAÇÃO DA POPULAÇÃO 8

	MA_1	MB_2	MA_3	MB_4	MA_5	MB_6
MA_1	1					
MB_2	0,7587	1				
MA_3	0,6649	0,9455	1			
MB_4	0,7588	0,7193	0,7369	1		
MA_5	-0,0805	-0,0960	0,0597	0,1913	1	
MB_6	0,8192	0,9023	0,9093	0,8805	0,2327	1

Na matriz das correlações referente à população 8 que possuem igual ou acima de 70% tem correlação significativa.

TABELA 28 – MATRIZ DOS RESÍDUOS DA POPULAÇÃO 8

	MA_1	MB_2	MA_3	MB_4	MA_5	MB_6
MA_1	0					
MB_2	-0,0848	0				
MA_3	-0,1399	0,0694	0			
MB_4	0,0188	-0,0852	-0,0842	0		
MA_5	0,0166	0,0154	0,0062	-0,0424	0	
MB_6	-0,0037	0,0075	0,0010	-0,0179	0,0021	0

A maioria dos dados na matriz dos resíduos da população 8 têm resíduos abaixo de 15%, então se tem um bom ajuste do modelo.

Na tabela 29 estão os fatores referentes a população 9 pelo método das componentes principais e rotação varimax:

TABELA 29 - FATORES DA ROTAÇÃO VARIMAX E O MÉTODO DAS COMPONENTES PRINCIPAIS DA POPULAÇÃO 9

		Método das componentes		Rotação Varimax		
Potreiros	Variáveis	Y ₁	Y ₂	F ₁	F ₂	comunalidades
1	Massa alta	0,8726	-0,1645	0,8836	0,0880	0,7918
2	Massa baixa	0,9198	-0,3510	0,9814	-0,0777	0,9779
3	Massa alta	0,8581	-0,1817	0,8745	0,0674	0,7179
4	Massa baixa	0,0458	0,9389	-0,2206	0,9137	0,8798
5	Massa alta	0,2993	0,7419	0,0782	0,7962	0,5726
6	Massa baixa	0,5045	0,7081	0,2846	0,8215	0,7624
Autovalores		2,6898	2,1165	2,6443	2,1620	
Proporção Acumulada da Variância		44,83%	80,11%	44,07%	80,10%	

A população 9 é dada pelas variáveis medidas na pastagem (taxa de acúmulo) que obteve dois fatores nos dois métodos dados na tabela 20 com seus autovalores maiores que 1. O modelo ajustado em conjunto no método das componentes principais explicou 80,11% e na rotação varimax 80,10%. As variáveis que melhor foram explicadas pelos fatores foi o potreiro 2 (massa baixa) com 97,79% sua variação captada pelos fatores e a menor explicação foi o potreiro 5 (massa alta) 57,26% de sua variação explicada pelos fatores estudados. As demais variáveis apresentaram comunalidades entre 76% e 88%, o que mostra que suas variações foram de modo geral, captadas de forma satisfatória pelos fatores ajustados.

Nas figuras 29 e 30 estão ordenados os fatores pelo método das componentes principais e rotação varimax e na figura 31 o dendrograma da população 9.

FIGURA 29 –ORDENAÇÃO DAS 6 VARIÁVEIS PELO MÉTODO DAS COMPONENTES PRINCIPAIS DA POPULAÇÃO 9

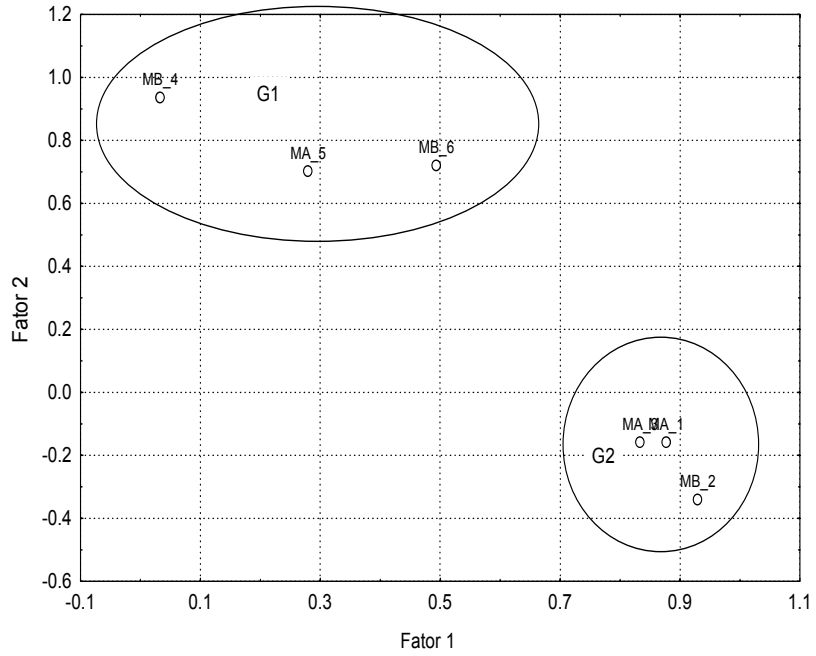


FIGURA 30 –ORDENAÇÃO DAS 6 VARIÁVEIS PELA ROTAÇÃO VARIMAX

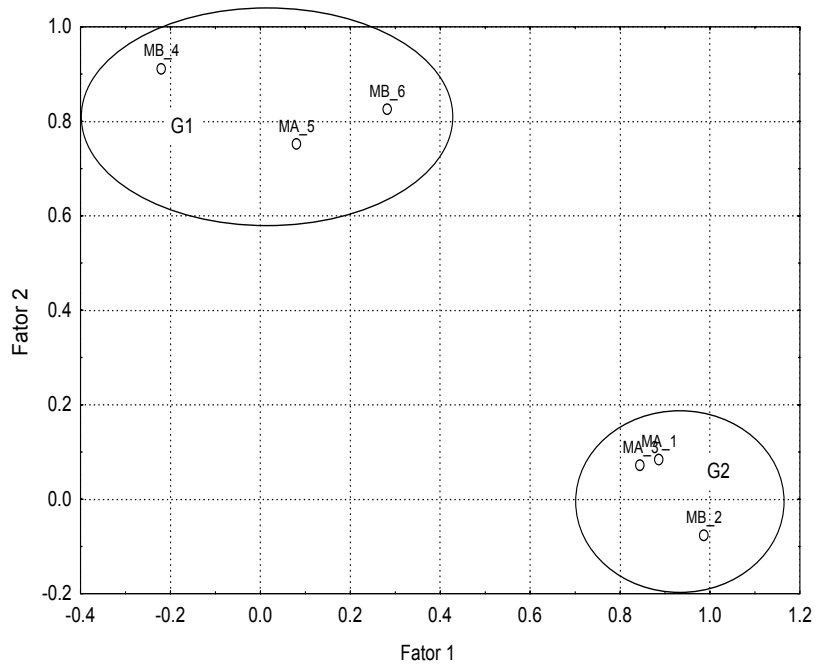
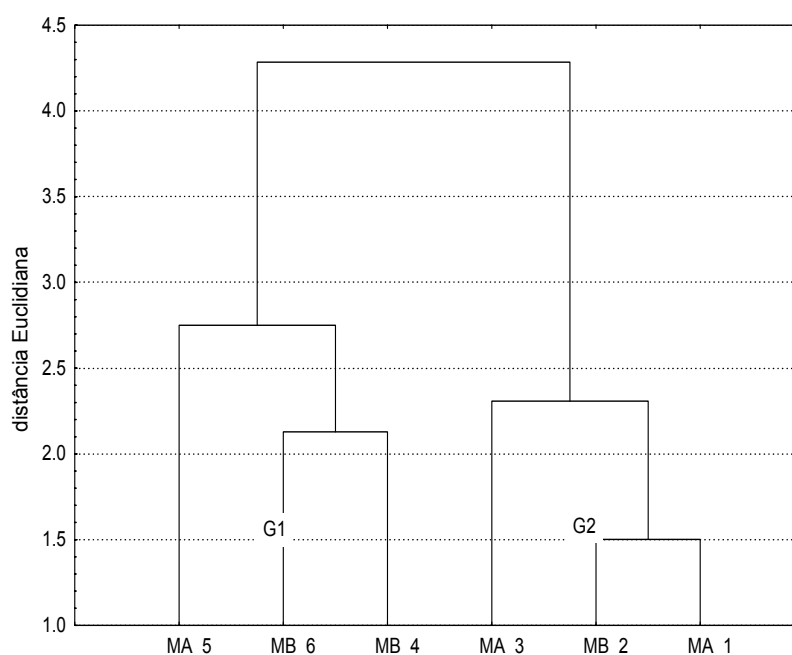


FIGURA 31 – DENDROGRAMA DA POPULAÇÃO 9



O método de componentes principais confirma a análise de agrupamento quanto a formação de grupos homogêneos segundo as suas similaridades (correlações) e suas dissimilaridades (distância média). Os grupos formados conforme suas distâncias médias são o grupo G1 com sua distância média de 2,6 e o grupo G2 com uma distância média de 2,4.

Na rotação varimax os grupos permaneceram os mesmos. Nas tabelas seguintes são apresentadas as matrizes das correlações e dos resíduos:

TABELA 30 – MATRIZ DE CORRELAÇÃO DA POPULAÇÃO 9

	MA_1	MB_2	MA_3	MB_4	MA_5	MB_6
MA_1	1					
MB_2	0,8389	1				
MA_3	0,6194	0,8155	1			
MB_4	-0,1474	-0,3115	-0,0264	1		
MA_5	0,3472	-0,0161	-0,0478	0,5810	1	
MB_6	0,1471	0,2805	0,3467	0,6763	0,4599	1

As correlações da população que estiverem iguais ou acima de 70% são significativas.

TABELA 31 – MATRIZ DOS RESÍDUOS DA POPULAÇÃO 9

	MA 1	MB 2	MA 3	MB 4	MA 5	MB 6
MA 1	0					
MB 2	-0,0281	0				
MA 3	-0,1345	-0,0112	0			
MB 4	-0,0290	-0,0246	0,0932	0		
MA 5	0,2133	-0,0371	-0,1698	-0,0875	0	
MB 6	-0,1720	0,0660	0,0488	-0,0152	-0,1844	0

A maioria dos resíduos está abaixo de 15% , então se aceita como um bom ajuste do modelo.

Na tabela 32 estão os fatores referentes a população 10 (carga animal) pelo método das componentes principais e rotação varimax:

TABELA 32 - FATORES DA ROTAÇÃO VARIMAX E O MÉTODO DAS COMPONENTES PRINCIPAIS DA POPULAÇÃO 10

		Método das componentes		Rotação Varimax		
Potreiros	variáveis	Y ₁	Y ₂	F ₁	F ₂	comunalidades
1	Massa alta	0,8988	-0,3437	0,9612	0,0443	0,9259
2	Massa baixa	-0,7657	-0,5450	-0,5559	0,7578	0,8833
3	Massa alta	0,8743	-0,3566	0,9420	0,0642	0,8915
4	Massa baixa	-0,2436	-0,8105	0,0231	0,8460	0,7163
5	Massa alta	0,9202	-0,1482	0,9202	-0,1481	0,8688
6	Massa baixa	-0,1526	-0,9332	0,1480	0,9339	0,8940
Autovalores		3,09	2,09	2,99	2,19	
Proporção Acumulada Da Variância		51,46%	86,26%	49,83%	86,36%	

A população 10 é dada pelas variáveis de animais (carga animal) que obteve dois fatores nos dois métodos dados na tabela 23 com seus autovalores maiores que 1. O modelo ajustado em conjunto no método das componentes principais explicou 86,26% e na rotação varimax 86,26%. As variáveis que melhor foram explicadas pelos fatores foi o potreiro 1 (massa alta) com 92,59% sua variação captada pelos fatores e a menor explicação foi o potreiro 4 (massa alta) 71,63% de sua variação explicada pelos fatores estudados. As demais variáveis apresentaram comunalidades entre 86% e 90%, o que mostra que suas variações foram de modo geral, captadas de forma satisfatória pelos fatores ajustados.

Nas figuras 32 e 33 estão ordenados os fatores pelo método das componentes principais e a rotação varimax e na figura 34 tem se o dendrograma da população 10.

FIGURA 32 – ORDENAÇÃO DAS VARIÁVEIS PELO MÉTODO DAS COMPONENTES PRINCIPAIS DA POPULAÇÃO 10

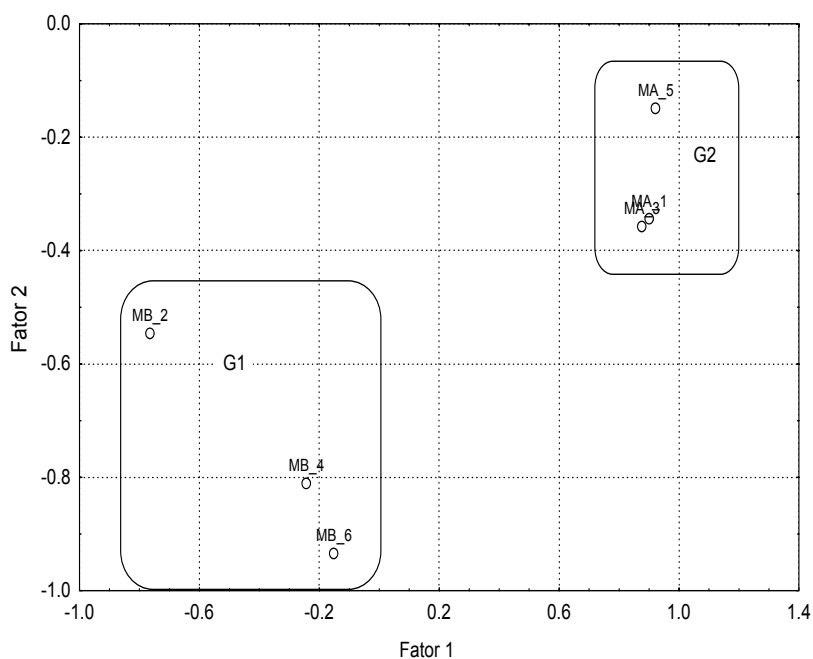


FIGURA 33 – ORDENAÇÃO DAS VARIÁVEIS PELA ROTAÇÃO VARIMAX DA POPULAÇÃO 10

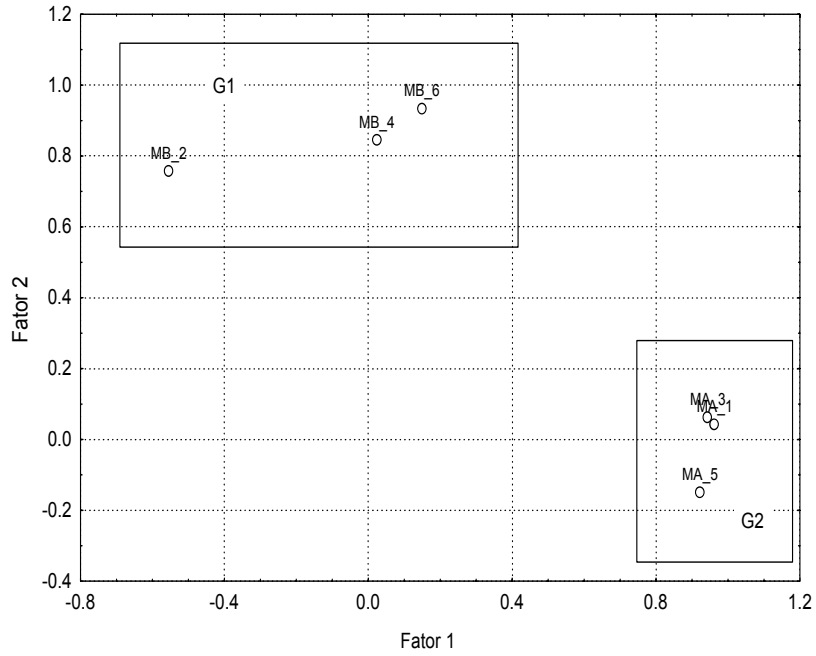
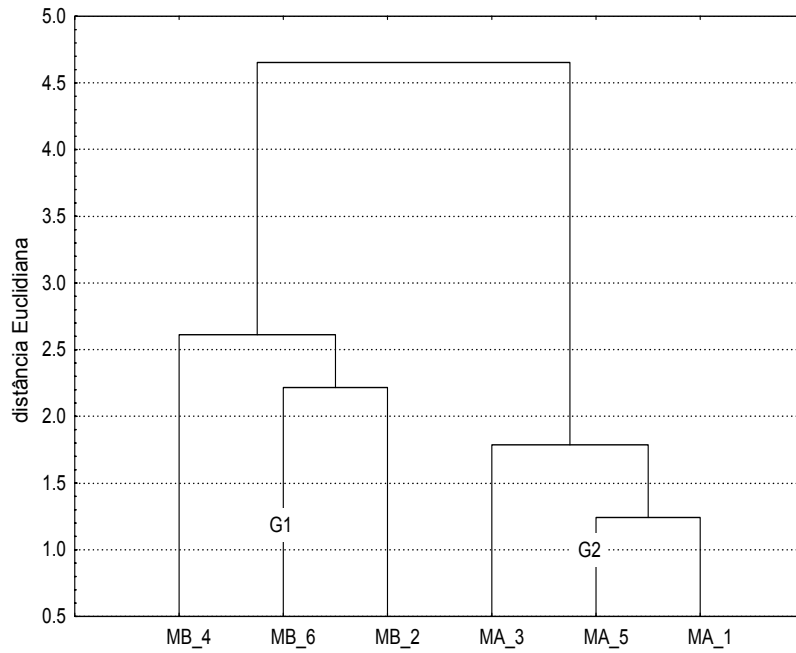


FIGURA 34 – DENDROGRAMA DA POPULAÇÃO 10



O método de componentes principais confirma a análise de agrupamento quanto a formação de grupos homogêneos segundo as suas similaridades (correlações) e suas dissimilaridades (distância média). Os grupos formados conforme suas distâncias médias são o grupo G1 com sua distância média de 2,6 e o grupo G2 com uma distância média de 1,6.

Na rotação varimax os grupos permaneceram os mesmos. Nas tabelas seguintes são apresentadas as matrizes das correlações e dos resíduos:

TABELA 33 – MATRIZ DE CORRELAÇÃO DA POPULAÇÃO 10

	MA_1	MB_2	MA_3	MB_4	MA_5	MB_6
MA_1	1					
MB_2	-0,4616	1				
MA_3	0,8465	-0,5228	1			
MB_4	0,0392	0,5127	0,0190	1		
MA_5	0,8899	-0,5463	0,7718	-0,0764	1	
MB_6	0,1730	0,6491	0,2718	0,6396	-0,0638	1

Correlações na matriz de dados da população 10 que estão iguais ou acima de 70% são significativas.

TABELA 34 – MATRIZ DOS RESÍDUOS DA POPULAÇÃO 10

	MA_1	MB_2	MA_3	MB_4	MA_5	MB_6
MA_1	0					
MB_2	0,0392	0				
MA_3	-0,0619	-0,0477	0			
MB_4	-0,0205	-0,1156	-0,0571	0		
MA_5	0,0119	0,0775	-0,0855	0,0276	0	
MB_6	-0,0107	0,0237	0,0724	-0,1539	-0,0618	0

Os resíduos dados na tabela 34 a maioria estão abaixo de 15%, portanto se aceita como um bom ajuste do modelo.

Na tabela 35 estão os fatores pelo método das componentes principais e pela rotação varimax da população 11 (taxa de lotação):

TABELA 35 - FATORES DA ROTAÇÃO VARIMAX E O MÉTODO DAS COMPONENTES PRINCIPAIS DA POPULAÇÃO 11

		Método das componentes		Rotação Varimax		
Potreiros	variáveis	Y ₁	Y ₂	F ₁	F ₂	comunalidades
1	Massa alta	0,7099	-0,5132	-0,8740	-0,0591	0,7673
2	Massa baixa	-0,8416	-0,1605	0,5446	0,6614	0,7340
3	Massa alta	-0,6249	-0,5630	0,1206	0,8325	0,7075
4	Massa baixa	-0,4490	-0,7054	-0,1056	0,8295	0,6992
5	Massa alta	0,6532	-0,5923	-0,8809	0,0379	0,7775
6	Massa baixa	-0,8458	0,0620	0,6901	0,4930	0,7192
Autovalores		2,94	1,45	2,34	2,07	
Proporção Acumulada Da Variância		49,11%	73,42%	38,97%	73,41%	

A população 11 é dada pelas variáveis de animais (taxa de lotação) que obteve dois fatores nos dois métodos dados na tabela 26 com seus autovalores maiores que 1. O modelo ajustado em conjunto no método das componentes principais explicou 73,42% e na rotação varimax 73,41%. As variáveis que melhor foram explicadas pelos fatores foi o potreiro 5 com 77,75% sua variação captada pelos fatores e a menor explicação foi o potreiro 4 com 69,92% de sua variação explicada pelos fatores estudados. As demais variáveis apresentaram comunalidades entre 70% a 77%, o que mostra que suas variações foram de modo geral, captadas de forma satisfatória pelos fatores ajustados.

Nas figuras 35 e 36 estão ordenados os fatores da população 11 pelo método das componentes principais e a rotação varimax, e na figura 37 o dendrograma.

FIGURA 35 - ORDENAÇÃO DAS VARIÁVEIS PELO MÉTODO DAS COMPONENTES PRINCIPAIS DA POPULAÇÃO 11

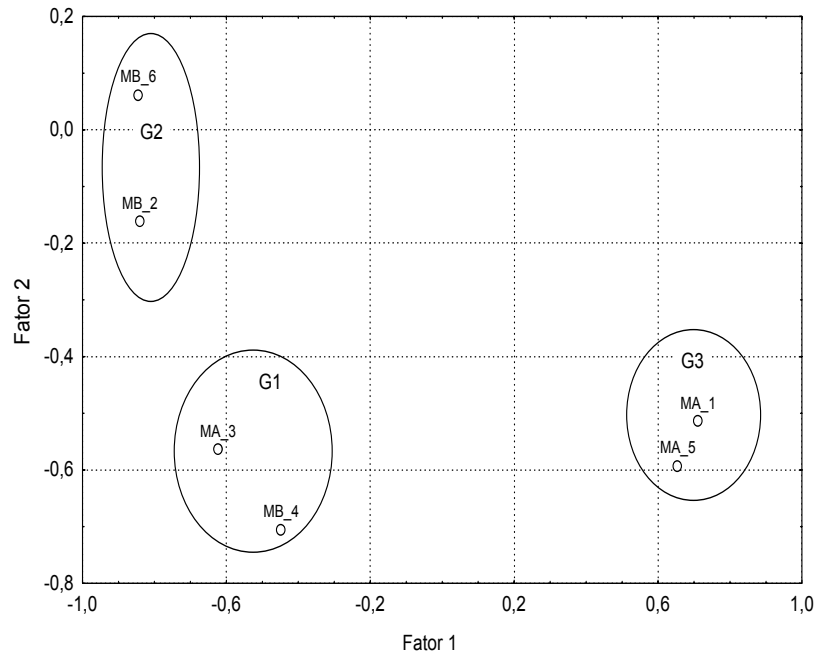


FIGURA 36 - ORDENAÇÃO DAS VARIÁVEIS PELA ROTAÇÃO VARIMAX DA POPULAÇÃO 11

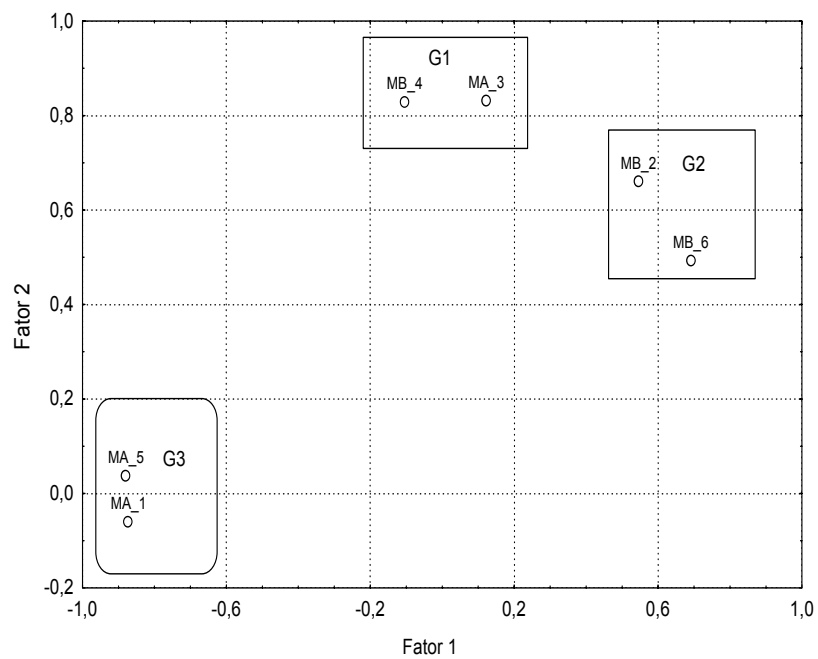
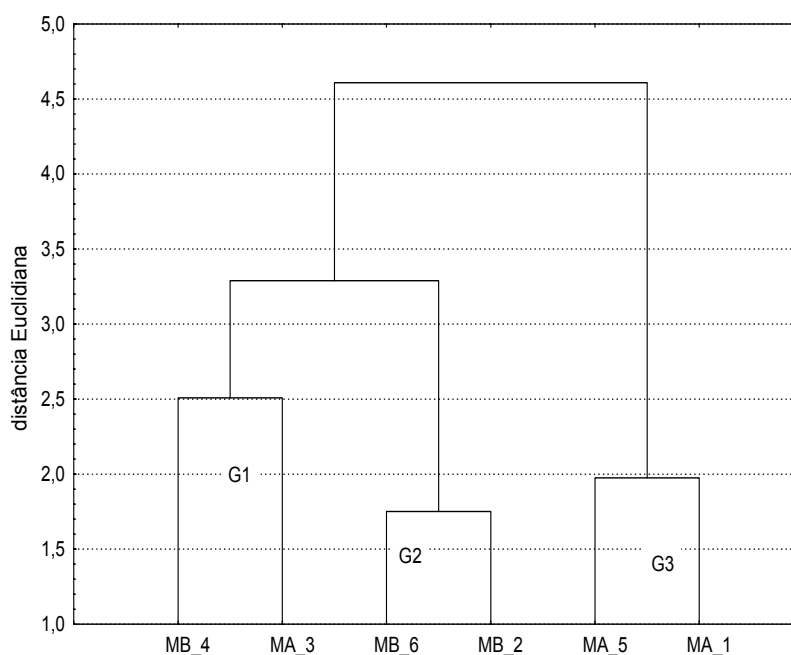


FIGURA 37 - DENDROGRAMA DA POPULAÇÃO 11



O método de componentes principais confirma a análise de agrupamento quanto à formação de grupos homogêneos segundo as suas similaridades (correlações) e suas dissimilaridades (distância média). Os grupos formados conforme suas distâncias médias são o grupo G1 com sua distância média de 2,5, o grupo G2 com uma distância média de 1,6 e grupo G3 com distância média de 2,0.

Na rotação varimax os grupos permaneceram os mesmos. Nas tabelas a seguintes são apresentadas as matrizes das correlações e dos resíduos:

TABELA 36 – MATRIZ DE CORRELAÇÃO DA POPULAÇÃO 11

	MA 1	MB 2	MA 3	MB 4	MA 5	MB 6
MA 1	1					
MB 2	-0,4666	1				
MA 3	-0,2150	0,4875	1			
MB 4	-0,0327	0,3567	0,5505	1		
MA 5	0,7212	-0,2779	-0,1255	-0,0219	1	
MB 6	-0,4362	0,7808	0,3834	0,2274	-0,5170	1

As correlações iguais ou acima de 70% são significativas.

TABELA 37 - MATRIZ DOS RESÍDUOS DA POPULAÇÃO 11

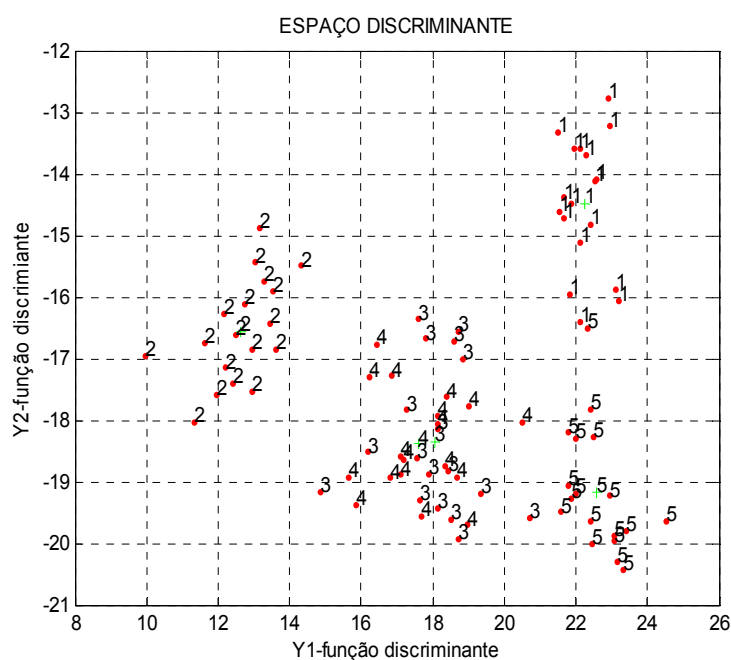
	MA 1	MB 2	MA 3	MB 4	MA 5	MB 6
MA_1	0					
MB_2	0,0484	0				
MA_3	-0,0603	-0,1288	0			
MB_4	-0,0759	-0,1344	-0,1273	0		
MA_5	-0,0465	0,1767	-0,0508	-0,1464	0	
MB_6	0,1961	0,0790	-0,1102	-0,1086	0,0722	0

Os dados da matriz dos resíduos da população 11 estão abaixo de 15%, se aceita como um bom ajuste do modelo.

4.4 ANÁLISE DISCRIMINANTE

Na figura 38 a classificação das populações 1 até 5 provenientes de resultados de análises de solos:

FIGURA 38 - ESPAÇO DISCRIMINANTE DAS 5 POPULAÇÕES



As amostras de cada população foram calculadas através das funções discriminantes Y1 e Y2 relacionadas na tabela 38. Quanto à alocação, a população 2 está bem definida e

quanto as populações 1, 3, 4 e 5 há necessidade de verificar a alocação de suas amostras. Na tabela seguinte estão os centróides de cada população.

TABELA 38 - CENTRÓIDES DAS 5 POPULAÇÕES

Populações	Centróides	
	1	22,2450
2	12,6289	-16,5584
3	18,0613	-18,3546
4	17,6149	-18,3943
5	22,5923	-19,1709

TABELA 39 - FUNÇÕES DISCRIMINANTES

Variáveis	Funções Discriminantes	
	Y1	Y2
pH	6,4630	-3,4911
MO	0,0208	0,0056
Al	-0,6560	1,7952
H+Al	-0,0418	-0,5794
Ca	-0,7939	0,4509
Mg	-0,06512	-0,2453
K	0,2515	-2,9168
P	-0,0043	0,1315
V	-0,1321	-0,1373
NH ₄	-0,1057	-0,0037
NO ₃	0,0076	0,0539

Nas tabelas seguintes têm-se as distâncias euclidianas de cada amostra aos centróides de cada população:

TABELA 40 – CLASSIFICAÇÃO DAS AMOSTRAS DA POPULAÇÃO 5

Amostras da população 5	População 1	População 2	População 3	População 4	População 5
1	5,1828	11,7557	5,5555	6,0533	0,9277
2	4,7319	11,3820	5,1999	5,7057	0,5126
3	5,6766	12,3648	6,1698	6,6827	1,5531
4	3,2269	10,4623	4,0566	5,0336	1,5329
5	4,7152	10,2225	4,0300	4,5123	0,7788
6	5,0696	10,9483	4,7570	5,2564	0,4840
7	3,7113	10,8779	4,8356	5,3620	1,0611
8	1,9407	10,6610	5,1975	5,6221	3,1460
9	4,5431	10,0154	3,8148	4,3266	0,9778
10	5,3667	11,8372	5,6242	6,1244	0,9578
11	5,8152	11,7268	5,5078	5,9990	1,3040
12	6,1021	12,4770	6,1781	6,6756	1,8245
13	4,6296	10,1331	4,1295	4,6294	0,6435
14	5,0178	9,9007	3,6870	4,1594	1,2970
15	5,5691	12,0954	5,8331	6,3269	1,3279
16	3,7844	10,1334	4,0345	4,5619	1,1238
17	3,7053	9,7684	3,6502	4,1876	1,5032
18	5,4649	11,1854	4,9905	5,4791	0,8926

Na tabela 40 estão as distâncias euclidianas de cada amostra e o centróide de cada população. A amostra 8 da população 5 que está alocada na população 1

TABELA 41 – CLASSIFICAÇÃO DAS AMOSTRAS DA POPULAÇÃO 4

Amostras da população 4	População 1	População 2	População 3	População 4	População 5
1	6,6630	4,3890	2,2884	1,9128	6,7369
2	6,3538	4,3887	4,0453	4,0773	8,4212
3	11,6946	4,9469	5,7670	5,1470	9,9126
4	4,5160	7,3429	1,3521	2,0243	3,7546
5	7,0717	5,3241	1,3279	0,9102	5,8202
6	8,2316	4,2459	2,5834	2,1208	7,2394
7	5,5615	7,3047	1,0166	1,6083	3,7947
8	5,4870	6,0290	0,5189	0,7993	4,8574
9	5,6052	6,0777	0,4780	0,6052	5,057
10	8,1161	4,6867	2,4711	1,9171	6,9780
11	5,7689	5,6481	0,8956	0,6390	5,4420
12	6,7463	5,3475	1,1510	0,5460	5,8480
13	6,8418	5,5227	1,0780	0,6991	5,5733
14	5,7071	6,7421	0,5892	1,0218	4,2032
15	6,2118	4,7483	1,7014	1,4401	6,2192
16	6,2141	4,7723	1,6211	1,3130	6,1475
17	5,6243	6,0447	0,3196	0,6318	4,8411
18	3,9704	9,0974	2,8050	3,5760	2,4991

Com relação às classificações das amostras da população 4 com as demais populações da tabela 41, as amostras 2, 4, 7,8 ,9, 14 e 17 foram alocadas na população 3 e a amostra 18 alocadas na população 5.

TABELA 42 – CLASSIFICAÇÃO DAS AMOSTRAS DA POPULAÇÃO 3

Amostras da população 3	População 1	População 2	População 3	População 4	População 5
1	5,6758	11,8747	5,5659	5,4149	1,0455
2	4,8770	6,7574	1,4204	1,6591	5,4284
3	6,1779	6,9515	0,7874	0,5349	5,1179
4	5,5640	8,1726	1,8222	1,7137	3,8961
5	6,8688	5,0405	1,2964	1,4771	6,9291
6	6,4089	7,4882	1,4923	1,2240	4,7749
7	6,5553	7,8319	1,9103	1,6407	4,5893
8	6,4040	6,5005	0,5058	0,2217	5,5580
9	4,3931	6,4202	2,0700	2,0740	4,8215
10	7,3432	4,8447	1,5754	1,1506	6,2835
11	6,5804	6,5606	0,7952	1,1072	4,6891
12	3,5083	6,9670	3,3380	3,3374	5,1813
13	5,5015	7,1299	0,9232	1,4217	3,8085
14	4,8100	6,3027	1,4981	1,5271	4,7023
15	7,4333	4,3633	1,9388	1,4514	6,6643
16	6,8845	6,0759	1,0569	0,2611	5,6545
17	5,5938	6,1027	1,7180	1,7248	5,8058
18	5,8978	5,6786	2,1400	1,9815	6,3032

Na tabela 42 a população 3 tem algumas de suas amostras como : 3, 4, 6 7,8, 10, 12, 15 , 16 e 18 foram classificadas como população 4., e a amostra 1 foi classificada na população 5.

Para avaliar a função de classificação das populações 1, 2, 3, 4 e 5, a taxa estimada de erro é calculada a partir da “matriz de confusão”:

		Classificado em					Total
		π_1	π_2	π_3	π_4	π_5	
População verdadeira	π_1	17	0	0	0	0	17
	π_2	0	18	0	0	0	18
	π_3	0	0	11	10	1	22
	π_4	0	0	7	8	1	16
	π_5	1	0	0	0	16	17
	total	18	18	18	18	18	90

As taxas aparentes de erro são: no total de classificação é de 22,22% e da população de 3 para 4 é 11,11% e da população 3 para 5 é de 1,11% e da população 4 para 3 é de 7,78%. Portanto, a população que teve o maior erro percentual de classificação foi a população de 3 para 4 .

5. CONCLUSÃO

As técnicas da análise multivariada aplicadas neste trabalho auxiliaram como uma das ferramentas fundamentais para as tomadas de decisão e previsão dos dados.

As populações de 1 até 5 na verificação da normalidade dados é uma aproximação adequada e útil da verdadeira distribuição normal multivariada, pois 50% dos seus dados estão em torno de uma linha reta, sendo que a maioria dos dados está centrada na média de sua população. As populações de 6 até 11 que são variáveis derivadas de animais e variáveis medidas na pastagem não obedecem a uma distribuição normal multivariada pois seus dados amostrais não estão centrados na média de sua população.

Na comparação entre vetores de médias das populações de 1 até 4 através do teste T^2 de Hotteling, ao nível de significância de 5%, as variáveis que tiveram diferença significativa foram matéria orgânica e amônia.

Com as técnicas de análise fatorial confirmou-se que a análise de agrupamento onde se pode observar as variáveis das populações estudadas em grupos quanto a correlação e distância média. As variabilidades permaneceram praticamente as mesmas com o método das componentes principais e rotação varimax, contribuindo assim para formação dos grupos de cada população.

Quanto a matriz dos resíduos a maioria dos dados das populações obtiveram resultados abaixo de 15% considerando um bom ajuste dos modelos.

Na análise discriminante diferenciaram-se algumas amostras de cada população, mesmo com as taxas de erro de classificação sendo consideradas baixas, em torno de 11,11%, 1,11% e 7,78%. Portanto não é necessário introduzir novas variáveis.

A técnica de análise de agrupamento nas populações de 6 até 11 referente às variáveis derivadas de animais e variáveis medidas na pastagem pode-se observar a existência de dependência entre a população 10 (carga animal) e a população 11 (taxa de lotação) e entre as populações 7 (produção animal) e 8 (massa de forragem).

REFERÊNCIAS

BARROSO, L.P. **Análise Multivariada**. 48ª Reunião da RBRAS e 10º SEGRO –7 a –Lavras MG. Departamento de Ciências e Exatas. Universidade Federal de Lavras.

BUENO E.A .C. et al. Pastagem Nativa Melhorada Sob Distintas Intensidades de Pastejo. In. IX Seminário de Iniciação Científica e Tecnológica, 2004. Curitiba PR **Anais...** Pato Branco. CEFET PR 2004. .p.70-73.

CARVALHO, W. A. et. al. Aplicação de Análise Multivariada na Discriminação de Unidades de Solo, Mediante Parâmetros Morfométricos de Bacias Hidrográficas, em Botucatu, SP. **Revista Brasileira de Ciências do Solo**. v.14. p.195-203. Campinas, 1990.

DANTAS, Adriana Cibele de Mesquita et al. Tolerância Ao Alumínio Em Porta Enxertos Somacionais De Macieira Cultivados Em Solução Nutritiva. **Pesquisa Agropecuária Brasileira**. Brasília, v.36, n. 4 , p 615-623 , abr 2001.

FERRAUDO, A. S. et. al. **Aplicação da Análise Multivariada em Atributos Diagnósticos dos Solos no Estado de São Paulo**. Disponível em < [http://www.sigfcav.hpg.ig.com.br / trabalho_7.htm](http://www.sigfcav.hpg.ig.com.br/trabalho_7.htm) > Acessado em: 08/03/2004.

FERREIRA, D. F. **Análise Multivariada**. Lavras.MG. Departamento de Exatas. apostila 400 f. Universidade Federal de Lavras. 1996.

GOMES, J.B.V. et. al. Análise de Componentes Principais de Atributos Físicos, Químicos e Mineralógicos de Solos do Bioma Cerrado. **Revista Brasileira do Solo**. v.28, p 137 -153, 2004.

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

LIMA, J.D. **A Análise Econômico-Financeira De Empresas Sob a Ótica Da Estatística Multivariada.** Curitiba. 2002.178 f.Dissertação (Mestrado em Métodos Numéricos em Engenharia)-Setores de Tecnologia e de Ciências Exatas, Universidade Federal do Paraná.

NANNI, Marcos Rafael. Análise Discriminante Dos Solos Por Meio De Uma Resposta Espectral No Nível Terrestre. **Pesquisa Agropecuária Brasileira** vol 39 no 10 Brasília oct 2004.

NETO, J.M. MOITA. **Estatística Multivariada Uma Visão Didática Metodológica.** Disponível em <<http://www.criticanarede.com>> Acessado em: 21/09/2005

VIDIGAL, M. C.G, et. al.Divergência Genética entre Cultivares de Mandioca por Meio de Estatística Multivariada. **Bragantia.** vol. 56. Campinas, 1997

ANEXOS

ANEXO 1 projeto referente a:

PRESSÕES DE PASTEJO EM CAMPO NATIVO MELHORADO SOBRE A PRODUÇÃO VEGETAL E ANIMAL NA REGIÃO “CAMPOS DE PALMAS”

Universidade Tecnológica Federal do Paraná

Centro Federal de Educação Tecnológica do Paraná (CEFET/UNED-PB)

1) IDENTIFICAÇÃO

Equipe executora

Orientador: Eng.º Agr.º Dr.: André Brugnara Soares (Coordenador) Professor adjunto do curso de agronomia do CEFET-PR- unidade de Pato Branco.

Eng.º Agr.º Dr.: Luis César Cassol (Co-orientador) Professor adjunto do curso de agronomia do CEFET-PR- unidade de Pato Branco.

Eng.º Agr.º Dra.: Tangriani Simeoni Assmann (Co-orientador) Professor adjunto do curso de agronomia do CEFET-PR- unidade de Pato Branco.

Eng.º Agr.º Dr.: João Alfredo Braidá (Co-orientador) Professor adjunto do curso de agronomia do CEFET-PR- unidade de Pato Branco.

Colaboradores: Acadêmicos do Curso de Agronomia:

Jean Carlos Mezzalira, acadêmico do 4º ano do curso de Agronomia do CEFET-PR – Unidade de Pato Branco (Colaborador).

Emanuel Antonio Centenaro Bueno, acadêmico do 4º ano do curso de Agronomia do CEFET-PR – Unidade de Pato Branco (Colaborador).

Leila Angela Tirelli, acadêmica do 4º ano do curso de Agronomia do CEFET-PR – Unidade de Pato Branco (Colaborador).

Haroldo Batistti Lorenzato, acadêmico do 4º ano do curso de Agronomia do CEFET-PR – Unidade de Pato Branco (Colaborador).

Cleimary de Fátima Zotti, acadêmica do 2º ano do curso de Agronomia do CEFET-PR – Unidade de Pato Branco (Colaboradora).

Luzia Vanessa Marceniuk, acadêmica do 4º ano do curso de Agronomia do CEFET-PR – Unidade de Pato Branco (Colaboradora).

O Centro Federal de Educação Tecnológica do Paraná (CEFET/UNED-PB) auxiliará, principalmente, com o fornecimento de infra-estrutura e recursos humanos para a realização do projeto. Também, como a distância do CEFET-PR até o experimento é de 190 km, teremos um gasto apreciável em gasolina que ficará por conta do CEFET-PR.

Parcerias

O produtor fornecerá toda a infra-estrutura da propriedade e ficará responsável pelos custos de manutenção dos animais (mineralização, vermífugos, vacinas, etc), cercas, bebedouros e maquinário.

RESUMO

O campo nativo caracteriza-se por ser um substrato pouco produtivo, principalmente no período frio, por apresentar, quase na sua totalidade, espécies de crescimento estival. Apesar de ser um patrimônio ecológico de valor inestimável, a pastagem natural vem perdendo área para cultivos agrícolas a cada ano. Desta forma, com introdução de espécies exóticas é possível manter o patrimônio natural e aumentar sua produção. O objetivo deste trabalho é identificar o melhor manejo da forragem, melhorar a distribuição da forragem ao longo do ano, avaliar o efeito de duas massas de forragem na produção animal e vegetal, usando duas repetições e identificar a melhor intensidade de pastejo. O trabalho será realizado no período compreendido entre junho de 2004 e junho de 2005, na propriedade do Sr. José Antonio Bueno, localizada no município de Água Doce, SC, região denominada “Campos de Palmas”. A área total destinada ao experimento será de aproximadamente 8 ha dividida em 6 poteiros. O trabalho propõe a quantificação do ganho de peso animal, e produção vegetal em função de distintos manejos da pastagem natural melhorada. O método de pastejo utilizado será o pastejo contínuo com carga variável usando o método "put and take" (Mott e Lucas, 1953). Espera-se confirmar a hipótese de que a massa de forragem a ser manejada influencia de forma marcante na produção vegetal e animal.

2) OBJETIVOS

Objetivo geral

Aumentar o conhecimento dos processos que atuam nesta vegetação, especialmente quando introduzidas espécies exóticas, e detectar respostas produtivas e ecológicas do campo natural melhorado frente a diferentes intensidades de pastejo.

Objetivos específicos

Identificar o melhor manejo da pastagem em termos de intensidade de pastejo, em relação à produção animal, tanto por área quanto individual.

Identificar a capacidade real de suporte da pastagem que permita que ela permaneça vários anos produzindo de forma econômica, social e ecologicamente sustentável;

Verificar o efeito da carga animal nos parâmetros do solo como densidade e resistência à penetração;

Verificar o efeito do nível alimentar nas características reprodutivas dos animais;
Verificar a melhoria na produção de forragem através da introdução das espécies exóticas.

3) IDENTIFICAÇÃO E CARACTERIZAÇÃO DO PROBLEMA

Na região sul do Brasil existe aproximadamente 15 milhões de hectares de pastagem natural que representam a base alimentar dos herbívoros domésticos. Apesar da produção animal em pastagem nativa ser uma atividade sustentável sob o ponto de vista ecológico, supostamente dependente do manejo, atualmente existe, uma pressão social (aumento das famílias e subdivisão das propriedades) para que se eleve o índice de produção animal. Neste contexto, o campo nativo é freqüente, e erroneamente, rotulado como um substrato pouco produtivo, e vem perdendo espaço para lavouras anuais, permanentes e pastagens cultivadas, especialmente nos últimos três anos.

Existem vários fatores que contribuem para esta diminuição de área, dentre os quais ressaltamos os baixos índices produtivos e a baixa rentabilidade dos rebanhos criados em pastagem natural.

Esse substrato foi, por muito tempo, negligenciado pela pesquisa devido, em parte, à grande complexidade envolvida na avaliação de suas características, causada pela grande heterogeneidade botânica, estrutural e de solos, e também pela cultura ainda existente no nosso meio científico de menor valorização dos germoplasmas nativos frente aos exóticos.

Baseado nesta realidade, esta formação campestre pode ser vislumbrada como um grande substrato para produção animal de produto diferenciado, atingindo mercados seletos de consumo. Como se não bastasse essa ótima perspectiva, as pastagens nativas do sul do Brasil possuem espécies de bom valor forrageiro, e que vem resistindo há anos sob condições de manejo muitas vezes adversas, forçando a uma desestabilização do sistema. Já existem informações de que a produção animal pode ser, no mínimo, duplicada simplesmente pelo manejo correto da desfolha.

Portanto, esta formação campestre clama por maior entendimento de seus processos ecológicos para que possamos alterar os índices produtivos e manter sustentável este ecossistema.

Uma das maneiras de aumentar a produtividade animal do campo nativo mantendo a sustentabilidade ecológica do sistema é o melhoramento através da correção do solo com adubação e calagem e principalmente a introdução de espécies que produzam especialmente nos meses de vazio forrageiro (inverno e início de primavera).

Este é o principal problema relacionado à produção animal em campo nativo, principalmente na região dos “Campos de Palmas”, baixos índices produtivos no inverno, devido a grande ocorrência de geadas e a não existência de algumas espécies nativas de inverno. Com o início da estação fria ocorre redução de produção e qualidade e, conseqüentemente, há perda de peso dos rebanhos que ocupam essas pastagens. Aliado a isso, os pecuaristas, com a necessidade de produzir mais, aumentam a lotação dos campos sem respeitar um manejo adequado. Diante dessa situação, o pastoreio intensificou-se, o pisoteio aumentou, favorecendo a infestação de espécies indesejáveis nos pastos, diminuindo a qualidade, dando início ao processo de degradação.

Para BARRETO et al. (1986), o melhoramento da pastagem natural via sobre-semeadura de espécies tem se mostrado uma alternativa de grande importância para aumentar-se o rendimento, por envolver baixos custos, manter a estrutura física do solo e não eliminar as espécies nativas, que podem contribuir para a melhoria da composição da forragem.

Além de influenciar de forma contundente no rendimento do cultivo de verão, a maneira de manejar a pastagem, mais especificamente, a massa de forragem mantida, é um dos principais fatores que determinam a produtividade animal tanto por área quanto individual. Existe uma série de trabalhos científicos que visaram delinear uma relação funcional entre a intensidade de pastejo e a produção animal sobre ela.

A presença de leguminosa na pastagem é de extrema importância quando analisamos o sistema de produção como um todo e também quando consideramos o preço do fertilizante nitrogenado sintético e sua oscilação de preço. Em outras palavras, o produtor não deve adotar um sistema de produção animal a pasto com alta dependência de fertilizantes químicos, especificamente o nitrogênio, por tratar-se do nutriente de custo mais elevado e de preço muito variável com o mercado internacional.

4) METODOLOGIA

Período de execução

O trabalho será realizado no período compreendido entre junho de 2004 e junho de 2005.

Local de execução

O experimento será realizado na propriedade do Sr. José Antonio Bueno, localizada no município de Água Doce, SC, região denominada “Campos de Palmas”. A área total destinada ao experimento será de aproximadamente 8 ha dividida em 6 poteiros.

O projeto propõe a quantificação da produção animal, entre outras variáveis respostas, em função de distintos manejos da pastagem natural melhorada. A pastagem será campo nativo sobre-semeado com espécies de inverno, dentre elas: aveia preta (*Avena strigosa*), azevém (*Lolium multiflorum* Lam.), trevo branco (*Trifolium repens*), trevo vermelho (*Trifolium pratense*) e cornichão (*Lotus corniculatus*). Essas espécies foram eleitas para o estudo em função da sua representatividade na região e também por indicativos de estudos preliminares que apontam essas espécies como promissoras para produção animal e consórcio com gramíneas.

As intensidades de pastejos constarão de duas massas de forragem no manejo da pastagem, sendo elas, 1000 e 2000kg de Matéria Seca (MS)/ha. O método de pastejo a ser adotado será o contínuo com lotação variável, utilizando a técnica “put-and-take” (MOTT e LUCAS, 1952).

Nesse período, serão quantificadas variáveis relativas ao solo, à planta e ao animal, com o objetivo de gerar informações e também de explicar os resultados obtidos. Sempre no intuito de buscar uma abordagem sistêmica e interdisciplinar do tema em questão.

Serão avaliadas variáveis de densidade de solo, para a estimativa do grau de compactação e teor de nitrogênio e matéria orgânica. Esses parâmetros são fundamentais de serem avaliados.

Serão coletadas amostras a cada vinte e oito dias durante todo o período experimental. As análises serão feitas no laboratório de análise de solos do CEFET – PR de Pato Branco.

Também será avaliado, a campo, a penetrabilidade do solo, um outro indicador de compactação. Esses dados serão posteriormente correlacionados com a carga animal, massa de forragem e produção animal obtida.

Será realizada a cada 28 dias a avaliação dos parâmetros relacionados à pastagem, massa de forragem através do método de dupla amostragem (HAYDOCK e SHAW, 1975).

Será feita através do método Botanal (TOTHILL et al., 1992) a composição botânica. Essa medida é importante pois o objetivo de toda mescla forrageira é manter os componentes desejáveis de forma relativamente constante através dos anos. A contribuição de leguminosas também é importante porque é uma informação usada posteriormente na estimativa da quantidade de nitrogênio fixado.

Será usada a técnica de gaiolas emparelhadas de exclusão ao pastejo (KLINGMAN et al., 1943), para medir a taxa de acúmulo de matéria seca, num total de três gaiolas por unidade experimental (potreiro). Será feita a avaliação de composição botânica (botanal) dentro e fora da gaiola para estimar a produção de forragem, não só total, mas de cada espécie na mistura.

Será avaliada com o uso de uma trena, a altura da pastagem, pois trata-se de uma informação que, embora seja de uso limitado na questão científica, é bem aceita e assimilada pelos produtores.

Será medido o ganho médio diário (GMD) pela diferença do peso dos animais, no início e fim de cada subperíodo, dividido pelo número de dias transcorridos no subperíodo (28 dias).

Será medido o ganho de peso vivo por hectare (GPV/ha), obtido multiplicando o ganho médio diário dos animais pelo número de dias/animal/ha em cada subperíodo.

A cada 28 dias será realizada a avaliação do estado corporal dos animais, num rank que varia de 1 (muito magra) a 5 (muito gorda).

Delineamento experimental e análise estatística

O delineamento experimental utilizado foi de blocos completamente casualizados (BCC), com três repetições por tratamento.

Os dados serão analisados primeiramente pela análise de variância para verificar a existência ou não de diferenças significativas ($P < 0,05$) entre os tratamentos e da possível interação ($P < 0,05$) entre os fatores. Caso ocorra interação significativa, todos os tratamentos serão analisados juntos, comparando-os através de um método de comparação múltipla (DMS). Caso não ocorra significância na interação entre os fatores, cada fator será analisado separadamente.

5) INFRA-ESTRUTURA DISPONÍVEL

O experimento, por constar de duas massas e três repetições, terá seis poteiros totalizando 8 hectares de área com 1600 metros de cerca que a delimita. A cerca terá dois fios de arame, totalizando 3200 m de arame. Além disso, serão usados para confecção da cerca, centenas de isoladores que são colocados nos mourões para fixação dos arames, porteiras e 200 mourões de madeira de lei.

Os animais utilizados serão novilhas de sobreano, da raça Blond d' Aquitaine. Será utilizado três animais testers por poteiro mais oito animais disponíveis para regular a massa de forragem, totalizando 26 animais.

O manejo com os animais envolve uma série de estruturas, como, saleiros, balança, tronco de contenção, seringas, etc. Todas essas estruturas serão providenciadas para o presente experimento.

Para a observação diária dos animais e cuidados com fornecimento de água, sal e a permanência dos animais nos respectivos poteiros será necessário um peão.

A maquinaria necessária para o plantio e manutenção da pastagem (roçadeira, pulverizadores) também fica por responsabilidade do produtor. Basicamente serão usadas uma semeadora de plantio direto da marca Fundiferro para plantio de espécies de sementes pequenas, uma roçadeira, um trator Massey Ferguson 275 e um pulverizador para controle de plantas invasoras.

Para o manejo dos animais também serão necessários uso de cavalos e arriamento, também incluídos na estrutura necessária.

Para a coleta de sangue dos animais para determinação de progesterona, uma avaliação extra ao projeto, serão usados tubos de ensaio para armazenamento do sangue e um freezer para manutenção do material até a análise. Convém salientar que essa determinação não está contemplada no projeto, pois sua realização fica na dependência de disponibilidade de laboratório do Curso de Medicina Veterinária da Universidade de São Paulo, que fará as análises.

6) RESULTADOS ESPERADOS

No estado do Paraná, as pastagens naturais estão desaparecendo de forma intensa a cada ano pela substituição por cultivos agrícolas como soja e milho e culturas permanentes como *Pinus*. A única região do Paraná que ainda mantém suas pastagens naturais de forma relativamente preservada é a região dos campos de altitude, chamado ecossistema Campos de Palmas. A preservação desse substrato natural é de vital importância, pois existe uma tendência mundial de demanda de alimentos produzidos de forma “limpa”, preservando os recursos naturais. Desta forma, as pastagens naturais podem ser vislumbradas como um grande alibi para a comercialização dos produtos de origem animal, podendo-se em breve agregar valor ao produto pela qualidade de sua origem.

Apesar dessa excelente perspectiva para a pastagem natural, ela ainda é pouco produtiva, principalmente se submetida a manejos tradicionais, e ainda não cumpre sua função social, pois a empregabilidade de pessoal em sistema pecuários baseados em campo nativo é mínima.

Baseado nesse contexto, a presente proposta poderá, ao seu final, propor uma forma de manejo da pastagem natural, que revolucione os sistemas de produção da região, aumentando a produção, evitando que o campo seja substituído pois torna-se atraente economicamente, aumentando assim a renda e qualidade de vida do produtor.

Os resultados desse experimento serão publicados em congressos, prioritariamente a Reunião anual da Sociedade Brasileira de Zootecnia, sendo que o coordenador do projeto é membro desta entidade. Pretende-se redigir três artigos na íntegra e publicar em periódicos

nacionais. Também serão publicados resumos expandidos em eventos científicos. Os dados também serão repassados aos extensionistas da EMATER (órgão oficial de extensão rural do Paraná) através de curso de treinamento e palestras. Para os produtores será organizados Dias-de-campo e palestra para treinamento em manejo de pastagem natural.

7) REFERENCIAS BIBLIOGRAFICAS

BARRETO, I.L.; VICENZI, M.L.; NABINGER, C. Melhoramento e renovação de pastagens. In: PEIXOTO, A.M.; MOURA, J.C.; FARIA, V.P. **Pastagens: fundamentos de exploração racional**: FEALQ, 1986. p. 295-309.

HAYDOCK, K.P.; SHAW, N.H. The comparative yield method for estimating dry matter yield of pasture. **Austr. J. Agr. And Aim. Husb.**, Melbourne. v. 15, p.66-70, 1975.

KLINGMAN, D.L.; MILES, S.R.; MOTT, G.O. The cage method for determining consumption and yield of pasture herbage. **Journal of the Animal Society of Agronomy**, Geneva, v. 35, p. 739-746, 1943.

LOBATO, J.F.P. BARCELOS, J.O. 1992. Efeito da utilização de pastagens melhoradas no pós-parto e no desmame aos 100 ou 180 dias de idade no desempenho reprodutivo de vacas de corte. **Revista SBZ**. 21(3): 385-395. 1992.

MOHRDIECK, K.H. 1980. Formações campestres do Rio Grande do Sul. In: SEMINARIO SOBRE PASTAGENS, Porto Alegre, 17-19 junho 1980. “De que pastagens necessitamos”. FARSUL. 1980. P. 18-27.

MOLETTA, J.L. 2001. Avaliação para diferentes sistemas para uso de culturas de inverno em um sistema de Integração agricultura-pecuária. Dados parciais ainda não publicados, IAPAR-estação Experimental Fazenda Modelo, Ponta Grossa - PR, 2001.

MOTT, G.O.; LUCAS, H.L. The design, conduct, and interpretation of grazing trials on cultivated and improved pastures. In: INTERNATIONAL GRASSLAND CONGRESS, 6, 1952, Pensylvania. **Proceedings...** Pensylvania: [s.n.], 1952. p.1380-1385.

NABINGER, C. & PAIM, N.R. Alternativas para o uso de espécies forrageiras de produção hibernal. **Revista Lavoura Arrozeira**, v. 38, n.360, p. 47-54, 1985.

PELLISSARI, A. et al., Manejo de plantas daninhas em sistemas de integração lavoura-pecuária. In: **Anais do I Encontro de Integração Lavoura Pecuária no Sul do Brasil /**

editores Nilvânia Aparecida de Mello e Tangriani Simioni Assmann, Pato Branco: CEFET-PR, 2002.

TOTHILL, J.C.; HARGRAVES, J.N.G.; JONES, R.M. et al. BOTANAL – A comprehensive sampling and computing procedure for estimating pasture yield and composition. 1. Field sampling. **Tropical Agronomy**, Queensland, v.78, 24p. 1992.

VICENZI, M.L. **Estabelecimento de leguminosas tropicais consorciadas ou não com capim de Rhodes, introduzidas em pastagem natural com preparo superficial do solo.** 1974. 166 p. (Dissertação de Mestrado) – Universidade Federal do Rio Grande do Sul, 1974.

ANEXOS

ORÇAMENTO

Orçamento		
Discriminação	Valor (R\$)	Responsável
Adubação	1.380,00	Fundação Araucária
Calagem	2.800,00	Fundação Araucária
Implantação pastagem	2.800,00	Fundação Araucária
Eletrificador	1.000,00	Fundação Araucária
TOTAL	7.980,00	Fundação Araucária

Antes da implantação do experimento será realizada uma caracterização da área experimental, corrigida a acidez e a deficiência nutricional para posteriormente se introduzir os tratamentos.

A recomendação de adubação e calagem foi feita segundo as recomendações oficiais, Constando de trinta sacas da fórmula 5-20-20 e quarenta toneladas de calcário. O custo de transporte dos insumos não está sendo incluído no custo pois ficará por conta do produtor.

A implantação da pastagem consta do custo de operação da máquina de plantio direto mais o trator, sem a inclusão nesta proposta, o custo de herbicidas e roçadas.

O custo das sementes foi feito da seguinte forma:

Componente	Custo/unidade	Unidade/ha	Custo/ha	ha	Total
aveia	0,5	60	30	8	240
Azevém	0,8	30	24	8	192
Trevo branco	14	4	56	8	448
Cornichão	14	10	140	8	1120
Trevo vermelho	14	5	70	8	560
Semeadura	30			8	240
mais trator					
Total					2800

O eletrificador, em função da dificuldade de rede de energia elétrica próxima ao local do experimento, será optado pela compra de um aparelho solar. Este equipamento será usado para energizar os arames da cerca para evitar que animais de diferentes lotes misturem-se.

Orçamento contrapartida:		
Discriminação	Valor (R\$)	Responsável
Gasolina	1.820,00	CEFET-Pato Branco
Gaiolas de exclusão ao pastejo	1.440,00	CEFET-Pato Branco
Automóvel	Sem custo	CEFET-Pato Branco
Hora/máquina	900,00	Produtor
Materiais para poteiros	998,00	Produtor
Mão-de-obra diária	3.000,00	Produtor
Vermífugo		Produtor
Mangueira e brete de manejo		Produtor
Cerca	3.200,00	Produtor
TOTAL	11.358,00	

**ANEXO 2 POPULAÇÃO 1 -MATRIZ DE DADOS ORIGINAL
ANÁLISE DE SOLO DA CARACTERIZAÇÃO DA ÁREA**

**Projeto Guamirim / Água Doce - SC Análise de Solo 30/08
CARACTERIZAÇÃO DA ÁREA**

100

Bloco	Identificação	Prof.	pH	MO	Al⁺³	H+Al	Ca	Mg	K	P	V	NH4	NO3
		Cm	CaCl ₂	g dm ⁻³cmol _c ⁽⁺⁾ dm ⁻³					mg dm ⁻³		mg/kg	mg/kg
1	Trincheira	0-5	4,20	53,61	2,20	14,44	2,70	2,20	0,23	16,94	0,26	4,20	57,87
1	Trincheira	5-10	4,00	53,61	3,80	15,77	0,63	1,14	0,10	0,93	0,11	4,20	36,17
1	Trincheira	10-20	4,00	0,00	4,05	13,06	0,40	0,87	0,08	0,93	0,09	4,20	20,67
1	Composta	0-5	4,2	53,6	3,1	15,8	1,5	1,4	0,1	10,6	0,16	3,6	20,7
1	Composta	5-10	4,0	53,6	4,1	17,2	0,6	1,3	0,1	0,9	0,11	1,8	13,4
1	Composta	10-20	4,0	46,9	4,7	15,2	0,5	0,7	0,1	0,9	0,08	2,4	20,7
2	Trincheira	0-5	4,00	56,61	4,24	15,77	0,66	0,94	0,10	1,25	0,10	1,80	33,07
2	Trincheira	5-10	4,00	44,23	4,09	13,06	0,35	0,55	0,05	0,61	0,07	2,10	32,03
2	Trincheira	10-20	4,00	40,21	3,75	13,06	0,33	0,67	0,05	0,61	0,07	2,70	35,13
2	Composta	0-5	4,00	53,61	3,11	14,44	1,18	1,72	0,10	6,13	0,17	3,00	24,80
2	Composta	5-10	3,90	53,61	4,13	15,77	0,53	1,17	0,10	1,25	0,10	2,70	20,67
2	Composta	10-20	3,90	53,61	5,10	18,82	0,40	0,50	0,08	0,61	0,05	1,80	21,70
3	Trincheira	0-5	4,00	60,31	4,58	18,82	0,90	1,17	0,15	1,57	0,11	1,80	21,70
3	Trincheira	5-10	3,90	53,61	4,53	15,77	0,32	0,58	0,08	0,93	0,06	2,40	10,33
3	Trincheira	10-20	3,90	46,91	4,25	14,08	0,20	0,40	0,05	0,00	0,04	2,10	16,53
3	Composta	0-5	4,20	60,31	2,25	14,44	2,30	2,00	0,18	15,36	0,24	2,10	26,87
3	Composta	5-10	4,00	60,31	5,07	17,23	0,75	0,85	0,13	1,57	0,09	2,10	31,00
3	Composta	10-20	3,90	60,31	5,15	18,82	0,40	0,50	0,10	0,61	0,05	1,80	25,83

Fonte: Laboratório de Solos da Unidade do Cefet Pato Branco

ANEXO 3 POPULAÇÃO 2 – MATRIZ DE DADOS ORIGINAL

Projeto Guamirim / Água Doce - SC Análise de Solo 04/10

Potreiro	MASSA	Bloco	Prof.	pH	MO	Al ³⁺	H+Al	Ca	Mg	K	P	V	NH4	NO3
1	ALTO	1	0-5	4,40	87,11	0,61	13,22	7,09	2,46	0,50	15,36	0,43	103,80	31,00
3	ALTO	2	0-5	4,10	79,07	1,63	13,22	3,90	2,60	0,35	8,91	0,34	98,04	35,13
5	ALTO	3	0-5	4,00	80,41	2,63	15,77	2,90	3,15	0,50	13,86	0,29	81,00	31,00
1	ALTO	1	5-10	3,90	80,41	3,47	18,82	1,90	2,00	0,25	8,50	0,18	78,00	26,87
3	ALTO	2	5-10	3,80	67,01	4,18	20,55	0,90	1,42	0,20	2,23	0,11	82,50	26,87
5	ALTO	3	5-10	3,80	54,95	4,65	18,82	0,90	2,16	0,28	2,56	0,15	66,00	28,93
1	ALTO	1	10-20	3,90	60,31	3,35	18,82	2,47	2,60	0,15	1,57	0,22	66,60	24,80
3	ALTO	2	10-20	3,80	60,31	4,85	20,55	0,70	0,90	0,15	1,25	0,08	69,00	24,80
5	ALTO	3	10-20	3,70	67,01	5,22	20,65	0,70	1,10	0,20	1,57	0,09	63,00	25,83
2	BAIXO	1	0-5	4,30	77,73	1,10	14,44	5,90	2,70	0,45	21,61	0,39	112,50	34,10
4	BAIXO	2	0-5	4,20	73,71	1,29	14,44	4,90	2,43	0,43	34,86	0,35	82,80	31,00
6	BAIXO	3	0-5	4,20	83,09	1,75	14,44	5,30	2,40	0,53	10,18	0,36	87,30	32,03
2	BAIXO	1	5-10	3,90	69,69	3,47	18,82	1,90	2,40	0,23	2,56	0,19	83,40	27,90
4	BAIXO	2	5-10	3,90	71,03	3,75	18,82	1,60	2,30	0,20	3,25	0,18	69,00	29,97
6	BAIXO	3	5-10	3,80	58,97	4,75	20,55	1,00	1,55	0,23	2,23	0,12	71,10	26,87
2	BAIXO	1	10-20	3,90	60,31	3,48	17,23	1,90	2,00	0,18	1,57	0,19	70,20	27,90
4	BAIXO	2	10-20	3,80	53,61	4,95	20,55	0,90	1,70	0,20	1,90	0,12	63,30	26,87
6	BAIXO	3	10-20	3,80	61,65	4,65	18,82	0,70	1,00	0,18	1,57	0,09	63,00	21,70

Fonte: Laboratório de Solos da Unidade do Cefet Pato Branco

ANEXO 4 POPULAÇÃO 3 – MATRIZ DE DADOS ORIGINAL

Coleta 13-12-03

Potreiro	MASSA	Bloco	Prof.	pH	MO	Al ³⁺	H+Al	Ca	Mg	K	P	V	NH4	NO3
			Cm	CaCl ₂	g dm ⁻³cmol _c ⁽⁺⁾ dm ⁻³					mg dm ⁻³		mg/kg	mg/kg
1	ALTO	1	0-5	5,50	67,01	0,05	5,01	7,18	7,32	0,28	6,51	0,75	53,70	29,97
3	ALTO	2	0-5	4,40	80,41	0,75	12,10	5,70	2,70	0,30	13,37	0,42	39,90	35,13
5	ALTO	3	0-5	4,30	80,41	0,95	14,44	5,17	2,23	0,33	13,86	0,35	46,20	15,50
1	ALTO	1	5-10	4,00	67,01	2,35	17,23	1,90	2,60	0,18	1,25	0,21	24,60	21,70
3	ALTO	2	5-10	3,90	69,69	3,95	22,45	1,07	1,58	0,20	2,23	0,11	41,10	45,47
5	ALTO	3	5-10	4,00	76,39	3,50	20,55	1,28	1,74	0,18	2,23	0,13	39,00	15,50
1	ALTO	1	10-20	3,90	53,61	2,90	20,55	1,10	1,30	0,13	0,93	0,11	30,90	16,53
3	ALTO	2	10-20	3,90	62,99	4,15	24,52	0,80	1,40	0,15	1,25	0,09	29,70	42,37
5	ALTO	3	10-20	3,90	60,31	3,85	17,23	0,70	1,32	0,13	1,57	0,11	31,20	11,37
2	BAIXO	1	0-5	4,30	80,41	0,93	14,44	4,90	2,50	0,35	6,90	0,35	65,40	45,47
4	BAIXO	2	0-5	4,60	76,39	0,35	11,08	5,90	4,37	0,28	6,51	0,49	53,70	29,97
6	BAIXO	3	0-5	4,30	77,73	1,35	14,44	3,85	2,55	0,35	11,96	0,32	46,80	63,03
2	BAIXO	1	5-10	4,00	67,01	2,95	20,55	0,90	1,70	0,20	1,57	0,12	41,70	35,13
4	BAIXO	2	5-10	4,00	69,69	2,95	20,55	1,90	2,00	0,23	2,23	0,17	37,80	55,80
6	BAIXO	3	5-10	3,90	67,01	3,59	20,55	1,14	2,26	0,23	1,90	0,15	59,70	18,60
2	BAIXO	1	10-20	3,90	53,61	3,43	18,82	0,70	1,60	0,13	0,93	0,11	39,90	10,33
4	BAIXO	2	10-20	3,90	67,01	3,45	18,82	1,10	2,20	0,15	1,57	0,15	35,70	45,47
6	BAIXO	3	10-20	3,90	62,99	4,55	20,55	0,60	1,44	0,15	0,93	0,10	37,20	35,13

Fonte: Laboratório de Solos da Unidade do Cefet Pato Branco

ANEXO 5 POPULAÇÃO 4 MATRIZ DE DADOS ORIGINAL

Análise Solo Rotina + NO₃ NH₄
Coleta 15-01-04

Potreiro	MASSA	Bloco	Prof.	pH	MO	Al ³⁺	H+Al	Ca	Mg	K	P	V	NH ₄	NO ₃
1	ALTO	1	0-5	4,90	80,41	0,11	8,50	6,60	5,68	0,28	19,77	0,60	73,80	46,50
3	ALTO	2	0-5	5,80	73,71	0,00	3,84	8,10	8,30	0,20	4,65	0,81	107,40	82,66
5	ALTO	3	0-5	4,90	73,71	0,05	7,79	4,00	3,40	0,28	5,76	0,50	94,80	35,13
1	ALTO	1	5-10	4,10	67,01	2,55	15,77	1,46	2,14	0,15	2,23	0,19	40,20	31,00
3	ALTO	2	5-10	4,00	67,01	3,40	18,82	1,27	1,63	0,18	1,57	0,14	51,60	10,33
5	ALTO	3	5-10	4,10	71,03	2,45	17,23	2,65	0,65	0,25	1,90	0,17	72,60	20,64
1	ALTO	1	10-20	4,00	53,61	2,85	18,82	1,30	1,39	0,18	1,25	0,13	37,20	25,83
3	ALTO	2	10-20	3,90	67,01	4,10	18,82	0,70	1,45	0,28	1,25	0,11	34,20	8,27
5	ALTO	3	10-20	3,90	64,33	3,59	22,45	0,80	0,57	0,13	0,93	0,06	32,70	18,60
2	BAIXO	1	0-5	4,90	77,73	0,05	7,79	6,40	6,30	0,25	4,30	0,62	76,80	33,07
4	BAIXO	2	0-5	4,60	73,71	0,42	10,15	7,19	1,51	0,28	7,29	0,47	53,40	22,73
6	BAIXO	3	0-5	4,20	69,69	1,15	14,44	3,90	2,00	0,38	12,42	0,30	55,80	22,73
2	BAIXO	1	5-10	4,00	67,01	3,75	20,55	1,07	2,07	0,15	1,25	0,14	45,60	20,67
4	BAIXO	2	5-10	4,00	58,97	3,35	18,82	1,24	1,23	0,15	0,93	0,12	38,40	12,40
6	BAIXO	3	5-10	4,00	57,63	3,35	17,23	1,50	1,00	0,20	2,56	0,14	52,80	18,60
2	BAIXO	1	10-20	3,90	57,63	3,95	20,55	0,90	1,70	0,15	1,25	0,12	38,10	10,33
4	BAIXO	2	10-20	3,90	58,97	3,75	18,82	0,90	1,64	0,15	0,61	0,13	31,80	10,33
6	BAIXO	3	10-20	3,90	64,33	4,23	20,55	0,80	0,50	0,18	1,25	0,07	14,40	8,27

Fonte: Laboratório de Solos da Unidade do Cefet Pato Branco

ANEXO 6 POPULAÇÃO 5 - MATRIZ DE DADOS ORIGINAL

Análise Solo Rotina
Coleta 19-03-04

Potreiro	MASSA	Bloco	Prof.	pH	MO	Al ³⁺	H+Al	Ca	Mg	K	P	V
1	ALTO	1	0-5	4,80	71,03	0,12	7,79	6,50	3,64	0,28	5,09	0,57
3	ALTO	2	0-5	4,30	77,73	0,94	13,22	5,90	0,90	0,25	4,00	0,35
5	ALTO	3	0-5	4,60	69,69	0,42	11,08	7,40	2,40	0,83	15,06	0,49
1	ALTO	1	5-10	4,00	57,63	3,24	17,23	1,10	1,39	0,20	8,20	0,14
3	ALTO	2	5-10	3,90	64,33	3,30	18,82	0,90	1,56	0,18	1,59	0,12
5	ALTO	3	5-10	4,00	60,31	3,37	18,82	1,08	1,35	0,30	1,93	0,13
1	ALTO	1	10-20	3,90	52,27	3,21	17,23	0,60	0,80	0,13	1,26	0,08
3	ALTO	2	10-20	3,90	45,57	3,95	16,33	0,30	0,54	0,13	0,94	0,06
5	ALTO	3	10-20	3,90	58,97	3,95	20,55	0,70	1,00	0,18	1,26	0,08
2	BAIXO	1	0-5	4,30	80,41	1,09	13,22	4,30	2,13	0,30	6,61	0,34
4	BAIXO	2	0-5	4,30	69,69	0,95	13,22	4,10	2,55	0,30	5,84	0,34
6	BAIXO	3	0-5	4,20	73,71	1,64	15,77	3,04	1,86	0,23	4,72	0,25
2	BAIXO	1	5-10	3,90	64,33	3,21	18,82	1,14	1,20	0,18	1,93	0,12
4	BAIXO	2	5-10	3,90	64,33	3,70	20,55	0,90	1,52	0,15	1,26	0,11
6	BAIXO	3	5-10	4,00	67,01	2,94	18,82	1,24	0,76	0,18	1,26	0,10
2	BAIXO	1	10-20	3,90	64,33	3,72	18,82	0,90	0,98	0,13	0,94	0,10
4	BAIXO	2	10-20	3,90	64,33	4,45	20,55	0,53	0,92	0,15	0,94	0,07
6	BAIXO	3	10-20	3,90	53,61	3,35	20,55	0,49	0,66	0,13	0,94	0,06

Fonte: Laboratório de Solos da Unidade do Cefet Pato Branco

ANEXO 7 POPULAÇÃO 6 – MATRIZ DE DADOS ORIGINAL

GANHO MÉDIO DIÁRIO
kg PV/animal/dia
Potreiros

	massa alta	massa baixa	massa alta	massa baixa	massa alta	massa baixa	
Período	1	2	3	4	5	6	Médias
17/08 a 13/09	0,768	0,421	0,643	0,482	0,125	0,343	0,464
13/09 a 05/10	0,205	-0,195	0,023	-0,193	-0,091	-0,509	-0,127
05/10 a 08/11	0,882	-0,032	0,382	-0,010	0,706	0,212	0,357
08/11 a 13/12	0,721	0,362	0,721	0,300	1,043	-0,014	0,522
13/12 a 16/01	0,757	0,412	0,419	0,324	0,618	0,627	0,526
16/01 a 21/02	0,965	0,801	0,694	0,690	0,509	0,398	0,676
21/02 a 19/03	0,179	0,463	0,565	0,562	0,185	0,639	0,432
Médias	0,640	0,319	0,493	0,308	0,442	0,242	

Fonte: Projeto Guamirim –Cefet Pato Branco

ANEXO 8 POPULAÇÃO 7 – MATRIZ DE DADOS ORIGINAL

PRODUÇÃO ANIMAL

kg PV/há

Potreiros

	massa alta	massa baixa	massa alta	massa baixa	massa alta	massa baixa	
Período	1	2	3	4	5	6	Médias
17/08 a 13/09	27	58	37	54	6	42	37
13/09 a 05/10	6	-21	1	-17	-3	-49	-14
05/10 a 08/11	38	-5	27	-2	40	31	21
08/11 a 13/12	32	37	52	21	60	-2	33
13/12 a 16/01	32	41	29	33	38	56	38
16/01 a 21/02	44	84	51	74	45	38	56
21/02 a 19/03	9	37	31	45	8	60	32
19/03 a 21/04	21	47	11	15	11	12	19
Médias	26	35	30	28	26	24	

Fonte: Projeto Guamirim Cefet Pato Branco

ANEXO 9 POPULAÇÃO 8 –MATRIZ DE DADOS ORIGINAL

MASSA DE FORRAGEM

kg MS /há

Potreiros

	massa alta	massa baixa	massa alta	massa baixa	massa alta	massa baixa	
Período	1	2	3	4	5	6	Médias
17/08 a 13/09	2133	1647	1708	1213	1803	1574	1680
13/09 a 04/10	2241	1656	1757	990	1534	1384	1594
04/10 a 08/11	1847	1438	1781	1012	2040	1470	1598
08/11 a 13/12	1731	1284	1588	733	1883	956	1362
13/12 a 16/01	1895	880	1385	780	1840	897	1279
16/01 a 21/02	1898	733	1446	908	2133	972	1348
21/02 a 19/03	1661	747	1437	845	1905	691	1214
19/03 a 21/04	1573	621	1347	690	1470	406	1018
Médias	1872	1126	1556	896	1826	1044	

Fonte: Projeto Guamirim Cefet Pato Branco

ANEXO 10 POPULAÇÃO 9 –MATRIZ DE DADOS ORIGINAL

TAXA DE ACÚMULO
kg MS/ha/dia

	massa alta	massa baixa	massa alta	massa baixa	massa alta	massa baixa	
Período	1	2	3	4	5	6	Média
17/08 a 13/09	-25	-2	19	6	-21	-11	-6
13/09 a 04/10	-45	-84	-42	25	50	-8	-18
04/10 a 08/11	23	15	-9	-10	6	-38	-2
08/11 a 13/12	47	53	63	17	51	14	41
13/12 a 16/01	-12	2	-5	9	37	18	8
16/01 a 21/02	9	27	5	13	37	16	18
21/02 a 19/03	-37	-9	8	9	-15	2	-7
19/03 a 21/04	23	16	5	22	28	16	18
Médias	-2	2	5	11	22	1	

Fonte: Projeto Guamirim Cefet Pato Branco

ANEXO 11 POPULAÇÃO 10 – MATRIZ DE DADOS ORIGINAL

CARGA ANIMAL

kg PV/ ha

Potreiros

	massa alta	massa baixa	massa alta	massa baixa	massa alta	massa baixa	
Período	1	2	3	4	5	6	Médias
17/08 a 13/09	241	899	412	782	294	795	571
13/09 a 04/10	257	917	431	801	295	791	582
04/10 a 08/11	279	904	445	398	338	782	524
08/11 a 13/12	313	555	485	332	388	375	408
13/12 a 16/01	345	589	526	547	413	553	496
16/01 a 21/02	572	652	566	600	686	600	613
21/02 a 19/03	584	713	607	660	485	928	663
19/03 a 21/04	422	755	629	691	495	964	659
Médias	377	748	513	601	424	723	

Fonte: Projeto Guamirim Cefet Pato Branco

ANEXO 12 POPULAÇÃO 11 – MATRIZ DE DADOS ORIGINAL

TAXA DE LOTAÇÃO							
Animais/ha							
	massa alta	massa baixa	massa alta	massa baixa	massa alta	massa baixa	
Período	1	2	3	4	5	6	Médias
17/08 a 13/09	1,26	4,90	2,16	4,00	1,65	4,39	3,06
13/09 a 04/10	1,26	4,90	2,06	4,00	1,65	4,39	3,04
04/10 a 08/11	1,26	4,90	2,06	1,96	1,65	4,39	2,70
08/11 a 13/12	1,26	2,97	2,06	2,13	1,65	3,51	2,26
13/12 a 16/01	1,26	2,94	2,06	3,00	1,57	2,63	2,24
16/01 a 21/02	1,87	2,84	2,06	2,95	2,42	2,62	2,46
21/02 a 19/03	1,79	2,85	2,06	2,95	1,65	3,80	2,52
19/03 a 21/04	1,25	2,85	2,06	2,93	1,65	3,80	2,42
Médias	1,40	3,64	2,07	2,99	1,74	3,69	

Fonte: Projeto Guamirim Cefet Pato Branco

Livros Grátis

(<http://www.livrosgratis.com.br>)

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)
[Baixar livros de Literatura de Cordel](#)
[Baixar livros de Literatura Infantil](#)
[Baixar livros de Matemática](#)
[Baixar livros de Medicina](#)
[Baixar livros de Medicina Veterinária](#)
[Baixar livros de Meio Ambiente](#)
[Baixar livros de Meteorologia](#)
[Baixar Monografias e TCC](#)
[Baixar livros Multidisciplinar](#)
[Baixar livros de Música](#)
[Baixar livros de Psicologia](#)
[Baixar livros de Química](#)
[Baixar livros de Saúde Coletiva](#)
[Baixar livros de Serviço Social](#)
[Baixar livros de Sociologia](#)
[Baixar livros de Teologia](#)
[Baixar livros de Trabalho](#)
[Baixar livros de Turismo](#)