UNIVERSIDADE ESTADUAL PAULISTA "JÚLIO DE MESQUITA FILHO"
FACULDADE DE CIÊNCIAS AGRONÔMICAS
CÂMPUS DE BOTUCATU

APLICATIVO COMPUTACIONAL PARA UTILIZAÇÃO DE COMPONENTES PRINCIPAIS EM EXPERIMENTAÇÃO AGRONÔMICA

NILZA REGINA DA SILVA

Dissertação apresentada à Faculdade de Ciências Agronômicas da UNESP – Câmpus de Botucatu, para obtenção do título de Mestre em Agronomia (Energia na Agricultura).

BOTUCATU - SP
Setembro - 2005

UNIVERSIDADE ESTADUAL PAULISTA "JÚLIO DE MESQUITA FILHO"
FACULDADE DE CIÊNCIAS AGRONÔMICAS
CÂMPUS DE BOTUCATU

APLICATIVO COMPUTACIONAL PARA UTILIZAÇÃO DE COMPONENTES PRINCIPAIS EM EXPERIMENTAÇÃO AGRONÔMICA

NILZA REGINA DA SILVA

Orientador: Prof. Dr. Carlos Roberto Padovani

Dissertação apresentada à Faculdade de Ciências Agronômicas da UNESP – Câmpus de Botucatu, para obtenção do título de Mestre em Agronomia (Energia na Agricultura)

BOTUCATU - SP
Setembro - 2005

DEDICATÓRIA

A DEUS,
pela minha vida,
capacidade e perseverança.

À minha querida mãe que sempre lutou
para que seus filhos pudessem
realizar seus sonhos.

A meu pai,
meus irmãos e minhas irmãs
por estarem sempre ao meu lado.

A Carlos Alberto Aiello e Nilcéia,
pelo incentivo.

Ao Antonio Carlos,
pela ajuda, presteza e constante
incentivo durante todas as etapas do projeto.

AGRADECIMENTOS

Ao ***Prof. Dr. Carlos Roberto Padovani***, pela orientação e presença constante no desenvolvimento deste trabalho e durante o curso, período em que pude reconhecer sua competência e dedicação profissional, bem como o respeito e o carinho dedicados aos seus orientados.

Ao ***Prof. Dr. Flávio Ferrari Aragon***, Departamento de Bioestatística. IB-UNESP, Botucatu-SP, pelo auxílio nos procedimentos de matemática computacional.

Ao ***Prof. Dr. Guilherme Jordão de Magalhães Rosa***, Departament of Animal Science, Michigan State University, Michigan – USA, pelas sugestões apresentadas e revisão do resumo na língua inglesa.

Aos ***funcionários da Biblioteca e à seção de Pós-Graduação***, sempre prestativos.

No desenvolvimento deste trabalho outras pessoas também foram importantes, dentre elas gostaria de alguns destaques em especial:
Ana Helena Neuber de Oliveira
José Higino Rossingnoli
Laura Aparecida Lopes Lotufo
Lisabete Bueno
Silvia Helena Pereira Padovani

A ***todos*** aqueles que,direta ou indiretamente, colaboraram para a realização deste trabalho.

SUMÁRIO

Página

1. RESUMO 01
2. SUMMARY 03
3. INTRODUÇÃO 05
4. REVISÃO DE LITERATURA 07
5. DESENVOLVIMENTO METODOLÓGICO 20
5.1 Introdução à técnica dos componentes principais 20
5.2 Conceitos matemáticos 21
5.2.1 Traço de uma matriz 21
5.2.2 Posto de uma matriz 21
5.2.3 Autovetor e autovalor 21
5.2.4 Polinômio e equação característica 22
5.2.5 Solução do sistema linear homogêneo 22
5.3 Componentes principais 22
5.3.1 Considerações preliminares 22
5.4 Dedução dos componentes principais 25
5.5 Contribuição dos componentes principais na variação total 27
5.6 Componentes principais obtidos de variáveis padronizadas 29
5.7 Correlação das variáveis com os componentes principais 32
5.8 Critérios de escolha do número de componentes 34
5.8.1 Critério de Kaiser 35
5.8.2 Diagrama de autovalores 37
5.8.3 Fatores interpretáveis 38
5.8.4 Método de simulação de Lèbart 39
5.9 Sistema computacional ACOMP 40
5.9.1 Cadastro básico dos experimentos 41
5.9.2 Digitação dos dados do experimento no sistema ACOMP 42
5.9.3 Importação dos dados da planilha EXCEL para o sistema ACOMP 42
5.9.4 Listagem para conferência 43

5.9.5 Exporta experimento para MAPLE V.3..... 43
5.9.6 Dentro do ambiente MAPLE V.3..... 43
5.9.7 Relatório retorno do MAPLE..... 44
5.9.8 Fluxograma do sistema "ACOMP"..... 45
6. RESULTADOS E DISCUSSÃO..... 46
6.1 Convenções para operacionalização do sistema 46
6.2 Exemplos de aplicação da análise de componentes principais em ciências agronômicas..... 51
6.2.1 Agrupamento de dados – ordenação gráfica 51
6.2.2 Delineamento experimental multivariado..... 61
7. CONCLUSÕES..... 74
8. REFERÊNCIAS BIBLIOGRÁFICAS..... 75
APÊNDICE..... 84

1. RESUMO

Os experimentos agronômicos, em geral, apresentam uma quantidade razoável de variáveis observadas e uma complexa estrutura de variação entre e dentro dessas variáveis. Essa estrutura de variação acarreta uma dificuldade para a utilização dos procedimentos requeridos pelo modelo estatístico, em virtude do difícil acesso a programas computacionais para a análise dos dados multivariados. Uma alternativa para redimensionar a quantidade de variáveis consiste na técnica dos componentes principais, que consegue descrever um conjunto com um número menor de variáveis não correlacionadas entre si, ordenadas de maneira decrescente pelas magnitudes das variâncias, de tal forma que a variância total do conjunto inicial seja preservada. Em síntese, a prática da análise de componentes principais é considerada sob o objetivo da redução do espaço paramétrico. Uma das dificuldades encontrada pelos pesquisadores no uso da técnica dos componentes principais, consiste na determinação do número de componentes que deve ser utilizado na redução do espaço paramétrico. Dentre alguns métodos exploratórios discutidos foram apresentados quatro critérios para a escolha do número de componentes principais os quais retem de forma qualificada, a informação contida nas variáveis originais. Neste sentido, foi proposto no presente estudo, a elaboração de um programa computacional, desenvolvido em linguagem MAPLE V.3 e CLIPPER 5.1, de fácil manuseio e acessível a todos os pesquisadores das áreas agronômicas. Visando a operacionalização do aplicativo e a utilização dos procedimentos de análise multivariada, finalizou-se o estudo apresentando dois exemplos

envolvendo situações observadas na literatura agronômica, onde no primeiro faz-se uma abordagem pela metodologia univariada e pela utilização de componentes principais por processo gráfico, e no segundo, faz-se a discussão da técnica da MANOVA e o uso de componentes principais como gerador de uma variável resposta unidimensional transformada a partir do conjunto de todas as variáveis observadas.

COMPUTATIONAL APPLICATION FOR UTILIZATION OF PRINCIPAL COMPONENTS IN AGRONOMICAL EXPERIMENTATION. Botucatu, 2005. 101 p. Dissertação (Mestrado em Agronomia/Energia na Agricultura) – Faculdade de Ciências Agronômicas, Universidade Estadual Paulista.

Author: NILZA REGINA DA SILVA

Adviser: CARLOS ROBERTO PADOVANI

2. SUMMARY

The agronomical experiments, in general, introduce a reasonable quantity of observed variables and a variation complex structure between and within these variables. This variation structure carries a difficulty for the utilization of the procedures required by the statistical model, in view of the difficult access for computational programs for the analysis of the multivariate data. An option for redimensionate the quantity of variable consists in the technique of the principal components, which manages to describe a set with a smaller number of variable not correlated to each other, ordenate of decreasing way by the magnitudes of the variances, of such a form that the total variance of the initial set be preserved. In synthesis, the practice of the analysis of principal components is considered under the objective of the reduction of the parametric space. One of the difficulties found by the researchers in the use of the technique of the principal components, it consists in the determination of the number of components that should be used in the reduction of the parametric space. Among some argued exploratory methods were introduced four criteria for the choice of the number of principal components the ones retain of form qualified, the information contained in the original variables. In this sense, it was proposed at study present, the elaboration of a computational program, developed in language MAPLE V.3 and CLIPPER 5.1, of easy handling and accessible to all the researchers of the agronomical areas. Aiming at operationalization of the application and the utilization of the multivariate analysis procedures, it was concluded the study introducing two examples involving situations observed in the agronomical literature, where in the first an approach is done by the univariate

methodology and by the utilization of principal components for prosecute graph, and in the second, the discussion of the technique of MANOVA is done and the use of principal components as generator of a variable answer unidimensional transformed from the set of all the observed variables.

Keywords: Principal Components, Kaiser Criteria, Eigenvalues Diagram.

3. INTRODUÇÃO

A Análise de Componentes Principais (ACP) é uma técnica estatística de Análise Multivariada (Análise de Interdependência – estuda as relações de um conjunto de variáveis entre si) que transforma linearmente um grupo de variáveis num conjunto substancialmente menor de variáveis não correlacionadas, responsável pela maior parte da informação do conjunto original. Esta idéia foi desenvolvida por Hotelling (1936), embora Pearson (1901) já a tivesse lançada sob uma forma geométrica.

O objetivo da análise de componentes principais é similar ao da análise de fatores (AF). Ambas as técnicas tentam explicar parte da variação de um conjunto de variáveis a partir de um número menor de variáveis subjacentes. A principal diferença entre as duas técnicas é a de que a ACP parte da ausência de um modelo estatístico subjacente na divisão das variáveis observadas e focaliza a explicação da variação total baseando-se nas propriedades da variância máxima dos componentes principais.

A análise de componentes principais pode ser representada geometricamente sob a forma de diagrama de dispersão de pontos relativos às observações no espaço das variáveis. Os fatores ou eixos principais extraídos de uma análise de componentes principais fornecem imagens aproximadas desse conjunto de pontos e esta se propõe a medir a qualidade dessa aproximação.

Numa análise de componentes principais considerando as variáveis padronizadas, a inércia total da dispersão de pontos é igual ao número de variáveis (p). O

autovalor associado a cada fator (autovetor), representa a inércia da nuvem segundo a direção desse eixo. O primeiro autovalor está sempre compreendido entre um e p; se ele está próximo de um, as variáveis não são correlacionadas duas a duas e a ACP não permite a redução da dimensão de dados. Quanto maior for o autovalor, maior será sua capacidade de resumir as variáveis sem perder a informação destes dados e, portanto, mais importante torna-se o fator correspondente ao especificado autovalor. Um autovalor inferior a um indica que o eixo sintetiza menos informação do que uma variável isolada. Considerar o número de autovalores quase nulos permitirá calcular a dimensão real dos dados (FLECK; BOURDEL, 1998).

Uma das principais questões que se coloca na ACP é o critério de escolha do número de componentes principais a manter. Existem alguns métodos clássicos para orientar essa escolha. Os mais conhecidos são: Critério de Kaiser, Diagrama de Autovalores, Fatores Interpretáveis e Critério de Simulação de Lèbart.

Nesse sentido, objetiva-se no presente estudo, considerar os critérios de escolha do número de componentes principais, para desenvolver um programa computacional de fácil manuseio e respectivo manual explicativo do usuário, que possibilite aos pesquisadores da área aplicada a utilização desses critérios no cotidiano experimental. Como ilustração, serão utilizados acervo bibliográfico e banco de dados relativos às experimentações desenvolvidas na área de Agronomia.

4. REVISÃO DE LITERATURA

Segundo Kendall (1950), as técnicas de análise multivariada podem ser classificadas em dois grupos, ou seja:

1. Análise de interdependência: estuda as relações de um conjunto de variáveis entre si.

1.1 Análise de Agrupamento.

1.2 Análise de Componentes Principais.

1.3 Análise de Fatores.

2. Análise de dependência: estuda a dependência de uma ou mais variáveis em relação às outras.

2.1 Análise Discriminante.

2.2 Análise de Variância.

2.3 Análise de Medidas Repetidas.

2.4 Análise de Regressão.

2.5 Análise de Correlação Canônica.

No contexto da literatura especializada destaca-se Anderson (1958) que discutiu vários tópicos teóricos de análise multivariada. Atenção especial dada à

distribuição multinormal de probabilidades, cuja fundamentação é a base de todos os testes estatísticos dos diversos métodos multidimensionais. Atualmente, entre os autores mais citados na discussão teórica dos procedimentos multivariados, encontra-se a obra de Johnson e Wichern apresentada, na primeira edição, em 1982, em que possibilita um enfoque envolvendo uma teoria consistente e aprofundada, complementada com exemplos práticos e respectivos indicativos computacionais para suas realizações (PADOVANI, 2004).

Várias situações práticas envolvendo a análise de componentes principais são descritas em assuntos relacionados às áreas de ciências agronômicas e outras afins. Dentre estas aplicações, algumas são resumidamente apresentadas a seguir.

Jolicouer e Mosimann (1960) discutem a aplicação da técnica dos componentes principais no estudo morfométrico de carapaças de tartarugas envolvendo medidas de largura, comprimento e altura. Na interpretação prática do fator resultante tem-se o volume da carapaça expresso no primeiro eixo canônico obtido a partir da matriz de correlação.

Curi e Padovani (1980) apresentam um estudo, no qual fazem uma revisão teórica enfatizando a análise de discriminação ou classificação complementando-a, com quatro exemplos biológicos. No primeiro, observam sob oito características biológicas, cinco espécies de psitaciformes; no segundo, 10 espécies de columbiformes estão envolvidas nesta pesquisa em que 10 características foram estudadas; no terceiro, 12 indicadores morfométricos são avaliados em três espécies de abelhas e, por último, quatro variáveis observáveis são analisadas envolvendo dois grupos experimentais na área de bovinocultura.

Rao et al. (1981) ao examinarem a divergência genética de arroz, utilizando a técnica dos componentes principais, apresentam como resposta os dois eixos canônicos como os responsáveis pela retenção de 87,5% da variação total dos dados envolvendo nove características biológicas.

A aplicação em experimentos biológicos da análise de agrupamento complementada com a ordenação, usando-se os componentes principais, é amplamente discutida por Curi (1982). Este enfatiza que as análises de variância multivariada e a de Fisher dependem, para sua aplicação, de pressuposições muito rígidas, como a normalidade dos dados e a homogeneidade das matrizes de covariâncias. Infelizmente, estas pressuposições nem sempre estão presentes em pesquisas clínicas, uma vez que há heterogeneidade, na

diversidade individual e nos fatores que influenciam as variáveis estudadas. Baseando-se na linha de aplicação destes procedimentos multivariados, apresenta uma ordenação gráfica utilizando para isto eixos discriminantes (discute até os três primeiros componentes principais) e sugere uma alternativa exploratória destes dados observacionais definindo agrupamentos e criando uma alternativa para posterior análise inferencial dos dados.

Mallants e Feyen (1990) utilizam a análise de componentes principais para abordar a composição de regiões homogêneas de precipitação no oeste da Bélgica e norte da França. O fator relativo ao primeiro eixo (componente principal) reteve 75% da variação total dos dados permitindo dividir a área de estudo em quatro regiões homogêneas.

Demey, Adams e Freites (1994) estabelecem o ponto de corte de 80% da variância total explicada a partir da matriz de correlação, no estudo que aborda a análise, interpretação e classificação dos sistemas de produção agrícola, utilizando como caso particular o sistema de produção de arroz Sistema de Regar Rio Guárico (SRRG), Venezuela.

Zambrano, Demey e González (1995) utilizando o critério de Kaiser na escolha dos componentes principais, estudam a dinâmica do desenvolvimento do tecido caloso de seis cultivares de cana-de-açúcar, a fim de estabelecer um protocolo mais eficiente e a sincronização das atividades do laboratório para a produção "in vitro" de calo sã e vigoroso a ser usado em atividades como o estabelecimento de suspensões celulares de densidades similares, regeneração de plantas e programas de melhoramento onde se usam cultivares distintos de cana-de-açúcar.

Vida et al. (1996) buscam a resposta ótima para a fertilização com os nutrientes NPK em diferentes níveis de irrigação, de cultivares de trigo direcionados para o processo de panificação. A técnica dos componentes principais mostrou-se como um interessante suporte da análise de otimização e apresentou como resultado várias composições entre os níveis de NPK e irrigação quanto à qualidade final do produto na transformação em pães.

Silvarolla et al. (1997) na avaliação de progênies derivadas do híbrido de timor com resistência ao agente da ferrugem, utilizam a técnica dos componentes principais. Foram analisadas 57 progênies onde verificou-se uma alta correlação entre todas as variáveis estudadas sendo os dois eixos principais responsáveis por 91% da variação total do experimento, com 82,4% explicado pelo primeiro eixo e 8,6%, pelo segundo eixo.

Boza e Horii (1998) buscam a resposta ótima da operação de destilação da aguardente, a partir de dados previamente estabelecidos pela correlação entre a composição química e a qualidade sensorial da aguardente. A análise de fatores verificou que os dois primeiros eixos, em conjunto, conseguem captar apenas 49,7% da variação da análise sensorial. Neste sentido, foi necessário acrescentar novas características da aguardente, para aumentar a associação entre os dados da composição química e sensorial, viabilizando a agilização no controle e/ou monitoramento das condições de condução da operação de destilação, bem como das demais operações que compõem a produção de aguardente.

O estudo, em diferentes amostras, das propriedades psicométricas da versão em português do Inventário de Depressão de Beck (BDI) por Gorestein e Andrade (1998), fez uso do critério de Kaiser na análise de componentes principais onde se encontrou que mulheres combinam afeto e autodepreciação na mesma dimensão, enquanto homens combinam conjuntamente a dimensão somática e a autodepressão.

A Escala de Depressão do Centro de Estudos Epidemiológicos (CES-D) visa identificar humor depressivo em estudos populacionais. Silveira e Jorge (1998) utilizando-se da análise de componentes principais através do critério de Kaiser, estudaram as propriedades psicométricas da escala de rastreamento populacional para depressão CES-D em populações clínica e não-clínica de adolescentes e adultos jovens.

Morgano, Queiroz e Ferreira (1999) para avaliarem duas metodologias diferentes de preparação de amostra para determinação da concentração de minerais (Ca, P, Na, K, Mg, Zn, Fe, Mn e Cu), por espectrometria de emissão óptica em plasma indutivamente acoplado (ICP-OES), em sucos de frutas e aplicar, a mais adequada, na análise de sete diferentes variedades de suco de frutas (abacaxi, acerola, caju, goiaba, manga, maracujá e uva), fazem uso da técnica dos componentes principais. Os fatores relativos aos eixos canônicos descrevem 69% da variância total dos dados, sendo o primeiro relacionado principalmente aos minerais Ca, Mg, Fe e P, enquanto o segundo aos minerais Mn, Cu, P e K.

Peroni, Martins e Ando (1999) discutem não só a diversidade inter e intra-especifica presente num sistema itinerante de uma propriedade rural de agricultores tradicionais autóctones da região de Cananéia, litoral sul do Estado de São Paulo, como também as etnovariedades de mandioca presentes neste sistema. Usando a técnica dos componentes principais, e utilizando os dados de mandioca cultivadas nas roças, isto é, de

caracteres avaliados *in situ*, sem que haja qualquer tratamento experimental, o estudo mostra que os dois primeiros componentes principais representam 65,69% da variância total, extraídos da matriz de correlação referentes a 21 caracteres morfológicos de 20 amostras de mandioca de uma roça.

Behrens e Silva (2000) utilizam a técnica dos componentes principais para desenvolver terminologia descritiva e perfil sensorial de vinhos brancos nacionais largamente comercializados no mercado brasileiro, pertencentes aos varietais Chardonnay, Riesling e Gewürztraminer. O estudo mostra que os dois primeiros eixos, conjuntamente, explicam 67,2% da variabilidade total observada entre os vinhos. Concluiu-se que, através da análise de componentes principais, há dois segmentos distintos de vinhos brancos varietais no mercado consumidor de vinhos nacionais, um primeiro segmento constituído por amostras de vinhos com maior intensidade de gosto doce (vinhos suaves) e de sabor e aroma frutado e um segundo segmento, em oposição ao primeiro, constituído por amostras de vinhos com maior intensidade de gostos ácido e amargo, adstringência, sabor e aroma fermentado e sabor alcoólico.

Cardello, Silva e Damásio (2000) para traçar o perfil de aspartame, extrato de folhas de estévia e a mistura ciclamato/ sacarina 2:1 e sacarose, em diferentes níveis de doçura, fazem uso da análise de componentes principais. A porcentagem da explicação das variações ocorridas entre as amostras dadas pelo primeiro eixo, foi bastante alta em todos os níveis de doçura. A análise de fatores mostrou que houve mudança no perfil sensorial dos edulcorantes com o aumento da concentração, porém nos dois níveis mais altos, ficaram semelhantes.

Cavalcanti et al. (2000) utilizam a técnica dos componentes principais para estimar o coeficiente de repetibilidade de alguns caracteres de clones de cajueiro-anão precoce; comparar os métodos utilizados quanto à sua eficiência no processo de estimação da repetibilidade desses caracteres e determinar o número de avaliações adequado para um eficiente processo seletivo. Concluiu-se que a metodologia dos componentes principais, utilizando a matriz de variâncias e covariâncias fenotípicas, é mais eficiente para a estimação do coeficiente de repetibilidade dos caracteres avaliados em clones de cajueiro-anão precoce. Resultou também que, na seleção de clones, deve-se realizar três medições do caráter

produção de castanha e duas outras da altura da planta e diâmetro de copa, a partir do segundo ano de idade das plantas.

Mascioli et al. (2000) buscam estimativas de parâmetros genéticos e fenotípicos e estimam componentes principais para essas características de crescimento nos bovinos de corte da raça Canchim. Foi observado que os dois primeiros componentes principais foram responsáveis por mais de 75% da variação total e constatou-se que a variação nos pesos e ganhos de peso, foi bem sumarizada por três componentes principais que envolveram, principalmente, os pesos na desmama aos 24 meses de idade e os ganhos de peso entre essas idades.

Messetti e Padovani (2000) a partir de características agronômicas das fases vegetativa e reprodutiva, investigam a variabilidade genotípica de 12 populações de girassol e apresentam um procedimento multivariado para identificar as similaridades biológicas entre espécies, utilizando a distância generalizada de Mahalanóbis, como indicador de semelhança. Destacam, para a cultura do girassol, dois aspectos conclusivos ao avanço tecnológico no melhoramento genético: o uso da distância generalizada de Mahalanóbis, como coeficiente de dissimilaridade e a robustez gráfica da primeira variável canônica da divergência genotípica.

Strapasson, Vencovsky e Batista (2000) para selecionar os descritores na caracterização de germoplasma *Paspalum sp*, utilizam-se da análise de componentes principais. O estudo envolveu 27 acessos do gênero *Paspalum,* pertencentes ao grupo Plicatula e concluiu-se que não há necessidade de se trabalhar com número excessivo de descritores e também, que os descritores agronômicos por serem mais influenciados por fatores ambientais, tiveram menor participação na caracterização dos acessos, uma vez que, dos oito descritores selecionados, três foram reprodutivos e três vegetativos.

Brune e Melo (2001) para identificar um método rápido de avaliação da resistência ao esverdeamento dos tubérculos de batata, fazem uso da técnica dos componentes principais. Com a técnica utilizada, pôde-se verificar, que as quatro avaliações contribuíram de forma definitiva, para a composição do primeiro componente principal, que reuniu 87% da variabilidade apresentada pelos dados, e que, pela sua natureza, representou basicamente a curva de esverdeamento dos tubérculos ao longo do tempo.

Lopes et al. (2001) utilizam os componentes principais para estimar o coeficiente de repetibilidade de características físicas e químicas de frutos de aceroleira e, na seqüência, determinam o número de medições necessárias para predizer o valor real dos indivíduos com o valor de determinação genotípica desejado. Concluiu-se que, as estimativas dos coeficientes de repetibilidade das características da altura média, diâmetro médio e peso médio de fruto, vitamina C e acidez titulável, demonstram alta regularidade na superioridade dos indivíduos de um ciclo para outro, e que dois ciclos de avaliação são suficientes para predizer o valor real dos indivíduos com nível de certeza de 90%. Também se chegou à conclusão que o aumento no número de medições pouco acrescenta em precisão e que três medições permitem alcançar coeficientes de determinação próximos a 80%.

Silva et al. (2001) para avaliar a forma de ocupação da terra como um dos fatores determinantes da distribuição espacial da raiva bovina em Minas Gerais, utilizam-se da análise de componentes principais. A pesquisa feita em três períodos obteve mais de 70% da inércia total nos três eixos (componentes principais). Concluiu-se que o uso da terra é um fator determinante da distribuição temporal e espacial da raiva bovina em Minas Gerais. Constatou-se maior associação entre a ocorrência da raiva bovina com o efetivo bovino, áreas de lavouras permanentes e temporárias e pastagens naturais e plantadas. As áreas de matas naturais e plantadas, lavouras em descanso e terras produtivas não utilizadas, apresentam menor associação com a ocorrência e distribuição espacial da raiva bovina.

Oliveira e Padovani (2001) consideram o delineamento experimental multivariado planejado em blocos completos com tratamentos casualizados e discutem uma opção para condensar as múltiplas medidas efetuadas sobre uma mesma parcela e, numa única combinação linear dessas variáveis com a mínima perda de informação (primeiro componente principal) e, posteriormente, abordam a comparação dos tratamentos na variável gerada pela função linear.

Ferreira et al. (2002) consideram a análise exploratória dos teores de constituintes inorgânicos em sucos e refrigerantes de uva por meio da técnica dos componentes principais. Os dados resultantes dos sete analitos determinados em triplicatas de 18 amostras, foram descritos por cinco componentes principais, os quais acumulam 98,7% da variabilidade das informações, e evidenciam quais as características comuns e discrepantes

entre as diferentes amostras, que se mostram importantes para o controle de qualidade dos produtos.

Para avaliar as propriedades psicométricas do Inventário da Depressão Infantil (CDI), adaptado para o Brasil, numa amostra de 287 escolares na faixa etária de 7 a 14 anos de Ribeirão Preto, Golfeto et al (2002) utilizaram o critério de Kaiser para a escolha dos autovalores, na aplicação da análise de componentes principais. Com os resultados obtidos na análise estatística, concluiu-se que o CDI tem propriedades psicométricas relevantes na pesquisa da sintomatologia depressiva.

Machado et al. (2002) para avaliar a divergência genética entre cultivares/linhagens de feijão; determinar o grau de concordância das estimativas e discriminar os caracteres mais importantes na avaliação da divergência genética, discutem a aplicação da técnica de componentes principais. A partir dos 12 genótipos de feijoeiros avaliados em quatro épocas por meio de 10 características agromorfológicas, os três primeiros eixos canônicos, permitem concluir que o genótipo ESAL 693 é o genitor mais promissor para a divergência genética.

Martinello et al. (2002) utilizam a análise de componentes principais para estimar a diversidade genética e discriminar os caracteres morfoagronômicos de maior importância na caracterização da variabilidade de acessos do gênero *Abelmoschus*. A técnica reduziu os 13 descritores em três eixos canônicos que explicaram 76,7% da variância total chegando-se à conclusão do modelo que foi capaz de estabelecer padrões e classificações da diversidade genética em grupos. Verificou-se também, que as características morfométricas delineiam estes grupos auxiliando na pesquisa de novas cultivares e busca de espécies silvestres requeridas como base genética para programas de melhoramento da cultura.

Motta et al. (2002) relacionam a ocorrência da macaúba, no estado de Minas Gerais, a atributos climáticos, pedológicos e vegetacionais. Verificou-se que as variáveis de solo empregadas, discriminaram fortemente os solos, com e sem macaúba, já no primeiro eixo de ordenação (componente principal). Constatou-se, também, que o uso de atributos químicos e físicos do solo permitiu separar de forma eficiente os conjuntos de segmentos da paisagem (sítios) com e sem macaúba, dos quais as correlações com os atributos químicos são as mais relevantes.

Rodrigues et al. (2002) para caracterizar a variabilidade genética de parte do germoplasma existente em poder de produtores de feijão do Rio Grande do Sul, e de cultivares produzidas pela pesquisa, e reuni-las em grupos de similaridade genética, utilizam a análise de componentes principais . O estudo possibilitou identificar descritores ineficientes ou redundantes na discussão da variabilidade genética; reunir as cultivares em quatro grupos distintos de similaridade genética e, ainda, na análise da divergência genética, os quatro primeiros componentes principais explicaram 69,88% da variação.

Shimoya et al. (2002) buscam a análise de componentes principais para estimar o coeficiente de repetibilidade de características forrageiras do capim-elefante; o número mínimo de medições necessárias para predizer o valor real dos genótipos com base em cinco coeficientes de determinação pré-estabelecidos, e ainda, calcular o coeficiente de determinação para cada característica com base na média dos quatro cortes e na estimação dos coeficientes de repetibilidade. A técnica mostrou que os quatro cortes foram suficientes para obter um coeficiente de determinação de 80% e também que, com a adição de três cortes, perfazendo um total de sete, podem-se obter predições com confiabilidade em torno de 90% para as características avaliadas.

Silva et al. (2002) utilizam a técnica dos componentes principais na discriminação geográfica de águas minerais do estado de São Paulo e verificam que os dois eixos canônicos (componentes principais) representam 78,6% da variância total dos dados fornecendo informações discriminatórias claras das amostras.

Alves et al. (2003) selecionam descritores botânico-agronômicos quantitativos para caracterização de germoplasma de cupuaçuzeiro. A técnica dos componentes principais foi empregada em duas etapas para descartar 64% dos 53 descritores associados a quatro características. Concluiu-se que os primeiros eixos canônicos explicavam grande parte da variação existente no banco de germoplasma e que o método de descarte deveria ser suficiente.

Andrade et al. (2003) comparam variedades de cana-de-açúcar para uso na alimentação de ruminantes, utilizando a técnica dos componentes principais. Foram avaliadas 39 variedades de cana-de-açúcar sendo que os três primeiros eixos explicaram 79% da variância total disponível entre os caracteres estudados. Em canas com elevada

porcentagem de fibra insolúvel em detergente neutro e baixa porcentagem de carboidratos totais disponíveis ocorreu maior degradabilidade da fibra.

Azevedo et al. (2003) buscam a análise de componentes principais para avaliar a divergência nutricional de variedades de cana-de-açúcar onde visam identificar três variedades representativas dessa divergência. Os três primeiros componentes principais explicaram 87,8% da variação total e sua utilização foi eficiente em selecionar as variedades com diferentes épocas de colheita.

Farias Neto, Lins e Müller (2003) estimam os coeficientes de repetibilidade para produção de fruto e de albúmem sólido em coqueiro, utilizando a análise de fatores. Pode-se concluir que existe variabilidade significativa entre os híbridos nos caracteres estudados e que a superioridade relativa de híbridos de coqueiro com 90% de certeza do valor real dos indivíduos, com relação à produção de frutos e de albúmem fresco, pode ser conseguida com cinco e três avaliações, respectivamente, após o quinto ano de plantio.

Freire Filho et al. (2003) no estudo da adaptabilidade e estabilidade da produtividade de grãos de 16 genótipos de feijão caupi, ao utilizarem a técnica dos componentes principais, apresentam como resposta os dois primeiros eixos canônicos, explicando por 61,25% da soma de quadrados da interação genótipo versus ambiente. A representação gráfica dos efeitos multiplicativos da interação genótipo versus ambiente nos dois primeiros eixos sugere pouca variação entre os genótipos e alta variação entre os ambientes. Ou seja, a interação genótipo versus ambiente é variável dentro e, homogênea, entre os estados do Piauí e Maranhão em termos de qualidade ambiental.

Guarienti et al. (2003) utilizam o método da análise fatorial com o objetivo de verificar quais são as principais variáveis meteorológicas que afetam a qualidade industrial e o rendimento de grãos de trigo, nas condições da região tritícola sul-brasileira. Verificou-se que a maior parte dos períodos estudados, apresentou duas a três variáveis meteorológicas, afetando simultaneamente as características de qualidade e rendimento de grãos de trigo e que os dois eixos (componentes principais) explicaram, juntos, de 37,97% a 47,91% da variabilidade dos dados.

Marques e Coelho (2003) utilizam o “scree-plot” para representação gráfica dos autovalores, quando averiguam se os fatores relacionais de uma amostra de 188 empresas industriais portuguesas são reconhecidos como fatores competitivos.

Martel et al. (2003) consideram a análise de componentes principais no processo de caracterização morfométrica de raças e populações de pupunha ao longo dos rios Amazonas e Solimões. Os dois componentes determinados pelos dois principais eixos canônicos permitiram a representação bidimensional das 16 populações de pupunha das três raças envolvidas na caracterização morfométrica.

Medeiros (2003) examinando as relações existentes entre características organizacionais, dimensões latentes do comprometimento organizacional e desempenho das organizações e aplicando o critério de Kaiser na escolha dos autovalores, observou que do lado teórico, esse estudo verificou explicitamente a hipótese de que o comprometimento organizacional leva as organizações a um melhor desempenho, uma hipótese até então sem comprovação empírica.

Teixeira, Spehar e Souza (2003) caracterizam e avaliam 48 acessos de três espécies de amaranto para cultivo na entressafra no cerrado, utilizando a análise de componentes principais. Verificou-se que os dois primeiros eixos canônicos representaram 92,18% da variação e permitiu concluir que o amaranto se adapta às condições climáticas e edafológicas do Brasil Central e, além de se ter potencial para se tornar uma opção de cultivo na entressafra, apresenta características agronômicas desejáveis.

Fonseca e Fonseca (2004) procuraram encontrar um modelo multivariado que possibilitasse caracterizar as fases do desenvolvimento do mosaico sucessional de um trecho de floresta estacional semidecidual através de variáveis estruturais. Utilizaram procedimentos de análise de componentes principais e análise de agrupamento. Verificou-se que os dois primeiros eixos canônicos representavam 69,62% da variância total e quando combinado com a análise de agrupamento, permitiu identificar as fases do desenvolvimento. Além disso indicam que a melhora da qualidade de discriminação das fases do desenvolvimento do mosaico sucessional necessita de incrementações.

Gomes et al. (2004) utilizam a análise de componentes principais com o intuito de avaliar diferenças de comportamento entre solos das superfícies geomórficas sul-americana e velhas da região dos cerrados, considerando características físicas, químicas e

mineralógicas sempre sob uso atual de cobertura vegetal nativa. Utilizaram apenas os dois primeiros eixos para explicar os dados e também auxiliar o entendimento das diferenças e similaridades dos ambientes pedológicos separados no campo.

Lopes et al. (2004) avaliam o potencial da técnica dos componentes principais no exame dos relacionamentos entre variáveis supostamente envolvidas na regulação do consumo de pastagem de capim-elefante por vacas mestiças Holandês x Zebu em lactação. Resultou-se que o primeiro componente principal (33,7% da inércia dos dados) representou o uso da suplementação volumosa (cana x uréia) da pastagem; o segundo componente principal (15,3% da inércia) foi relacionado ao consumo de nutrientes do concentrado e o terceiro componente principal (8,5% da inércia) representou efeitos do manejo sobre a composição química da pastagem. Concluiu-se que a análise de componentes principais possibilitou uma visão global dos principais relacionamentos entre as variáveis contextualizadas no estudo.

Monteiro e Pinheiro (2004) para determinar a implementação de tecnologias de suprimento de água potável em municípios cearenses afetados pelo alto teor de sal nos mananciais que abastecem as suas populações, indicam como técnica de redução do número de parâmetros da análise dos componentes principais, e como avaliadores das cargas (pesos) de cada uma das variáveis, a análise de fatores. A partir dos indicadores socioeconômicos e ambientais de cada município, foram construídos índices a partir dos quais efetuou-se uma hierarquização dos municípios estudados. Dentre as 20 variáveis socioeconômicas e ambientais referentes a cada um dos 170 municípios avaliados, foram selecionados sete fatores que, em conjunto, explicam 72,22% das variâncias das variáveis selecionadas. Os resultados mostraram haver forte correlação entre os índices socioeconômicos e o teor de sal encontrados nos poços dos municípios. O estudo mostrou também, que a construção de índices dos municípios a partir da análise fatorial, com base em indicadores socioeconômicos da população, representa critério tecnicamente consistente para estabelecer prioridades na instalação de dessalinizadores ou tecnologias outras, que viabilizem o abastecimento de água potável para consumo humano.

Pinto e Ávila-Pires (2004) discutem a aplicação da técnica dos componentes principais para definir como as cinco espécies de Mabuya se diferenciam em termos de proporções corporais; verificam também se as diferenças observadas poderiam estar

associadas aos habitats ocupados pelas espécies e que papel um possível crescimento alométrico desempenha no desenvolvimento da forma adulta de cada espécie.

Soares, Santos e Ferraudo (2004) consideram 18 variáveis morfométricas de espécimes inteiros e partes do corpo de fêmeas de 58 populações de *Rotylenchulus* spp., obtidas de solo de 13 diferentes cultivos sob a técnica de componentes principais. O resultado do procedimento estatístico possibilitou identificar dois grandes grupos de similaridade nas 58 populações e identificar quais as características morfométricas mais discrepantes entre eles.

Souza-Sobrinho et al. (2004) buscam estimativas de repetibilidade para a produtividade de matéria seca de alfafa na região sudeste do Brasil. A técnica dos componentes principais mostrou-se como um interessante suporte entre as diferentes metodologias de estimação aplicadas em todos os ensaios; apresentou como resultado ser suficiente a realização de quatro cortes para a estimativa do real valor genotípico das cultivares de alfafa, com 85% de confiabilidade.

5. DESENVOLVIMENTO METODOLÓGICO

5.1 Introdução à técnica dos componentes principais

A análise dos componentes principais teve sua origem com Karl Pearson (1901), mas foi Hotelling (1933, 1936) quem a usou com o propósito determinado de analisar as estruturas de correlação.

É um método estatístico multivariado que permite transformar um conjunto de variáveis iniciais correlacionadas entre si, num outro conjunto de variáveis não correlacionadas, que resultam em combinações lineares ortogonais do conjunto inicial. Os componentes principais são apresentados por ordem decrescente de importância, isto é, o primeiro explica o máximo possível da variabilidade dos dados originais, já o segundo o máximo possível da variabilidade ainda não explicada após o efeito do primeiro componente, e assim por diante. O último componente será o que menor contribuição dará para a explicação da variabilidade total dos dados originais. O conjunto final das combinações envolvendo os coeficientes normalizados constitui uma solução única, pelo fato das mesmas explicarem toda a variabilidade total além de serem ortogonais a qualquer outra combinação já definida.

Na prática a análise de componentes principais é utilizada com objetivo da redução do espaço paramétrico.

Quanto maior for a retenção da variação total em um número menor de combinações lineares, melhor será a aplicação prática desse procedimento aos dados experimentais.

5.2 Conceitos matemáticos

5.2.1 Traço de uma matriz

Dada uma matriz A =[a_{ij}] de ordem *n x n*, o traço de A é igual a soma dos elementos da diagonal principal de A, isto é:

$$\text{tr(A)} = a_{11} + a_{22} + ... + a_{nn} = \sum_{i=1}^{n} a_{ii}$$

5.2.2 Posto de uma matriz

Seja A uma matriz arbitrária *m* x *n* sobre um corpo *K*. As linhas de A (A_1, A_2, ..., A_m) geram um subespaço de K^n cuja dimensão é o posto linha de A e as colunas (A_1, A_2, ..., A_n) geram um subespaço de K^m denominado posto coluna de A. As dimensões do espaço de A são chamadas, respectivamente, *posto linha* e *posto coluna* de A e são iguais.

Assim, o posto linha de uma matriz dá o número máximo de linhas linearmente independentes e o posto coluna o número máximo de colunas linearmente independentes.

5.2.3 Autovetor e autovalor

Seja T: V → V um operador linear. Um vetor **μ** ∈ V, onde V é um espaço vetorial com dimensão finita, sendo **μ** ≠ **0**, é chamado de um *autovetor* de T se existir um número real λ tal que: T(**μ**) = λ **μ**. O número real λ é denominado *autovalor* de T associado ao vetor **μ**.

5.2.4 Polinômio e equação característica

Seja A uma matriz quadrada de ordem *n* e I a matriz identidade de mesma ordem que A, define-se o polinômio característico de A como:

$P(\lambda)=|A-\lambda I|$, cuja equação característica é representada por $|A-\lambda I| = \mathbf{0}$.

5.2.5 Solução do sistema linear homogêneo

Todo sistema linear homogêneo Ax = 0, cujo determinante associado é não nulo, admite somente a solução trivial, isto é:

A é uma matriz *n* x *n* e $|A| \neq 0 \Leftrightarrow$ solução única (0, 0, 0, ..., 0).

A condição necessária e suficiente para que um sistema linear homogêneo, de *n* equações e *n* incógnitas, admita solução própria (não nula) é que o determinante do sistema seja nulo.

Simbolicamente tem-se:

$|A| = 0 \Leftrightarrow$ Sistema indeterminado $\Leftrightarrow \exists$ solução não nula.

Assim, os autovalores são encontrados resolvendo-se a equação característica $|A-\lambda I| = 0$. Substitui-se os valores de λ na equação $(A-\lambda I)\,\boldsymbol{\mu} = \mathbf{0}$ e resolve-se o sistema linear, achando assim os autovetores.

5.3 Componentes principais

5.3.1 Considerações preliminares

Considere as seguintes observações multidimensionais de *p* variáveis aleatórias quantitativas $X_1, X_2, \ldots, X_p$.

Tabela 1. Valores genéricos correspondentes às respostas multidimensionais da n unidades experimentais.

Parcela (Unidade Experimental)	Variável			
	X_1	X_2	...	X_p
1	x_{11}	x_{12}	...	x_{1p}
2	x_{21}	x_{22}	...	x_{2p}
3	x_{31}	x_{32}	...	x_{3p}
...	...	...	...	...
...	...	...	...	...
N	x_{n1}	x_{n2}	...	x_{np}

onde:
x_{ij} representa a observação da j-ésima variável (X_j) na i-ésima parcela, com

$i = 1, 2, ..., n$ (parcela) e $j = 1, 2, ..., p$ (variável).

Considerando cada linha da Tabela 1 como um vetor de resposta multivariada, de ordem *p,* têm-se as seguintes indicações:

i) Representação matricial dos dados

$X = [\, x_{ij} \,]$, de ordem *n* x *p,* ou similarmente,

$$X = [\boldsymbol{x}_i^{'}]; \text{ com } \quad \boldsymbol{x}_i^{'} = \left[x_{i1} \quad x_{i2} \ . \ . \ . \ x_{ip} \right].$$

ii) Vetor de médias amostrais

$$\overline{\boldsymbol{x}'} = \left[\bar{x}_1 \quad \bar{x}_2 \, . \quad . \quad . \quad \bar{x}_p \right] \quad \text{com} \quad \bar{x}_j = \sum_{i=1}^{n} \frac{x_{ij}}{n} \quad \text{e} \quad j = 1, 2, ..., p.$$

iii) Matriz de variância e covariância amostral

Tomando-se:

$$1_p = \begin{bmatrix} 1 \\ 1 \\ ... \\ 1 \end{bmatrix}_{px1} \quad ; \quad \bar{x} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ ... \\ \bar{x}_p \end{bmatrix}_{px1} \quad ; \quad \overline{X} = 1_p \bar{x}^{'} \quad \text{ou} \quad \overline{X} = \begin{bmatrix} \bar{x}_1 & \bar{x}_2 & ... & \bar{x}_p \\ \bar{x}_1 & \bar{x}_2 & ... & \bar{x}_p \\ ... & ... & ... & ... \\ \bar{x}_1 & \bar{x}_2 & ... & \bar{x}_p \end{bmatrix} \quad \text{e}$$

$$A = (X - \overline{X})'(X - \overline{X}), \text{ ou seja,}$$

$$A = \begin{bmatrix} \sum_{i=1}^{n}(x_{i1}-\bar{x}_1)^2 & \sum_{i=1}^{n}(x_{i1}-\bar{x}_1)(x_{i2}-\bar{x}_2) & \dots & \sum_{i=1}^{n}(x_{i1}-\bar{x}_1)(x_{ip}-\bar{x}_p) \\ \sum_{i=1}^{n}(x_{i2}-\bar{x}_2)(x_{i1}-\bar{x}_1) & \sum_{i=1}^{n}(x_{i2}-\bar{x}_2)^2 & \dots & \sum_{i=1}^{n}(x_{i2}-\bar{x}_2)(x_{ip}-\bar{x}_p) \\ \dots & \dots & \dots & \dots \\ \sum_{i=1}^{n}(x_{ip}-\bar{x}_p)(x_{i1}-\bar{x}_1) & \sum_{i=1}^{n}(x_{ip}-\bar{x}_p)(x_{i2}-\bar{x}_2) & \dots & \sum_{i=1}^{n}(x_{ip}-\bar{x}_p)^2 \end{bmatrix};$$

define-se a matriz S de covariância amostral por:

$$S = \frac{A}{n-1}, \text{ ou, } S = \frac{1}{n-1}(X-\bar{X})'(X-\bar{X}).$$

A representação genérica de S pode ser expressa por:

$S = [s_{jj'}]$ $\quad j, j' = 1,..., p$, onde $\quad s_{jj'} = \dfrac{\sum_{i=1}^{n} x_{ij}x_{ij'} - n\bar{x}_j\,\bar{x}_{j'}}{n-1}$.

$$S = \frac{1}{n-1}\begin{bmatrix} \sum_{i=1}^{n}x_{i1}^2 - n\bar{x}_1^2 & \sum_{i=1}^{n}x_{i1}x_{i2} - n\bar{x}_1\,\bar{x}_2 & \dots & \sum_{i=1}^{n}x_{i1}x_{ip} - n\bar{x}_1\,\bar{x}_p \\ \sum_{i=1}^{n}x_{i1}x_{i2} - n\bar{x}_1\,\bar{x}_2 & \sum_{i=1}^{n}x_{i2}^2 - n\bar{x}_2^2 & \dots & \sum_{i=1}^{n}x_{i2}x_{ip} - n\bar{x}_2\,\bar{x}_p \\ \dots & \dots & \dots & \dots \\ \sum_{i=1}^{n}x_{i1}x_{ip} - n\bar{x}_1\,\bar{x}_p & \sum_{i=1}^{n}x_{i2}x_{ip} - n\bar{x}_2\,\bar{x}_p & \dots & \sum_{i=1}^{n}x_{ip}^2 - n\bar{x}_p^2 \end{bmatrix}$$

iv) Matriz de correlação amostral

A matriz de correlação amostral $R = [r_{jj'}]$ com $j, j' = 1,..., p$ é dada por:

$$r_{jj'} = \frac{\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)(x_{ij'}-\bar{x}_{j'})}{\sqrt{\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)^2\sum_{i=1}^{n}(x_{ij'}-\bar{x}_{j'})^2}}, \text{ ou similarmente, } r_{jj'} = \frac{s_{jj'}}{\sqrt{s_{jj}s_{j'j'}}}.$$

Considerando a transformação das variáveis X_j na variável padronizada Z_j, com os valores obtidos por: $z_{ij} = \dfrac{(x_{ij}-\bar{x}_j)}{s_{jj}}$, para todo j tem-se:

$$(x_{ij} - \bar{x}_j) = s_{jj} z_{ij} \quad \text{e} \quad r_{jj} = \frac{\sum_{i=1}^{n} s_{jj} z_{ij} s_{j'j'} z_{ij'}}{\sqrt{s_{jj}^2 \quad \sum_{i=1}^{n} z_{ij}^2 \quad s_{j'j'}^2 \quad \sum z_{ij'}^2}} \quad = \quad \frac{\sum_{i=1}^{n} z_{ij} z_{ij'}}{\sqrt{\sum_{i=1}^{n} z_{ij}^2 \sum_{i=1}^{n} z_{ij'}^2}}.$$

5.4 Dedução dos componentes principais

A análise de componentes principais consiste em considerar a partir das variáveis originais, combinações lineares não-correlacionadas, obtidas em número menor que o inicial, que consiga reter parte considerável das informações da variabilidade dos dados observados.

Seja X o vetor aleatório referente às observações das p variáveis originais, $X' = [X_1, X_2, \ldots, X_p]$, com vetor de média $\boldsymbol{\mu}$, matriz de covariância $\boldsymbol{\Sigma}$.

Define-se como primeiro componente principal de X a combinação linear Y_1 expressa por

$$Y_1 = a_{11}X_1 + a_{12}X_2 + \ldots + a_{1p}X_p, \text{ ou ainda,}$$

$$Y_1 = \boldsymbol{a}_1'\mathbf{x},$$

cuja variância, $Var\,(Y_1)$, obtida por $Var\,(Y_1) = \boldsymbol{a}_1'\boldsymbol{\Sigma}\,\boldsymbol{a}_1 = \sum_{j=1}^{p}\sum_{j'=1}^{p} a_{1j}a_{1j'}\sigma_{jj'}$, onde $\sigma_{jj'}$ é o elemento genérico da matriz de covariância $\boldsymbol{\Sigma}$. A expressão da variância de Y_1 pode ser representada pela forma quadrática $Var\,(Y_1) = \boldsymbol{a}_1'\,\boldsymbol{\Sigma}\,\boldsymbol{a}_1$, cujo valor é o maior para todos os vetores com coeficientes normalizados, isto é, $\boldsymbol{a}_1'\boldsymbol{a}_1 = \mathbf{1}$.

Para determinar esses coeficientes normalizados ($\boldsymbol{a}_1'\boldsymbol{a}_1=\mathbf{1}$) utiliza-se o multiplicador de Lagrange λ e deriva-se a expressão lagrangeana em relação a $\mathbf{a}_1$:

$$\frac{\partial}{\partial \boldsymbol{a}_1}\left[\, Var(Y_1) + \lambda(1 - \boldsymbol{a}_1'\,\boldsymbol{a}_1)\right] = \frac{\partial}{\partial \boldsymbol{a}_1}\left[\boldsymbol{a}_1'\boldsymbol{\Sigma}\,\boldsymbol{a}_1 + \lambda\,(\mathbf{1} - \boldsymbol{a}_1'\,\boldsymbol{a}_1)\right]$$

$$= 2\,(\boldsymbol{\Sigma} - \lambda\,\mathbf{I})\,\boldsymbol{a}_1$$

Os coeficientes que maximizam a $Var\,(Y_1)$, devem satisfazer as p equações lineares simultâneas $(\boldsymbol{\Sigma} - \lambda\,\mathbf{I})\,\boldsymbol{a}_1 = \mathbf{0}$.

Para a solução não-trivial, o valor de λ deve ser determinado tal que $|\mathbf{\Sigma} - \lambda \mathbf{I}| = 0$, onde λ é o autovalor da matriz de covariância $\mathbf{\Sigma}$, e $\boldsymbol{a}_1$ seu autovetor associado.

Para determinar qual das *p* raízes características deve ser usada como solução das condições para se ter o primeiro componente principal, multiplica-se o sistema de equações $(\mathbf{\Sigma} - \lambda \mathbf{I})\, \boldsymbol{a}_1 = \mathbf{0}$ por $\boldsymbol{a}_1'$, ou seja,

$$\boldsymbol{a}_1' (\mathbf{\Sigma} - \lambda \mathbf{I})\, \boldsymbol{a}_1 = \mathbf{0}$$

$$\boldsymbol{a}_1' \mathbf{\Sigma}\, \boldsymbol{a}_1 - \lambda \boldsymbol{a}_1' \boldsymbol{a}_1 = \mathbf{0}$$

Considerando $\boldsymbol{a}_1' \boldsymbol{a}_1 = 1$ e que a $Var(Y_1) = \boldsymbol{a}_1' \mathbf{\Sigma}\, \boldsymbol{a}_1$ deve atingir o seu valor máximo, o valor de λ precisa corresponder ao maior valor obtido para as raízes da equação polinomial $|\mathbf{\Sigma} - \lambda \mathbf{I}| = \mathbf{0}$.

Fazendo $\lambda_1 = \max_{1 \le l \le k} (\lambda_l)$, tem-se que:

$\lambda_1 = Var(Y_1)$ e $\boldsymbol{a}_1$ a correspondente solução não-trivial do sistema para maximizar $Var(Y_1)$, são os indicados para que se tenha o primeiro componente principal.

Neste sentido, caracteriza-se como Y_1, o primeiro componente principal obtido das variáveis $X_1, X_2, ..., X_p$, a combinação linear

$$Y_1 = a_{11}X_1 + a_{12}X_2 + ... + a_{1p}X_p,$$

cujos coeficientes são os elementos do autovetor associado ao maior autovalor λ_1 da matriz de covariância $\mathbf{\Sigma}$.

Define-se, como segundo componente principal, a combinação linear

$$Y_2 = a_{21}X_1 + a_{22}X_2 + ... + a_{2p}X_p, \text{ com } Var(Y_2) = \boldsymbol{a}_2' \Sigma\, \boldsymbol{a}_2,$$

cujos coeficientes estão sujeitos às restrições:

$\boldsymbol{a}_2' \boldsymbol{a}_2 = 1$,

$\boldsymbol{a}_1' \boldsymbol{a}_2 = 0$ e

que a variância de Y_2 seja máxima. A primeira restrição é, simplesmente, para assegurar a unicidade dos coeficientes, enquanto a segunda restrição requer que $\boldsymbol{a}_1$ e $\boldsymbol{a}_2$ sejam ortogonais. A conseqüência imediata da ortogonalidade é que o segundo componente principal é não-correlacionado com o primeiro e que a soma das variâncias dos *p* componentes sucessivos resultam na variância total das respostas (MORRISON, 2003). Encontram-se os coeficientes

do segundo componente pelos multiplicadores de Lagrange λ e θ, introduzindo as restrições $\boldsymbol{a}_2'\boldsymbol{a}_2 = 1$ e $\boldsymbol{a}_1'\boldsymbol{a}_2 = 0$ e diferenciando em relação a $\boldsymbol{a}_2$.

$$\frac{\partial}{\partial \boldsymbol{a}_2}\left[\boldsymbol{a}_2'\boldsymbol{\Sigma}\,\boldsymbol{a}_2 + \lambda\,(\mathbf{1} - \boldsymbol{a}_2'\boldsymbol{a}_2) + 2\,\theta\,\boldsymbol{a}_1'\boldsymbol{a}_2\right] = 2\,\boldsymbol{a}_2'(\boldsymbol{\Sigma} - \lambda\,\mathbf{I}) + 2\,\theta\,\boldsymbol{a}_1'$$

Fazendo-se $2\,\boldsymbol{a}_2'\,(\boldsymbol{\Sigma} - \lambda\,\mathbf{I}) + 2\,\theta\,\boldsymbol{a}_1'$ igual a $\mathbf{0}$ e pós-multiplicando a equação por $\boldsymbol{a}_2$

$$2\,\boldsymbol{a}_2'\boldsymbol{\Sigma}\,\boldsymbol{a}_2 + \lambda\,(-2\,\boldsymbol{a}_2')\,\boldsymbol{a}_2 + 2\,\theta\,\boldsymbol{a}_1'\,\boldsymbol{a}_2 = \mathbf{0}$$

$$\boldsymbol{a}_2'\boldsymbol{\Sigma}\,\boldsymbol{a}_2 - \lambda\,\boldsymbol{a}_2'\boldsymbol{a}_2 + \theta\,\boldsymbol{a}_1'\,\boldsymbol{a}_2 = \mathbf{0}$$

, com as restrições $\boldsymbol{a}_2'\boldsymbol{a}_2 = 1$ e $\boldsymbol{a}_1'\boldsymbol{a}_2 = 0$, tem-se:

$$\lambda = \boldsymbol{a}_2'\boldsymbol{\Sigma}\,\boldsymbol{a}_2 = Var\,(Y_2)$$

Utilizando-se o mesmo procedimento são determinados os outros componentes. Neste sentido, o *l-ésimo* componente principal, cujos coeficientes são os elementos do *l-ésimo* autovetor da matriz de covariância $\boldsymbol{\Sigma}$ correspondente ao *l-ésimo* maior autovalor λ_l, pode ser expresso pela combinação linear

$$Y_l = a_{l1}X_1 + a_{l2}X_2 + \ldots + a_{lp}X_p$$ com as restrições:

$\boldsymbol{a}_l'\boldsymbol{a}_l = 1$,

$\boldsymbol{a}_k'\boldsymbol{a}_l = 0$ para qualquer $k<l$ e $l = 1, 2, \ldots, p$.

5.5 Contribuição dos componentes principais na variação total

Em relação à matriz $\boldsymbol{\Sigma}$ e seus respectivos autovetores $\boldsymbol{a}_l'$, podem ser considerados:

i) $\boldsymbol{\Sigma}$ a matriz de ordem p x p e sua decomposição espectral dada por $\boldsymbol{\Sigma} = \sum_{i=1}^{p} \lambda_i a_i a_i'$, onde

(λ_i, $\boldsymbol{a}_i$) são os autovalores e autovetores de $\boldsymbol{\Sigma}$ (GRAYBILL, 1969);

ii) definindo-se $\boldsymbol{A} = [\,\boldsymbol{a}_1\ \boldsymbol{a}_2\ \ldots\ \boldsymbol{a}_p]$ como a matriz dos autovetores e $\boldsymbol{\Lambda} = \text{diag}\,(\lambda_1, \lambda_2, \ldots, \lambda_p)$ a matriz diagonal dos autovalores, com as propriedades relativas.

A decomposição espectral de A; o fato da matriz $\boldsymbol{A}$ ser ortogonal e a propriedade cíclica do traço, tem-se:

$\mathbf{\Sigma} = \sum_{i=1}^{p} \lambda_i a_i a_i^{'} = \boldsymbol{A} \Lambda \boldsymbol{A'}$, com $\boldsymbol{AA'} = \boldsymbol{A'A} = \boldsymbol{I}$, e também,

$$\sum_{i=1}^{p} Var[X_i] = \text{tr}\,(\mathbf{\Sigma}) = \text{tr}\,(\mathbf{A}'\Lambda\,\mathbf{A}) = \text{tr}\,(\Lambda\,\mathbf{A}\,\mathbf{A}') = \text{tr}\,(\Lambda) = \lambda_1 + \lambda_2 + \ldots + \lambda_p = \sum_{j=1}^{p} Var[Y_j]$$

Este resultado implica que a soma das variâncias das variáveis originais é igual à soma das variâncias dos componentes principais. Portanto, a contribuição do j-ésimo componente na explicação da variação total pode ser expressa por:

$$\text{Contr}\,(Y_j) = \frac{Var(Y_j)}{\sum_{j=1}^{p} Var(Y_j)} = \frac{\lambda_j}{\sum_{j=1}^{p} \lambda_j}.$$

A contribuição acumulada dos k primeiros componentes principais pode ser representada como a descrição a seguir.

Para o primeiro eixo canônico representado pelo primeiro componente principal, responsável pela variação máxima tem-se:

$$\text{Contr}\,(Y_1) = \frac{\sum_{j=1}^{1} \lambda_j}{\sum_{j=1}^{p} \lambda_j} = \frac{\lambda_1}{\sum_{j=1}^{p} \lambda_j}$$

A entrada do segundo componente principal no conjunto da redução do espaço paramétrico, acumula na informação o valor da $Var(Y_2)$, cujo resultado global dos dois componentes pode ser expresso por:

$$\text{Contr Acum}\,(Y_1, Y_2) = \text{Contr}\,(Y_1) + \text{Contr}\,(Y_2) = \frac{\sum_{j=1}^{2} \lambda_j}{\sum_{j=1}^{p} \lambda_j} = \frac{\lambda_1 + \lambda_2}{\sum_{j=1}^{p} \lambda_j}.$$

Sucessivamente, tem-se para os *k* primeiros componentes a expressão:

$$\text{Contr Acum } (Y_1, \ldots, Y_k) = \text{Contr } (Y_1) + \ldots + \text{Contr } (Y_k) = \frac{\sum_{j=1}^{k} \lambda_j}{\sum_{j=1}^{p} \lambda_j} .$$

Na dedução dos componentes principais, destacam-se propriedades sobre a indicação da confiabilidade numérica dos cálculos e da contribuição máxima do primeiro componente principal na explicação da variabilidade.

Em relação à confiabilidade tem-se:

$$\prod_{j=1}^{p} \text{Var } [Y_j] = \prod_{j=1}^{p} \lambda_j = |\boldsymbol{\Sigma}|,$$

isto é, o produto dos autovalores é igual ao determinante da matriz de covariância dos dados originais.

Quanto a contribuição máxima do primeiro componente principal, a maior raiz característica não pode exceder a maior soma da linha ou da coluna dos valores absolutos dos elementos de $\boldsymbol{\Sigma}$, ou seja:

$$\lambda_1 \leq \max_i \left(\sum_{j=1}^{p} |\sigma_{ij}| \right) \qquad i = 1, \ldots, p$$

Para a obtenção dos componentes principais, torna-se necessário verificar se as variáveis observadas nas respostas dos dados, estão sendo mensuradas na mesma unidade (escala). Caso essa situação não seja figurada, torna-se imprescindível utilizar variáveis padronizadas, cujas leituras dos dados são efetuadas na escala adimensional.

Nesse sentido apresenta-se, na seqüência, o procedimento da transformação para a escala adimensional.

5.6 Componentes principais obtidos de variáveis padronizadas

Quando as variáveis estão medidas em unidades diferentes, deve-se padronizar cada variável X_j , pois as combinações lineares das variáveis originais acarretariam

dificuldades de interpretação dos resultados, os quais com pouco significado biológico. Usa-se como estrutura de variabilidade a matriz de correlação dos dados originais no lugar da matriz de covariância.

Os componentes principais da matriz de correlação não são necessariamente os mesmos que os obtidos da matriz de covariância.

Esses componentes principais podem ser obtidos utilizando as variáveis padronizadas, Z_1, ..., Z_p, expressas por:

$$Z_j = \frac{(X_j - \mu_j)}{\sqrt{\sigma_{jj}}} \quad , \qquad \text{com} \quad j = 1, \ldots, p.$$

Em notação matricial,

$$\boldsymbol{Z} = \quad (\boldsymbol{V}^{1/2})^{-1} \, (\boldsymbol{X} - \boldsymbol{\mu})$$

onde $\boldsymbol{V}^{1/2}$ é a matriz diagonal envolvendo os desvios padrão das variáveis X_1, X_2, ..., X_p , isto é, $\boldsymbol{V}^{1/2} = diag(\sqrt{\sigma_{jj}})$.

Considerando-se a matriz $\mathbf{P} = [\rho_{jj'}]$ a matriz de correlação do vetor aleatório X, expressa em função das dispersões como:

$$\mathbf{P} = \left[\frac{\sigma_{jj'}}{\sqrt{\sigma_{jj}\sigma_{j'j'}}} \right], \text{ ou seja, } \mathbf{P} = (\mathbf{V}^{1/2})^{-1} \, \Sigma \, (\mathbf{V}^{1/2})^{-1},$$

sob a transformação $Z_j = \frac{(X_j - \mu_j)}{\sqrt{\sigma_{jj}}}$, ou $\boldsymbol{Z} = (\mathbf{V}^{1/2})^{-1}(\boldsymbol{X} - \boldsymbol{\mu})$, tem-se:

$$E(\boldsymbol{Z}) = \mathbf{0} \, ,$$

$$\text{Var } [\boldsymbol{Z}] = \mathbf{P}.$$

À semelhança dos componentes principais obtidos a partir da matriz

$\mathbf{\Sigma} = Var(X)$ pode-se, também, obter os componentes principais da matriz **P** de correlação de ***X***. Deve ser destacado que a matriz **P** envolvendo as correlações entre as variáveis originais (X_1, X_2, ..., X_p), é equivalente à matriz de variâncias e covariâncias das variáveis padronizadas. Neste sentido, todos os resultados obtidos anteriormente para a matriz **Σ**, aplicam-se, com algumas simplificações para a matriz **P**, principalmente à variância unitária de cada $\mathbf{Z}_i$ (JOHNSON; WICHERN, 1992).

O *j-ésimo* componente principal das variáveis padronizadas

Z'= [Z_1, Z_2, ...,Z_p], obtidos a partir de *Var* (**Z**) = **P**, pode ser representado por $Y_l^* = b_{l1}Z_1 + ... + b_{lp}Z_p$, ou equivalentemente,

$Y_l^* = \mathbf{b}_l'\mathbf{Z} = \mathbf{b}_l'(\mathbf{V}^{1/2})^{-1}(\mathbf{X} - \boldsymbol{\mu})$, $l = 1, 2, ..., p$, onde $\mathbf{b}_l$ refere-se ao *l-ésimo* autovetor de **P** correspondente ao autovalor θ_l.

Utilizando a correspondência entre θ_l e λ_l, respectivamente autovalores associados às matrizes **P** (matriz de cov Z) e **Σ** (matriz de cov X), tem-se:

$\sigma_{11} + \sigma_{22} + ... + \sigma_{pp} = \lambda_1 + \lambda_2 + ... + \lambda_p$, logo,

$\rho_{11} + \rho_{22} + ... + \rho_{pp} = \theta_1 + \theta_2 + ... + \theta_p$, ou ainda

$\theta_1 + \theta_2 + ... + \theta_p = p.$

Para o primeiro eixo canônico, representado pelo primeiro componente principal, responsável pela variação máxima, tem-se:

$$\text{Contr}(Y_1^*) = \frac{\sum_{j=1}^{1}\theta_j}{p} = \frac{\theta_1}{p}$$

A entrada do segundo componente principal no conjunto da redução do espaço paramétrico, acumula na informação o valor da $Var(Y_2^*)$, cujo resultado global dos dois componentes pode ser expresso por:

$$\text{Contr Acum}\,(Y_1^*, Y_2^*) = \text{Contr}\,(Y_1^*) + \text{Contr}\,(Y_2^*) = \frac{\sum_{j=1}^{2}\theta_j}{p} = \frac{\theta_1 + \theta_2}{p}.$$

Sucessivamente, tem-se para os *k* primeiros componentes a expressão:

$$\text{Contr Acum}\,(Y_1^*, Y_2^*, ...Y_k^*) = \text{Contr}\,(Y_1^*) + ... + \text{Contr}\,(Y_k^*) = \frac{\sum_{j=1}^{k}\theta_j}{p}.$$

5.7 Correlação das variáveis com os componentes principais

Uma importante contribuição prática referente à interpretação dos componentes principais, pode ser feita com base nas correlações entre as variáveis originais e os componentes principais. Essas correlações são medidas das contribuições individuais de cada variável e não consideram a contribuição multivariada das demais variáveis.

Considera-se o sistema linear homogêneo ($\mathbf{\Sigma} - \lambda_j\,\mathbf{I})\,\boldsymbol{a}_j = \mathbf{0}$, em relação ao *j-ésimo* autovalor λ_j e tem-se:

$$\mathbf{\Sigma}\,\boldsymbol{a}_j = \lambda_j\,\boldsymbol{a}_j$$

Além disso:

a) $Var(Y_j) = a_j' \sum a_j = \lambda_j \quad \rightarrow \quad a_j' \sum a_j a_j' = \lambda_j a_j' \Rightarrow a_j' \sum = \lambda_j a_j'$

$$\rightarrow \quad a_j a_j' \sum a_j = \lambda_j a_j \Rightarrow \sum a_j = \lambda_j a_j$$

b) $Var(X_k) = \sigma_{kk}$

c) $Cov(Y_j, X_k) = Cov(a_{j1}X_1 + ... + a_{jp}X_p, X_k) = Cov(a_{j1}X_1, X_k) + ... + Cov(a_{jp}X_p, X_k)$

$Cov\ (Y_j, X_k) = a_{j1}Cov(X_1, X_k) + ... + a_{jp}Cov(X_p, X_k) = a_{j1}\sigma_{1k} + ... + a_{jp}\sigma_{pk}$

$$Cov(Y_j, X_k) = \sum_{l=1}^{p} a_{jl}Cov(X_l, X_k)$$

$$Cov\ (Y_j, X_k) = \sum_{l=1}^{p} a_{jl}\sigma_{lk}$$

$$\sum a_j = \begin{bmatrix} \sigma_{11} & ..... & \sigma_{1p} \\ \sigma_{21} & ..... & \sigma_{2p} \\ ..... & ..... & ..... \\ \sigma_{p1} & ..... & \sigma_{pp} \end{bmatrix} \begin{bmatrix} a_{1j} \\ a_{2j} \\ .... \\ a_{pj} \end{bmatrix} = \begin{bmatrix} \sum_{l=1}^{p} \sigma_{1l}a_{1j} \\ ... \\ ... \\ \sum_{l=1}^{p} \sigma_{pl}a_{pj} \end{bmatrix} \Rightarrow \lambda_j a_{jk} = \sum_{l=1}^{p} a_{lj}\sigma_{lk}$$

$$\lambda_j a_j = \begin{bmatrix} \lambda_j a_{j1} \\ \lambda_j a_{j2} \\ ..... \\ \lambda_j a_{jp} \end{bmatrix}$$

Neste sentido, conclui-se que:

$$Cov(Y_j, X_k) = \sum_{l=1}^{p} a_{lj}\sigma_{lk} = \lambda_j a_{jk}\,.$$

A correlação entre o *j-ésimo* componente principal (Y_j) e a *k-ésima* variável original (X_k) é dada por:

$$Cor(Y_j, X_k) = \frac{Cov(Y_j, X_k)}{\sqrt{Var(Y_j)Var(X_k)}}, \text{ ou, equivalentemente}$$

$$Cor(Y_j, X_k) = \frac{\lambda_j a_{jk}}{\sqrt{\lambda_j \sigma_{kk}}} = \frac{a_{jk}\sqrt{\lambda_j}}{\sqrt{\sigma_{kk}}}.$$

Se os componentes forem extraídos da matriz **P**, a correlação da variável padronizada Z_k com o *j-ésimo* componente será:

$$Cor(Y_j^*, Z_k) = \frac{Cov(Y_j^*, Z_k)}{\sqrt{Var(Y_j^*)Var(Z_k)}}$$

$$Cor(Y_j^*, Z_k) = \frac{\theta_j b_{jk}}{\sqrt{\theta_j \rho_{kk}}} = b_{jk}\sqrt{\theta_j} \qquad j, k = 1, 2, ..., p$$

Neste caso, (θ_1,$\boldsymbol{b}_1$), (θ_2, $\boldsymbol{b}_2$), ..., (θ_p,$\boldsymbol{b}_p$) são os pares de autovalores e autovetores de **P** com $\theta_1 \geq \theta_2 \geq ... \geq \theta_p \geq 0$.

Todo o desenvolvimento teórico efetuado para **Σ**, também se aplica a **P**. Na prática são utilizados as estimativas de **Σ** e **P** , ou seja, **S** e **R**, respectivamente.

5.8 Critérios de escolha do número de componentes

É muito grande a dificuldade agronômica frente a problemas práticos no cotidiano experimental devido a três fatores básicos: número elevado de variáveis, as relações entre essas variáveis e as diferentes unidades que as mesmas se apresentam. Para contornar este problema, faz-se uso de análises, chamadas multivariadas, que levam em consideração essa estrutura biológica dos dados. Dentre os procedimentos multivariados, destacam-se a análise de componentes principais e suas diversas aplicações nas práticas experimentais. Uma das dificuldades encontrada para sua utilização, consiste na falta de critérios objetivos para determinar qual o número de componentes principais que deve ser

utilizado na análise dos dados experimentais. Dentre os critérios indicados na literatura que auxiliam nessa tomada de decisão, destacam-se:

- Critério de Kaiser.
- Diagrama de Autovalores.
- Fatores Interpretáveis.
- Critério de Simulação de Lèbart.

Deve ser também considerado que a escolha do critério a ser utilizado pode levar a resultados diferentes e tem característica subjetiva, ou seja, constitui-se em uma decisão exploratória a critério do pesquisador.

5.8.1 Critério de Kaiser

Segundo Jensen e Mark (1994) as contribuições de Henry Felix Kaiser incluem:

a) O método Varimax, que consiste na rotação ortogonal dos eixos dos componentes principais. Baseia-se na construção de um algoritmo para a transformação analítica de eixos canônicos. Mais tarde, projetou-se um algoritmo de computador para rotação oblíqua, permitindo os fatores correlatos e as análises hierárquicas.

b) Os critérios analíticos para a estimativa da comunalidade, que consiste em um procedimento matemático para determinar o número de fatores a partir da matriz de correlação (regra dos autovalores maiores que um) e um índice do grau para o qual uma matriz de correlação contribui de forma significativa com análise de fatores.

O procedimento para escolha do número de componentes principais, objeto primordial deste trabalho, tem como o mais comumente usado o estabelecido por Kaiser.

Llorà (2004) fez algumas ponderações sobre o método de Kaiser que serão descritos.

Quando se considera a matriz de correlação (**P**), toma-se como representativos os $q \leq p$ primeiros componentes associados aos autovalores superiores a um (1,0). Isto porque, devido à estrutura da matriz **P**, a soma dos autovalores é igual a p que é o número de variáveis observadas. De certo modo, o autovalor indica de quantas variáveis

originais o autovetor resume as informações. Assim, identificam-se os componentes com autovalor superior a um, como aqueles que contém informação acumulada de mais de uma variável original. Trata-se de componente que realmente condensa a informação da variação total dos dados.

O primeiro componente principal, Y_1, descreve um eixo que minimiza a soma das distâncias euclidianas ao quadrado de todos os pontos até aquele eixo. O segundo componente principal, Y_2, está dado pela projeção ortogonal ao primeiro componente, que retém a segunda maior quantidade de variância dos dados originais; o terceiro componente principal, Y_3, é um vetor ortogonal aos dois primeiros e que tem o terceiro posto na hierarquia estabelecida, de acordo com a retenção da variância, e assim, sucessivamente.

Em resumo, tratando-se o critério de Kaiser de um método exploratório para indicar a redução do espaço paramétrico, seu uso fica limitado às condições práticas de estudo e ao bom senso do pesquisador. Para Horn e Engstrom (1979), esse critério possui dois problemas: o primeiro é ser um tanto arbitrário pois, um autovalor 1,01 é retido, ao passo que outro com valor de 0,99, é rejeitado. O método ignora que os autovalores, como qualquer outro parâmetro em Estatística, são medidos com algum grau de erro. O segundo problema é que muitas vezes resulta em um número elevado de componentes principais quando se tem um grande número de variáveis (≥ 50) e em poucos, quando forem consideradas menos que 20 variáveis.

Segundo Artes (1998) este critério é também conhecido como critério da raiz latente, o qual considera que o número de fatores deve ser igual ao número de autovalores maiores ou iguais à média das variâncias das variáveis analisadas. Quando se utiliza a matriz de correlação (variáveis padronizadas) esse critério corresponde à exclusão dos fatores com autovalores inferiores a um.

Para Hair et al. (2005), a técnica mais comumente usada é o critério da raiz latente. O raciocínio para esse critério é que qualquer autovalor individual deve explicar a variância de pelo menos uma variável se o mesmo há de ser mantido para interpretação. Cada variável contribui com um valor um (1,0) do autovalor total. Logo, apenas os autovalores maiores que um são considerados significantes e aqueles menores que um (1,0) são considerados insignificantes e descartados. Dizem também que usar o autovalor para estabelecer um corte é mais confiável quando o número de variáveis está entre 20 e 50. Se o

número de variáveis é menor que 20, há uma tendência para que esse método extraia um número conservador (muito pouco) de autovalores, ao passo que, quando mais de 50 variáveis estão envolvidas, não é raro que muitos autovalores sejam extraídos.

5.8.2 Diagrama de Autovalores

Sua representação é feita de forma gráfica com o objetivo de relacionar a ordem de magnitude do autovalor com seu respectivo valor numérico. Segundo Fleck e Bourdel (1998) a observação deste diagrama permite conservar aqueles autovalores situados acima do ponto de ruptura do declive da curva da função, identificado quando há uma forte diminuição entre dois autovalores sucessivos, levando-se a conservar na resposta os fatores que o precederam.

Artes (1998) utiliza a expressão critério "scree test" e diz que, comumente, a diferença entre os primeiros autovalores é grande e diminui para os últimos. Dá como sugestão fazer o corte quando a variação passa a ser pequena.

Segundo Reis (1997), se alguma das variáveis iniciais for linearmente dependente das outras, alguns dos autovalores serão nulos e a matriz **P** terá uma característica $r < p$. Nesse caso, a variação total poderá ser completamente explicada pelos primeiros r componentes principais. Quando acontece da dependência linear entre algumas variáveis ser aproximada, os autovalores menores são muito próximos a zero e a sua contribuição para explicar a variância será muito pequena. Por isso, retirar da análise componentes correspondentes a estes valores, não implica uma perda significativa de informação. Em se configurando esta situação, reduz-se a dimensão dos dados e tornam-se os resultados mais simples e de interpretação mais clara. Pode-se representar graficamente a porcentagem de variância explicada pelas componentes principais individualmente. Quando há a redução da porcentagem, a curva passa a ser quase paralela ao eixo das abscissas, razão pela qual excluímos esses componentes principais.

Norman e Streiner (1994) referem-se ao diagrama de autovalores como "Scree Test de Cattell" e o definem como um teste estatístico poderoso que conta simplesmente com o globo ocular.

Para Hair et al. (2005), o teste "scree" é determinado fazendo-se o gráfico relacionado ao número de autovalores em sua ordem de extração, e a forma da curva resultante é usada para avaliar o ponto de corte. Começando com o primeiro autovalor, os ângulos de inclinação rapidamente decrescem no início e então lentamente se aproximam de uma reta horizontal. O ponto no qual o gráfico começa a ficar horizontal é considerado indicativo do número máximo de autovalores a serem extraídos.

5.8.3 Componentes significativos para a redução do espaço paramétrico (Fatores Interpretáveis)

Segundo Fleck e Bourdel (1998), é um critério empírico, que considera conservar os componentes que mais claramente podem dar um significado às respostas do problema. Embora subjetivo, seu uso tem sido recomendado por diferentes autores.

Para a utilização do procedimento observa-se a quantidade de variância acumulada e um nível de corte é estabelecido, representando o total da variância contabilizado pelos componentes selecionados.

Em relação ao número de componentes interpretáveis, têm-se diferentes indicativos sobre o número de componentes que devem ser utilizados na redução do espaço paramétrico.

O desenvolvimento do estudo mostrou quanto à revisão bibliográfica, que se destacam para o valor de 80% de retenção da informação, os autores: Sieglinde Brune, Ricardo Lopes, Aldo Shimoya, José Augusto Gomes Azevedo, Pedro Luiz Martins Soares e J. R. Demey e para o valor de 70% Arthur dos Santos Mascioli, José Ailton da Silva, Gilmar Efrem Martinello, Fernando Vitorino Silva e João Batista de Andrade.

Johnson e Wichern (1992) indicam que se a variabilidade total de um conjunto de p variáveis populacionais pode ser atribuída (de 80 a 90%) ao primeiro, ou aos dois primeiros, ou ainda, aos três primeiros componentes principais, então esses componentes poderão substituir as p variáveis originais com pouca perda de informação.

Segundo Artes (1998), determina-se o número de autovalores de modo que o conjunto de fatores explique uma porcentagem pré-definida da variabilidade global e, ele sugere um nível de explicação de pelo menos 70% da variação total dos dados.

Cardoso et al. (2004) sugerem utilizar os componentes que explicam uma proporção pré-fixada, tipicamente cerca de 70% da variação total.

Hair et al. (2005) denominam como critério de percentagem de variância e dizem que é uma abordagem baseada na conquista de um percentual cumulativo especificado da variância total extraída por autovalores sucessivos. O objetivo é garantir significância prática para os autovalores determinados, garantindo que expliquem pelo menos um montante especificado de variância. Salientam também que nenhuma base absoluta foi montada para todas as aplicações, no entanto, em ciências naturais, o procedimento de obtenção de fatores não deveria ser parado até os autovalores extraídos explicarem pelo menos 95% da variância ou até o último autovalor explicar apenas uma pequena parcela (menos que 5%). Em contraste, em ciências sociais, na qual as informações são menos precisas, não é raro considerar uma solução que explique 60% da variância total (e em alguns casos até menos) como satisfatória.

5.8.4 Método de Simulação de Lèbart

Segundo Fleck e Bourdel (1998), o método de simulação de Lèbart visa responder a seguinte questão: "Se conduzirmos uma análise de componentes principais sobre um conjunto de dados que na sua construção não continham nenhuma estrutura subjacente, o que acontecerá?". Para que um componente possa ser levado em consideração é intuitivamente claro, que os componentes extraídos de um conjunto de dados como este representa um limite inferior o qual deve ser ultrapassado.

Cria-se uma matriz de números aleatórios do mesmo tamanho, mas que respeitem a mesma distribuição (desvio-padrão) e a mesma média de cada variável da matriz em estudo. Após, faz-se a análise de componentes principais e obtém-se uma série de autovalores. Repete-se este processo "n" vezes. Para cada classificação na série de autovalores, conserva-se o autovalor máximo observado ao longo das "n" simulações. Esses valores máximos observados representam o limite inferior que deve ser ultrapassado para que um componente possa ser levado em conta. A probabilidade de se observar um autovalor maior que o valor máximo obtido durante as n-1 simulações precedentes é de 1/n. Assim, para

uma matriz de um tamanho determinado, há um risco de 1/n de se observar um autovalor de uma dada matriz que seja superior a este limite de confiança (FLECK; BOURDEL, 1998).

5.9 Sistema Computacional ACOMP

A intensidade e a velocidade do desenvolvimento da área computacional, realizou-se em associação com a introdução de novas técnicas quantitativas de análises de dados, principalmente os métodos gráficos e os envolvendo computação intensiva.

Segundo Bussab e Moretin (2003), para a operacionalização dessas técnicas, foram desenvolvidos pacotes estatísticos para uso no meio acadêmico, sendo estes genéricos ou específicos. Dentre os pacotes genéricos destacam-se Minitab, SPlus, BMDP, SPSS e o SAS, adequados para realizar uma série de procedimentos estatísticos, entre eles as técnicas multivariadas, em particular, os componentes principais. Por outro lado, para o acesso a estes pacotes estatísticos, exige-se do usuário uma razoável experiência computacional, envolvendo a linguagem própria do pacote e conhecimentos específicos de computação e estatística. Ademais, a acessibilidade a esses pacotes, em algumas situações, torna-se inviável pelo alto custo financeiro. Deve-se salientar também a existência de planilhas que possuem opções para certas técnicas estatísticas, como por exemplo EXCEL e LOTUS.

Segundo Hair et al. (2005), programas estatísticos não são mais primeiramente desenvolvidos para sistemas de grande porte e então adaptados para computadores pessoais; em vez disso, eles agora são inicialmente desenvolvidos para o microcomputador.

Neste sentido, a elaboração do sistema computacional ACOMP torna-se um fator interessante e útil aos pesquisadores da área experimental, pois sua construção será compatível para computadores pessoais e a elaboração criará um ambiente interativo com o usuário.

O sistema computacional ACOMP foi considerado a partir de dois programas, um em linguagem MAPLE V.3, segundo Abell e Braselton (1994), para ambiente WINDOWS 98, e outro em linguagem CLIPPER 5.1, segundo Vidal (1989), denominado "ACOMP.EXE" . O programa MAPLE V.3 refere-se à metodologia estatística utilizada na

análise do critério de escolha do número de componentes principais e o "ACOMP.EXE", em CLIPPER 5.1, trata da manutenção dos arquivos e emissão de relatórios.

Os procedimentos operacionais para uso do sistema acompanham os seguintes passos:

1) cadastramento básico do experimento, onde são fornecidos o código/nome do experimento (até sete caracteres alfanuméricos), a quantidade de variáveis observadas, a descrição das variáveis e a ordem de entrada dessas variáveis;
2) introdução dos valores das variáveis do experimento previamente cadastrado, via digitação direta, e também consulta, alteração, inclusão e exclusão de valores;
3) introdução dos valores das variáveis do experimento previamente cadastrado, por meio de importação de planilha EXCEL. Neste caso, há necessidade do cadastramento básico do experimento;
4) geração de arquivos de comunicação entre os ambientes CLIPPER 5.1 e MAPLE V.3;
5) procedimentos de cálculos matemáticos para análise estatística do número de componentes principais;
6) emissão do relatório de retorno do programa MAPLE V.3.

5.9.1 Cadastro Básico dos Experimentos

O primeiro passo consiste no cadastramento básico do(s) experimento(s). O usuário deve informar o nome do arquivo –no campo CÓDIGO- e fornecer também um nome ao experimento. Em seguida, indicar o número total de variáveis do experimento e a descrição destas variáveis, na seqüência em que as mesmas deverão ser solicitadas na rotina de coleta.

Essas informações são, obrigatoriamente, fornecidas antes da importação de dados da planilha EXCEL ou da digitação na opção COLETADOS.

O nome que constará no campo CÓDIGO do experimento, será utilizado para troca de informações entre a Base de Dados e o MAPLE V.3.

5.9.2 Digitação dos Dados do Experimento no Sistema ACOMP

Permite ao usuário digitar, incluir ou alterar os valores observados para cada variável diretamente no sistema ACOMP.

Poderão ter seus dados digitados somente os experimentos que já tenham o Cadastro Básico.

O número de variáveis solicitado para cada experimento será sempre o informado neste Cadastro Básico pelo pesquisador.

5.9.3 Importação dos Dados da Planilha EXCEL para o Sistema ACOMP

Antes da importação dos dados de uma planilha EXCEL para o Sistema ACOMP, é necessário tomar alguns cuidados na digitação dos valores do experimento na Planilha (Formato TXT), observando os seguintes ítens:

i) a planilha deve ter na célula A11(célula pertencente à primeira linha e primeira coluna) a palavra ACOMP. Caso as demais células contidas nessa primeira linha – após a palavra ACOMP- tenham alguma informação digitada, estas serão desprezadas pelo programa;

ii) a partir da segunda linha (linha de dados), na primeira coluna o usuário informará um registro (numérico ou nominal) para designar cada unidade experimental e nas demais colunas desta linha estarão os valores observados em cada uma das variáveis respostas;

iii) caso a planilha tenha mais variáveis do que o informado no Cadastro Básico do Experimento, as variáveis excedentes serão descartadas.

No Menu [Arquivo], opção [Salvar Como], informe no campo código, o mesmo nome da planilha que foi atribuído à pesquisa em seu cadastramento básico.

No campo "*Salvar como Tipo"*, selecione a opção [Texto – separado por tabulação]*.txt.

No campo *"Salvar em"*, selecione a pasta na qual o sistema ACOMP foi instalado –o sistema tem que estar na pasta [C:\ACOMP].

Ao pressionar o botão [Salvar], o Excel informará que o tipo de arquivo selecionado não suporta pastas de trabalho que contenham mais de uma planilha. Pressione o botão [OK] para que seja salva somente a planilha da pesquisa.

Ao fechar a planilha (ou sair do EXCEL), o sistema emitirá uma mensagem informando que o arquivo salvo não contém o formato Microsoft Excel e questiona se o usuário deseja salvar suas alterações. Pressione o botão [Não] para que o arquivo não seja modificado.

A planilha deve, obrigatoriamente, ser salva em formato TXT, com tabulação, sem a formatação padrão da MICROSOFT. Sempre que se exporta uma planilha para o padrão TXT, o EXCEL "questiona" se o usuário deseja que o arquivo seja formatado nos padrões dos utilitários MICROSOFT. Clique em “Não” para essa formatação.

5.9.4 Listagem para Conferência

Esta sub-rotina fornece a listagem dos dados que foram digitados via sistema ACOMP, ou importados via digitação na planilha EXCEL.

5.9.5 Exporta Experimento para MAPLE V.3

Consiste na formatação de um arquivo com os dados do experimento para ser processado no MAPLE V.3.

5.9.6 Dentro do Ambiente MAPLE V.3

Após gerar o arquivo de formatação com o MAPLE V.3 no ambiente CLIPPER, abra o programa MAPLE V.3 e selecione no menu *File* a opção *Open*....

Na caixa de diálogo Pastas, que se abre, selecione a pasta "MAPLE V.3\BIN" – normalmente é a pasta "DEFAULT" nesse procedimento.

Verifique se o campo Listar Arquivos do Tipo está preenchido com <MAPLE Section - *.MS>. Caso não esteja, selecione essa opção.

No título Nome do arquivo digite no campo código, o nome do arquivo referente aos dados do experimento ou selecione-o na barra de rolagem que se apresenta.

Confirme a operação para que o arquivo seja aberto e suas instruções iniciais apareçam na tela.

No menu *Format*, selecione a opção *Execute Worksheet* que fará com que a rotina de Análise de Componentes Principais seja executada, gerando um arquivo de integração com o ambiente CLIPPER, para posterior impressão dos resultados da análise. Também serão gerados os diagramas Autovalores de S e Autovalores de R, apresentados em janelas que se abrirão sobre a janela principal do MAPLE V.3.

Durante a execução do programa os resultados dos cálculos das rotinas vão "rolando" na tela, até que ao final apareça o comando ">Close();" indicando o término do processamento.

Feche o programa MAPLE V.3, selecionando no menu *File* a opção *Exit*.

O sistema emitirá uma mensagem informando que o arquivo foi executado e questionará se as alterações geradas por essa execução devem ser salvas. Informe <Não> para que o programa se feche.

5.9.7 Relatório Retorno do MAPLE

Lista os valores das estatísticas descritivas e os vetores dos coeficientes dos componentes principais segundo os três critérios de escolha definidos. Lembre-se que esta rotina só poderá ser utilizada após a execução do arquivo relativo ao experimento no ambiente MAPLE V.3.

5.9.8 Fluxograma do Sistema "ACOMP"

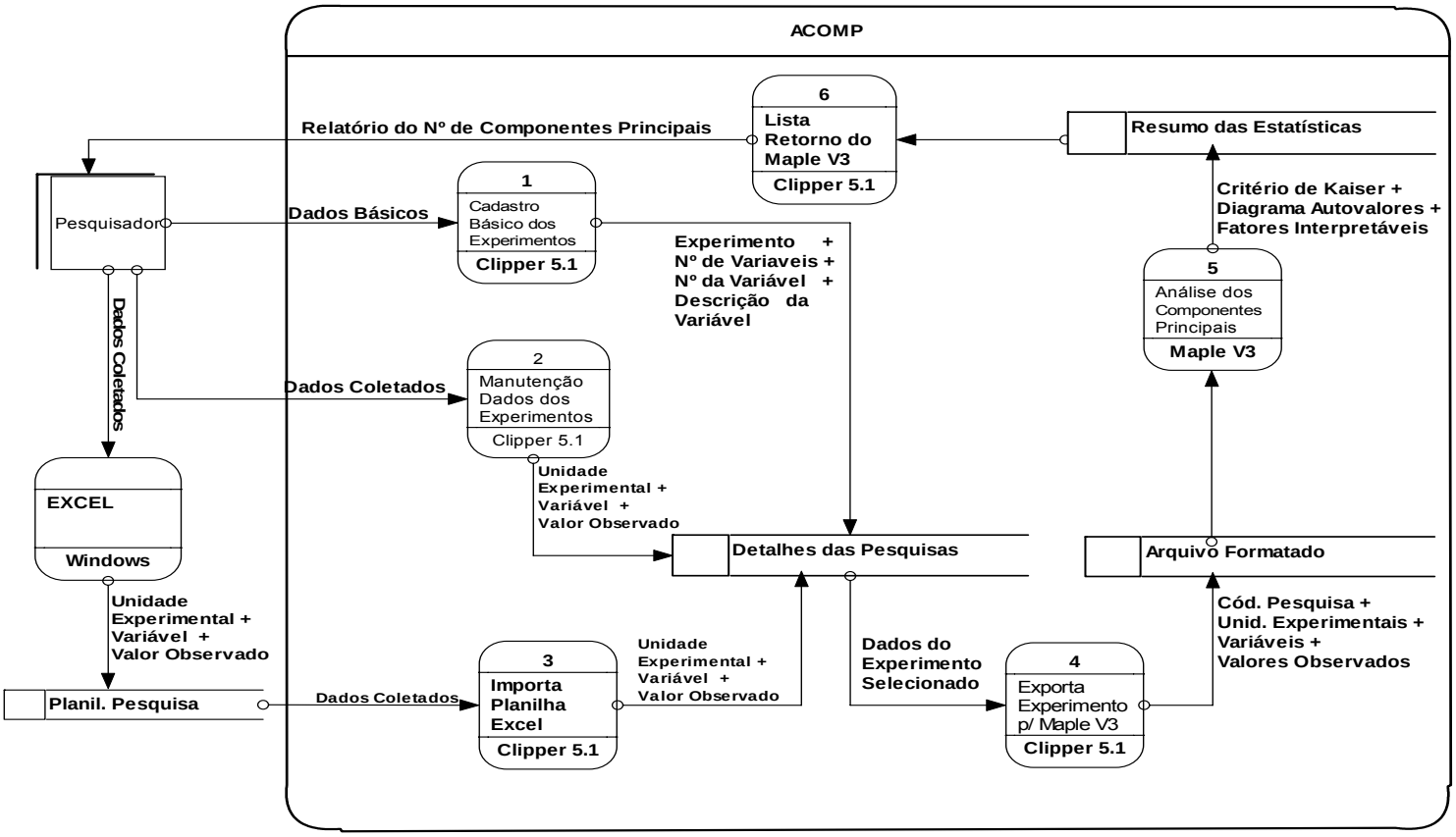


Figura 1. Fluxograma do Sistema Computacional "ACOMP"

O sistema ACOMP foi construído considerando os três primeiros critérios, ou seja, o Critério de Kaiser, o Diagrama de Autovalores e os Fatores Interpretáveis. O Critério de Simulação de Lèbart, por envolver procedimentos referentes a tópicos de matemática (especificamente análise numérica) e, não ser muito divulgado nas áreas aplicadas, desconsiderou-se sua participação no sistema computacional.

6. RESULTADOS E DISCUSSÃO

Para a apresentação e discussão dos resultados, considerou-se em primeiro plano todo o detalhamento da operacionalização do sistema computacional ACOMP e, em seguida dois exemplos de aplicação na área agronômica, sendo um relativo à ordenação gráfica pelos componentes principais e outro, sobre a técnica da ANOVA univariada a partir da redução do espaço paramétrico multidimensional.

6.1 Convenções para Operacionalização do Sistema

Para a operacionalização do sistema computacional ACOMP adotou-se alguns passos para o seu uso em ambiente CLIPPER e MAPLE. Destacam-se a seguir, quatro passos e logo após as convenções utilizadas.

Passo 1

Exige-se, obrigatoriamente, uma informação sobre os dados que constituirão o cadastro básico do(s) experimento(s). Ao acessar o ACOMP, tem-se acesso a uma tela na qual estão dispostos horizontalmente cinco menus (EXPERIMENTOS, MAPLE V.3, CONFIGURA, APOIO e FINALIZA). Com a ajuda das setas do teclado no menu EXPERIMENTOS, acesse a opção DADOS BÁSICOS. Usando-se a tecla Enter, clique-a e será montado em sua tela o CADASTRAMENTO BÁSICO DO EXPERIMENTO. O cursor estará aguardando que seja informado o CÓDIGO (lembre-se que o código é o nome que será

dado ao seu arquivo de dados). Digite um nome neste campo e tecle Enter. Forneça agora um outro nome para o campo TÍTULO e tecle Enter. O programa solicitará, em seguida, o NÚMERO DE VARIÁVEIS. Forneça-o e tecle Enter. De imediato, será solicitado que se faça a DESCRIÇÃO DAS VARIÁVEIS. Assim que realizar a digitação da descrição de cada variável tecle Enter, até que o cursor se desloque para Confirma Inclusão (S/N)?. Para confirmar, digite a letra S. Neste instante o programa abrirá uma nova tela para que seja feito um novo Cadastramento básico do experimento. Caso não tenha mais algum cadastramento a fazer, tecle ESC para retornar ao menu.

Passo 2

Ao retornar ao menu principal, haverá duas opções para que se possa entrar com os dados do experimento. A primeira, usando o próprio programa ACOMP, no menu EXPERIMENTOS, deve-se localizar a opção COLETADOS, tecle Enter e uma tela - Valores Coletados por Indivíduo/Variável- se abrirá. O cursor estará aguardando que se digite uma palavra e se tecle Enter para que a informação seja armazenada na variável código. Em seguida, será solicitado o Número de variáveis e tecle Enter. Após, a unidade experimental bem como sua descrição, serão solicitadas. Forneça as informações e tecle Enter. Como penúltimo passo será solicitado que se indique os valores obtidos para cada indivíduo/variável. Deve-se teclar Enter após cada dado digitado. Fornecidas estas informações, o cursor será deslocado para Confirma Inclusão? (S/N). Tecle S confirmando tudo.

Caso o pesquisador tenha digitado seus dados na opção planilha do EXCEL, esses dados poderão ser usados desde que se observem determinadas diretrizes. Esta planilha deverá, obrigatoriamente, ser salva como tipo **Texto (separado por tabulações *.txt)** em **C:\ACOMP**. Diga <Não> ao ser questionado se deseja que o arquivo seja formatado nos padrões dos utilitários MICROSOFT e, novamente, NÃO, para preservar os recursos, ignorando o recurso do utilitário Microsoft, clique em "Não".

Uma vez que o pesquisador tenha salvado a planilha com os cuidados citados anteriormente, esta poderá ser utilizada em processamento pelo programa ACOMP. Acesse o programa ACOMP e selecione a opção PLANILHA com a tecla Enter –lembre-se que esta opção faz parte do menu EXPERIMENTOS. Uma vez a tecla Enter acionada, abrirá uma outra tela – Importa Dados de Planilha (Formato TXT)- e o cursor estará aguardando que

se forneça o nome do código e tecle Enter. O próprio programa mostrará o número de variáveis. Uma mensagem surgirá e deverá ser digitada a letra S para "Confirma Execução". Tecle Enter e o programa abrirá uma nova tela para Importar Dados. Se não desejar fazer mais importação de dados, tecle ESC para voltar ao menu principal.

No menu principal deve ser localizado o menu MAPLE V.3 e, logo em seguida, a opção EXPORTA deve ser acionada pela tecla Enter. Neste momento abrirá uma tela –Exporta Arquivo para Cálculos no MAPLE V.3- solicitando o nome do código. Forneça-o e tecle Enter. Aparecerá o número de variáveis e o cursor ficará aguardando a letra S na pergunta "Confirma Execução? (S/N)". Após a digitação da letra S uma mensagem será descrita: ARQUIVO GERADO COM SUCESSO. Tecle <Enter>. Se não desejar exportar outro arquivo, tecle ESC para retornar ao menu principal.

Para finalizar a operação localize no menu principal o menu FINALIZA e tecle Enter para sair do ACOMP.

Passo 3

Entre no programa MAPLE V.3 e selecione no menu *File* a opção *Open*. Abrirá uma caixa de diálogo e o nome do arquivo de seus dados deverá ser escolhido (o nome sempre estará aparecendo com a extensão.ms). No canto direito superior desta tela, localize a tecla OK clicando-a. Será montada em seguida, uma tela "nome de seu arquivo".ms.

Acesse o menu FORMAT e clique na opção *Execute Worksheet* . A rotina da Análise de Componentes Principais será executada e, os diagramas dos autovalores das matrizes de covariância e correlação serão construídos com a opção de impressão. Ao final desta execução aparecerá o comando ">Close();" indicando o término do processamento. Feche o programa MAPLE V.3 selecionando no menu *File* a opção *Exit*. Responda <Não> à mensagem (*This worksheet has changed*) do sistema quando este questionar se serão salvas as alterações geradas pela execução.

Passo 4

Acesse novamente o ACOMP, localizando no menu principal o menu MAPLE V.3 e, neste, a opção RELATORIO. Tecle Enter e uma tela –Relatório com Dados Gerados no MAPLE V.3- será visualizada esperando que se forneça o nome do arquivo do

experimento no campo Código. Dada a informação, o programa fornecerá de imediato o número de variáveis e coloca-se à disposição para listar em vídeo ou impressora? Tecle V para vídeo ou I para que seja impresso o relatório.

Convenções adotadas para a operacionalização do sistema computacional ACOMP.

Convenções para uso do Sistema em ambiente CLIPPER:

<ESC> - Retorna ao menu.

[F1] - "Help" do sistema.

[F4] - Tabelas do sistema. Fornece a relação dos elementos cadastrados no arquivo/ tabela associado ao campo.

[F5] - Calculadora. Teclado numérico reduzido.

[↑] - Retorna ao campo anterior da tela (telas de diálogo).
No menu do sistema, permite "navegar" pelas opções do Módulo em que está posicionado.

[↓] - Avança ao campo seguinte da tela (telas de diálogo).
No menu do sistema, permite "navegar" pelas opções do Módulo em que está posicionado.

[←] - Move um caractere à esquerda dentro do campo.
No menu do sistema, permite "navegar" pelos Módulos disponíveis. Permite também 'navegar' pelos registros dos arquivos.

[→] - Move um caractere à direita dentro do campo.
No menu do sistema, permite "navegar" pelos Módulos disponíveis. Permite também 'navegar' pelos registros dos arquivos.

Nas rotinas de manutenção de cadastros e tabelas, sempre que informado o código de um elemento já cadastrado, o sistema apresentará seus atributos e

questionará se o usuário deseja alterar ou excluir o elemento informado, permitindo "navegar" pelo cadastro/ tabela com o uso das teclas [→] e [←].

Sempre que o usuário informar um valor fora dos limites previstos para o campo, o sistema apresentará uma mensagem elucidativa relativa ao mesmo e aguardará que seja pressionada alguma tecla.

As rotinas de emissão dos relatórios possibilitam que os mesmos sejam apresentados em vídeo ou impressora.

O relatório a ser emitido pelo sistema ACOMP, terá as seguintes informações:

1) vetor de médias;
2) matriz de covariância;
3) matriz de correlação;
4) autovalores de S (matriz de covariância) e autovetores de S;
5) autovalores de R (matriz de correlação) e autovetores de R;
6) autovalores e autovetores de S através do Critério de Kaiser;
7) autovalores e autovetores de R através do Critério de Kaiser;
8) fatores interpretáveis (matriz S) - contribuição individual e acumulada;
9) fatores interpretáveis (matriz R) - contribuição individual e acumulada;
10) correlação entre variáveis originais e componentes principais (matriz S);
11) correlação entre variáveis originais e componentes principais (matriz R);
12) diagrama de autovalores da matriz S;
13) diagrama de autovalores da matriz R.

6.2 Exemplos de Aplicação da Análise de Componentes Principais em Ciências Agronômicas

6.2.1 Agrupamento de Dados - Ordenação Gráfica

Uma aplicação muito importante do uso da técnica dos componentes principais consiste na ordenação gráfica de unidades experimentais multidimensionais, onde por meio do gráfico de dispersão dos dados transformados pelos eixos canônicos verificaram-se as similaridades dos vetores de respostas.

Exemplificando esta situação, foram adaptados os dados da tese apresentada por Souza (1985) à Escola Superior de Agricultura "Luiz de Queiroz" (ESALQ), da Universidade de São Paulo – Piracicaba – SP. O experimento desenvolveu-se em condições de casa-de-vegetação e objetivou avaliar o efeito da infestação do solo com isolados de *Pythium* na emergência e ocorrência de podridão em raízes de plântulas de milho. O exemplo a ser tratado na ordenação gráfica, refere-se ao efeito da variabilidade de isolados de patógenos, que teve como objetivo determinar possíveis diferenças na patogenicidade entre os isolados quanto ao desenvolvimento da planta da cultivar milho Doce de Cuba.

As cepas de *Pythium* foram isoladas a partir de raízes de plantas adultas que apresentavam sintomas necróticos associados, e em alguns casos, a destruição do sistema radicular.

Os dez isolados estudados, foram obtidos em lesões de raízes de plantas de milho e classificados, taxonomicamente, como pertencentes ao complexo *Pythium graminicolum – Pythium arrhenomanes* sensu WATANABE, 1974. Seis isolados (1G, 1P, 2P, 3G, 4P e 5G) foram obtidos de plantas de cultivar Piranão VF-1 junto ao campo experimental do Instituto de Genética – ESALQ – USP, Piracicaba – SP. Os outros quatro isolados (A, B, D e F) foram obtidos de plantas da cultivar (HM) oriundas do campo experimental do Departamento de Engenharia – Faculdade de Ciências Agronômicas (FCA) – UNESP, Botucatu-SP.

As variáveis consideradas nesta avaliação dos efeitos de *Pythium* sobre o desenvolvimento das plantas foram: altura da parte aérea (cm); volume das raízes (cm^3);

peso fresco das raízes (g); peso fresco da parte aérea (g); peso seco das raízes (g) e peso seco da parte aérea (g).

O procedimento para mensurar as respostas quantitativas avaliadas nas parcelas experimentais está descrito na seqüência.

Altura da parte aérea(cm)

Tomada por ocasião da coleta das plantas, sendo para tanto, considerada a distância existente do nível do solo à lígula da última folha.

Peso fresco do sistema radicular e da parte aérea(g)

A determinação do peso fresco do sistema radicular e da parte aérea foi feita logo após a chegada do material no laboratório e após o escorrimento do excesso de água das raízes.

Volume das raízes (cm^3)

Na determinação do volume das raízes foi usado um aparelho medidor de volume dotado de um sensor eletrônico de nível (EIRA e SOUZA, 1984). Esta medida foi realizada imediatamente após a avaliação do peso fresco.

Peso seco do sistema radicular e da parte aérea(g)

O peso seco da raiz e da parte aérea foi obtido mediante secagem realizada em estufa, à temperatura de 100ºC ± 2ºC até que atingisse peso constante.

A indicação genérica dos parâmetros de avaliação para o efeito da variabilidade de isolados de *Pythium* na cultivar de milho inoculado e não inoculado está indicada no Quadro 1. Em relação aos vetores médios dos isolados considerados, os valores estão apresentados na Tabela 1.

Quadro 1. Notação das variáveis respostas dos isolados.

Notação	Variável Resposta
X_1	Altura da parte aérea (cm)
X_2	Volume das raízes (cm^3)
X_3	Peso fresco das raízes (g)
X_4	Peso fresco da parte aérea (g)
X_5	Peso seco das raízes (g)
X_6	Peso seco da parte aérea (g)

Tabela 1. Valores observados nos perfis médios das variáveis quantificadas nos isolados de *Pythium* na cultivar de milho inoculado e não inoculado.

	Variáveis					
Isolados	X_1	X_2	X_3	X_4	X_5	X_6
A	11,03(*)	10,96	10,51	11,02	0,83	0,92
B	10,25	10,47	9,65	12,23	0,95	0,98
D	12,23	6,64	8,75	11,45	0,74	1,20
F	14,06	9,97	9,63	13,49	1,12	1,26
1G	11,83	8,36	10,02	12,68	0,98	1,12
1P	12,85	6,68	8,07	13,08	0,71	1,04
2P	12,20	6,40	9,13	10,51	0,92	0,81
3G	12,40	7,34	9,09	11,58	1,06	1,82
4P	7,20	4,71	3,33	4,68	0,92	0,68
5G	11,85	8,47	7,18	8,80	1,06	1,74
Testemunha	16,60	18,82	18,32	20,65	1,84	2,06

(*) representa média de quatro parcelas

Fazendo-se a análise de variância de cada uma das variáveis, isto é, comparando a variação devida aos isolados com a variação devido ao acaso ou resíduo, pode-se verificar se os agrupamentos assemelham-se ao teste de Tukey. As tabelas a seguir mostram

a análise de variância das variáveis: altura da parte aérea; volume das raízes; peso fresco das raízes; peso fresco da parte aérea; peso seco das raízes e peso seco da parte aérea.

Tabela 2. Análise de variância da variável altura da parte aérea.

Causa de Variação	**GL**	**SQ**	**QM**	**F**
Tratamento	10	213,804	21,380	P<0,001
Resíduo	33	120,293	3,645	
Total	43	334,097		

CV(%)=15,84%.

Tabela 3. Análise de variância da variável volume das raízes.

Causa de Variação	**GL**	**SQ**	**QM**	**F**
Tratamento	10	571,740	57,174	P<0,001
Resíduo	33	105,383	3,193	
Total	43	677,123		

CV(%)=19,90%.

Tabela 4. Análise de variância da variável peso fresco das raízes.

Causa de Variação	**GL**	**SQ**	**QM**	**F**
Tratamento	10	501,699	50,170	P<0,001
Resíduo	33	81,643	2,474	
Total	43	583,342		

CV(%)=16,68%.

Tabela 5. Análise de variância da variável peso fresco da parte aérea.

Causa de Variação	**GL**	**SQ**	**QM**	**F**
Tratamento	10	583,607	58,361	P<0,001
Resíduo	33	90,139	2,731	
Total	43	673,746		

CV(%)=13,97%.

Tabela 6. Análise de variância da variável peso seco das raízes.

Causa de Variação	**GL**	**SQ**	**QM**	**F**
Tratamento	10	3,688	0,369	P<0,001
Resíduo	33	1,782	0,054	
Total	43	5,470		

CV(%)=23,01%.

Tabela 7. Análise de variância da variável peso seco da parte aérea.

Causa de Variação	**GL**	**SQ**	**QM**	**F**
Tratamento	10	7,935	0,793	P<0,001
Resíduo	33	0,811	0,025	
Total	43	8,746		

CV(%)=12,65%.

Tabela 8. Resultados das comparações entre isolados por variável estudada (Teste de Tukey)

	Tratamento											
Variável	**A**	**B**	**D**	**F**	**1G**	**1P**	**2P**	**3G**	**4P**	**5G**	**Test**	**DMS (5%)**
$\mathbf{X_1}$	ab	ab	bc	bc	ab	bc	bc	bc	a	ab	c	4,70
$\mathbf{X_2}$	c	bc	abc	bc	abc	abc	ab	abc	a	abc	d	4,40
$\mathbf{X_3}$	bc	bc	bc	bc	bc	bc	bc	bc	a	bc	d	3,87
$\mathbf{X_4}$	bc	bc	bc	c	bc	c	bc	bc	a	b	d	4,07
$\mathbf{X_5}$	a	a	a	a	a	a	a	a	a	a	b	0,57
$\mathbf{X_6}$	Ab	ab	b	b	b	ab	a	c	a	c	c	0,39

Verifica-se na Tabela 8, que a testemunha se diferencia em X_2, X_3, X_4 e X_5 de todos os outros, constituindo-se nesse caso um conglomerado de elemento unitário. A mesma conclusão pode ser feita em relação ao tratamento isolado 4P em relação às variáveis X_3, X_4 .

Para a ordenação gráfica, serão considerados os vetores médios dos isolados como sendo a unidade experimental (unidade taxonômica operacional) representativa do tratamento.

As unidades das variáveis observadas no experimento são mensuradas em escalas diferentes (cm, cm^3 e g). Quando esta situação se apresenta, a melhor opção é padronizar cada variável X_j, pois, esta transformação facilita a interpretação dos resultados frente a adimensionalidade das novas variáveis.

Desta forma a nova estrutura de variação entre as variáveis será estabelecida pelas variações a partir das médias dos dados de cada grupo.

Assim, tem-se como estrutura de variabilidade no lugar da matriz de covariância das unidades taxonômicas, a matriz de correlação das variáveis originais.

Portanto, para os dados em consideração, tem-se:

$$R = \begin{bmatrix} 1,0000000000 & 0,6840035500 & 0,8291301088 & 0,8981615969 & 0,6322817723 & 0,6972603476 \\ 0,6840035500 & 1,0000000000 & 0,9187628816 & 0,8364778652 & 0,8432652320 & 0,5932769701 \\ 0,8291301088 & 0,9187628816 & 1,0000000000 & 0,9456672014 & 0,7638664882 & 0,5922570955 \\ 0,8981615969 & 0,8364778652 & 0,9456672014 & 1,0000000000 & 0,6833803804 & 0,5875466255 \\ 0,6322817723 & 0,8432652320 & 0,7638664882 & 0,6833803804 & 1,0000000000 & 0,7198613838 \\ 0,6972603476 & 0,5932769701 & 0,5922570955 & 0,5875466255 & 0,7198613838 & 1,0000000000 \end{bmatrix}$$

Neste sentido, estão apresentados a seguir os componentes principais e suas respectivas variâncias correspondentes aos autovetores e autovalores da matriz R.

Y_1=0,4074562543 X_1 + 0,4208980264 X_2 + 0,4367114183 X_3 + 0,4281834243 X_4 +

+ 0,3971116963X_5 + 0,3536504556X_6

Var (Y_1) = 4,760494897

Y_2 = - 0,1075965407 X_1 - 0,1527112568 X_2 - 0,3258971250 X_3 - 0,3773872144 X_4 -

- 0,3520540856 X_5 + 0,7697597890 X_6

Var (Y_2) = 0,57526998387

Y_3 = - 0,6020488836 X_1+ 0,4758337114X_2 + 0,08109018407X_3 - 0,2135938617 X_4 +

+ 0,5198003663 X_5 - 0,29787421117 X_6

Var (Y_3) = 0,4654439425

Y_4 = 0,4241223621 X_1 + 0,4121063737 X_2 +0,2195992115 X_3 + 0,02551600713 X_4 -

-0,6488664333 X_5 + 0,4247187670 X_6

Var (Y_4) = 0,1211282203

Y_5 = - 0,5201298130 X_1 - 0,5757688984 X_2 + 0,2783584556 X_3 + 0,5292591847 X_4 +

+ 0,1646496853 X_5 + 0,1150952749 X_6

Var (Y_5) = 0,04507675939

Y_6 = - 0,09759471119 X_1+ 0,2678897477 X_2 - 0,7555073398 X_3 + 0,5897873840 X_4 +

+ 0,005534798317 X_5 + 0,006262897129 X_6

Var (Y_6) = 0,03258633978

A importância de cada eixo canônico na redução do espaço paramétrico é igual ao quociente entre o valor da variância e a soma das variâncias de todos os componentes principais. Então, se uma porcentagem substancial da variabilidade total for explicada pelos k primeiros componentes principais, 80% ou 90%, pode-se usá-los para representar as variáveis originais praticamente sem perda de informação. A Tabela 9 mostra os autovalores e as respectivas contribuições percentuais individual e acumulada.

Tabela 9. Autovalores da matriz de correlação e porcentagem de explicação da variância total.

Componente Principal	**Autovalor**	**Explicação (%)**	**Explicação Acumulada (%)**
1	4,760494897	79,3416	79,3416
2	0,5752698387	9,5878	88,9294
3	0,4654439425	7,7574	96,6868
4	0,1211282203	2,0188	98,7056
5	0,04507675939	0,7513	99,4569
6	0,03258633978	0,5431	100,0000

Conforme observação da regra prática descrita na página 29, no item 5.5, o máximo da soma das linhas (ou colunas) da matriz R resultou em 5,049683773, indicando que a maior raiz característica será inferior a este valor e, portanto, a retenção dada pelo primeiro componente principal não excederá a 84,1614% da variação total. Fato que pode ser comprovado com os valores da Tabela 9.

Os valores das unidades experimentais apresentados na Tabela 1 (pág. 53) sob a transformação nos dois primeiros componentes principais estão apresentados na Tabela 10.

Tabela 10. Valores dos dois primeiros eixos canônicos obtidos a partir da matriz de correlação.

Isolados	Y1	Y2
A	19,07066	-9,44411
B	18,75801	-9,37329
D	17,22012	-8,31836
F	20,79728	-9,90048
1G	18,92939	-9,39315
1P	17,82206	-8,91843
2P	16,80390	-8,28442
3G	18,13450	-8,01351
4P	8,98009	-3,49805
5G	16,33326	-6,51687
Test	32,98682	-16,19013

Para a representação gráfica dos vetores de médias com o objetivo da ordenação gráfica dos isolados, serão consideradas duas situações:

1) ordenação gráfica pelo primeiro componente principal correspondente ao maior valor;

2) ordenação gráfica pelos dois primeiros componentes principais correspondentes aos dois maiores autovalores.

Considerando o primeiro componente principal, responsável por 79,34% da explicação da variação total, têm-se os seguintes agrupamentos de tratamentos:

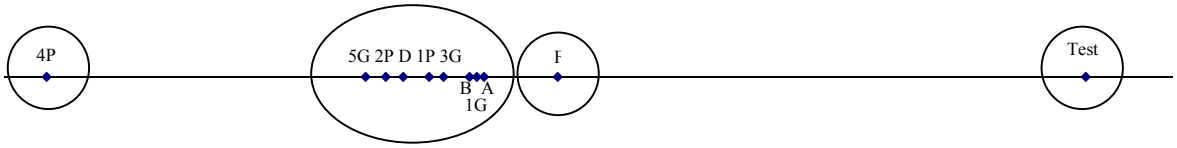


Figura 2. Representação dos valores do primeiro componente principal (Y_1).

A segunda consideração deve-se ao interesse em reter mais de 80% da variação total. Nesta circunstância, torna-se necessário utilizar os dois primeiros componentes principais, os quais explicam 88,93%, cuja representação será realizada por meio de um gráfico bidimensional (Figura 3).

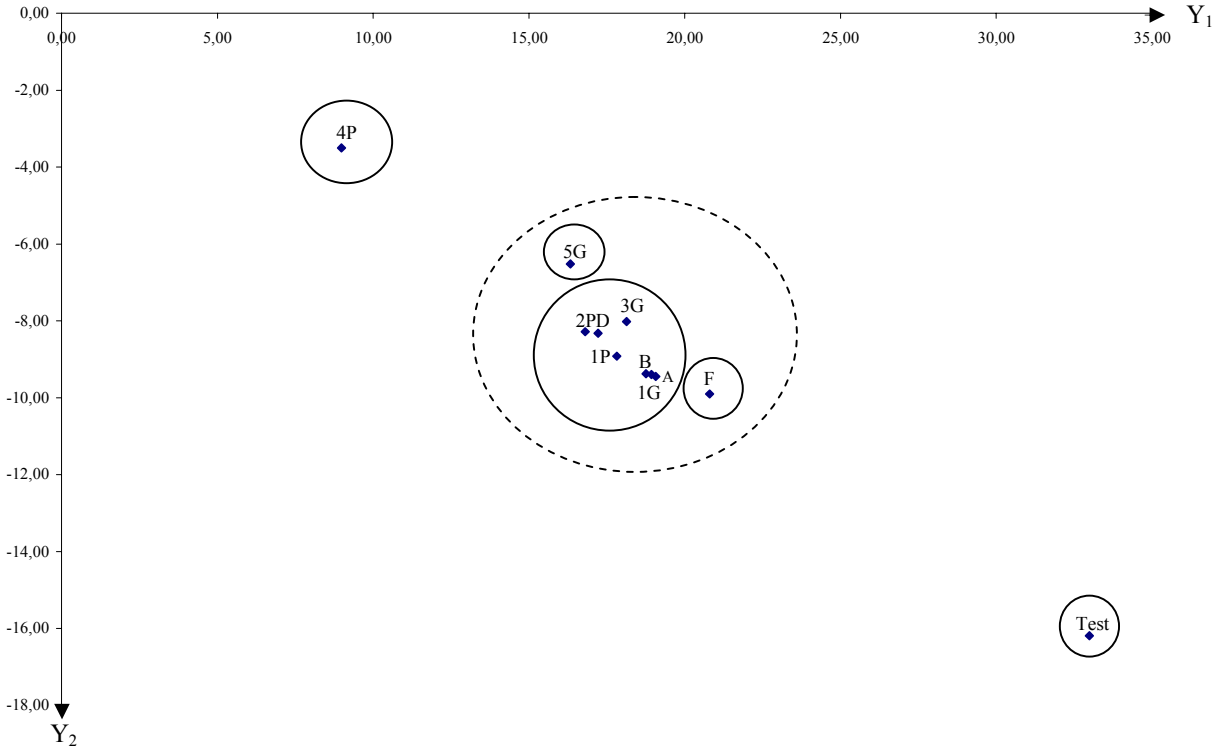


Figura 3. Diagrama de dispersão segundo componentes principais Y_1 e Y_2.

O diagrama de dispersão dos dados transformados pelos eixos canônicos, verificou as similaridades dos vetores de respostas.

O presente estudo considerou a análise conjunta das variáveis e pelo gráfico de dispersão, verificou-se a formação de três conglomerados onde a testemunha e o tratamento isolado 4P se diferenciaram dos demais grupos:

1º grupo - testemunha;
2º grupo - A, B, D, F, 1G, 1P, 2P, 3G e 5G com
- sub grupo 1 (5G)
- sub grupo 2 (A, B, D, 1G, 1P, 2P, 3G)
- sub grupo 3 (F)

3º grupo - 4P.

A representação bidimensional possibilitou com a entrada deste outro componente principal, a distinguir tratamentos não diferenciados pelo conglomerado com mais elementos da classificação anterior. Ou seja, fez-se um refinamento neste conglomerado.

Os resultados encontrados mostraram a existência de uma certa analogia entre a técnica da ordenação gráfica com a técnica univariada comum utilizada no delineamento. Isto ocorreu tanto nos conglomerados formados por tratamentos únicos (testemunha, 4P), como no formado com mais isolados.

Dois aspectos porém devem ser destacados quanto ao uso do procedimento gráfico, quais são:

1) algumas diferenças não reveladas para as análises univariadas, expressaram-se na ordenação gráfica;

2) as evidências das diferenças acentuam-se com a entrada do segundo eixo canônico.

Em síntese, embora sendo a análise gráfica somente de indicação exploratória, esta se mostrou mais informativa para a discussão dos resultados nas comparações realizadas envolvendo simultaneamente todas as características agronômicas do problema.

6.2.2 Delineamento Experimental Multivariado

Os dados multivariados foram adaptados conforme Stuker (1986) e os dados originais, referem-se a produção, em kg por parcela, de um experimento com a cultura de batatinha (*Solanum tuberosum L.*), fornecidos pela Empresa Catarinense de Pesquisa

Agropecuária SC – EMPASC – Itajaí – SC do ano de 1981, implantado segundo o delineamento em blocos completos.

Os tubérculos colhidos foram classificados segundo a norma utilizada pela Empresa Brasileira de Pesquisa Agropecuária – EMBRAPA, nos Ensaios Nacionais de Cultivares de Batata (Graúdo, Médio e Miúdo).

Ou seja,

G: Graúdo, tubérculos com o menor diâmetro superior a 45 mm;

ME: Médio, tubérculos com o menor diâmetro entre 35 mm e 45 mm;

MU: Miúdo, tubérculos com o menor diâmetro inferior a 35 mm.

O experimento em destaque tinha as seguintes características:

Delineamento: Blocos Completos

Número de blocos: 12

Número de cultivares: 5

Número de classes: 3

Unidade experimental: 5,6 m^2 com 20 plantas úteis.

A seguir estão apresentadas as cultivares utilizadas no experimento, bem como as suas origens, turgescência e brotação dos tubérculos, por ocasião do plantio.

Quadro 2. Cultivares utilizadas no experimento, suas origens, turgescência e brotação dos tubérculos, por ocasião do plantio

Cultivares	**Procedência**	**Turgescência**	**Brotação**
1. Anosta	Holanda	Completa	Média
2. Bintje	Holanda	Completa	Média
3. Elvira	Holanda	Completa	Média
4. Gelda	Alemanha	Completa	Boa
5. Recent	Holanda	Boa	Boa

A Tabela 11 a seguir, mostra as produções (kg por parcela) de um ensaio de cultivares de batatinha, no plantio de primavera, na safra 1981/82.

Tabela 11. Produções (kg por parcela) de um ensaio de competição de cultivares de batatinha, no plantio de primavera, na safra 1981/ 82.

Bloco	Anosta			Gelda			Recent			Bintje			Elvira		
	G	ME	MU	G	ME	MU	G	ME	UM	G	ME	MU	G	ME	MU
B_1	14,85	2,33	0,31	13,69	2,08	0,55	17,16	3,27	0,58	12,08	4,35	0,84	15,22	2,93	0,52
B_2	15,06	1,80	0,38	12,13	2,82	0,43	17,34	3,30	0,82	10,54	6,57	0,73	13,47	3,23	0,66
B_3	15,84	2,98	0,42	10,56	2,56	0,54	14,79	2,95	0,43	10,89	4,55	0,83	15,00	2,80	0,62
B_4	14,36	1,56	0,41	13,35	2,46	0,39	14,69	3,19	0,73	10,11	3,96	0,76	15,89	2,64	0,43
B_5	16,47	4,23	0,62	14,51	4,06	0,31	10,26	4,00	1,39	13,28	6,01	0,68	15,15	3,34	0,35
B_6	18,12	2,67	0,52	11,32	4,60	0,38	18,52	4,98	1,52	13,16	8,17	0,59	16,29	2,78	0,22
B_7	12,44	4,26	0,74	13,19	4,02	0,60	16,88	4,71	1,95	14,92	6,03	0,48	15,10	2,40	0,41
B_8	16,86	3,33	0,55	14,64	4,41	0,68	11,75	3,80	0,90	14,18	6,06	0,73	17,82	2,50	0,45
B_9	11,09	4,63	1,00	15,51	3,34	0,64	18,62	2,26	0,38	18,28	4,69	0,23	15,57	2,00	0,52
B_{10}	13,56	5,16	0,65	13,80	3,28	0,57	16,22	2,41	0,34	17,19	4,83	0,14	17,68	1,60	0,34
B_{11}	13,71	3,84	0,78	13,38	3,01	0,48	19,15	1,52	0,53	15,84	3,01	0,16	17,58	1,45	0,19
B_{12}	13,74	2,70	0,76	15,12	2,25	0,53	15,62	2,16	0,40	18,42	3,41	0,28	14,82	1,21	0,61

Têm-se, para o procedimento multivariado as seguintes matrizes de somas de quadrados e produtos:

a) Total

$$T = \begin{bmatrix} 317,3306 & -63,2835 & -10,5800 \\ -63,2835 & 112,3175 & 9,0888 \\ -10,5800 & 9,0888 & 5,6937 \end{bmatrix},$$

b) Tratamentos

$$H = \begin{bmatrix} 55,8414 & -26,7047 & 3,2492 \\ -26,7047 & 48,7032 & 0,2858 \\ 3,2492 & 0,2858 & 1,0647 \end{bmatrix},$$

c) Blocos

$$B = \begin{bmatrix} 47,0832 & -3,9175 & -2,2568 \\ -3,9175 & 29,9145 & 3,3140 \\ -2,2568 & 3,3140 & 0,7225 \end{bmatrix}$$

A matriz E (Resíduo) obtém-se por subtração:

$$E = T - H - B = \begin{bmatrix} 214,4060 & -32,6613 & -11,5724 \\ -32,6613 & 33,6998 & 5,4890 \\ -11,5724 & 5,4890 & 3,9065 \end{bmatrix}$$

Quadro 3. Quadro da análise de variância multivariada (MANOVA) da produção de tubérculos.

Fonte de Variação	G L	Matrizes
Tratamentos	4	$H = \begin{bmatrix} 55,8414 & -26,7047 & 3,2492 \\ -26,7047 & 48,7032 & 0,2858 \\ 3,2492 & 0,2858 & 1,0647 \end{bmatrix}$
Blocos	11	$B = \begin{bmatrix} 47,0832 & -3,9175 & -2,2568 \\ -3,9175 & 29,9145 & 3,3140 \\ -2,2568 & 3,3140 & 0,7225 \end{bmatrix}$
Resíduo	44	$E = \begin{bmatrix} 214,4060 & -32,6613 & -11,5724 \\ -32,6613 & 33,6998 & 5,4890 \\ -11,5724 & 5,4890 & 3,9065 \end{bmatrix}$
Total	59	$T = \begin{bmatrix} 317,3306 & -63,2835 & -10,5800 \\ -63,2835 & 112,3175 & 9,0888 \\ -10,5800 & 9,0888 & 5,6937 \end{bmatrix}$

O teste de hipóteses relativo a procedimento multivariado é realizado, considerando-se a estatística da razão de verossimilhança Λ de Wilks. Por meio de distribuição aproximada, pode-se utilizar a estatística F de Fisher-Snedecor para a decisão da estatística do teste multivariado. Segundo Gomes e Conagin (1987), a transformação estatística F pode ser estabelecida por:

$$\left[(\frac{1}{\Lambda})^{\frac{1}{b}} - 1 \right] \frac{ab - c}{pk_1} \quad \approx \quad F\ (pk_1,\ ab\text{-}c)$$

onde:

p = número de variáveis;

k_1 = número de graus de liberdade de tratamentos;

k_2 = número de graus de liberdade do resíduo.

e ainda,

$a = k_2 - 0{,}5\ (p - k_1 + 1)$;

$$b = \sqrt{\frac{p^2 k_1^2 - 4}{p^2 + k_1^2 - 5}}\ ,\ \text{no caso de}\ \ p^2 + k_1^2 \neq 5\ \ ;$$

$c = 0{,}5\ (pk_1 - 2)$.

Desta forma, têm-se:

$\left| E \right| \ = \ 17235{,}2857$

$\left| H + E \right| = 84170{,}5345$

$$\Lambda \ = \ \frac{|E|}{|H + E|} \ = \ \frac{17235{,}2857}{84170{,}5345} \ = \ 0{,}2048$$

$$\left[(\frac{1}{\Lambda})^{\frac{1}{b}} - 1 \right] \frac{ab - c}{pk_1} = 7{,}62\ (P < 0{,}01)$$

Os limites dos intervalos simultâneos de confiança para contrastes entre os pares de cultivares, no nível de 95% de confiança, serão construídos por:

$$LI = a'(\overline{Y_i} - \overline{Y_{i'}}) - \sqrt{\frac{\chi_\alpha}{1-\chi_\alpha} a' Ea(\frac{2}{b})}$$

$$LS = a'(\overline{Y_i} - \overline{Y_{i'}}) + \sqrt{\frac{\chi_\alpha}{1-\chi_\alpha} a' Ea(\frac{2}{b})}$$

onde,

g é o número de tratamentos, b é o número de blocos e p é o número de classes, com

$s = \min(g-1, p)$, $m = \frac{|g-1-p|-1}{2}$, $n = \frac{g(b-1)-b-p}{2}$,

e χ_α é o percentil de ordem 100 (1-α) % da distribuição da maior raiz característica, com parâmetros *s, m e n* (MORRISON, 2003).

No exemplo, tem-se:

s = 3 , m = 0 e n = 20 e o valor de $\chi_\alpha = 0{,}295$. Os quadros a seguir representam a comparação entre os valores médios das cinco cultivares.

Quadro 4. Comparação dos valores médios das cultivares Anosta e Gelda.

Anosta(A)	**Gelda(G)**	**A-G**	**a'**	**LI**	**LS**
14,68	13,43	1,25	[1 0 0]	-2,62	5,12
3,29	3,24	0,05	[0 1 0]	-1,48	1,58
0,60	0,51	0,09	[0 0 1]	-0,43	0,61

Quadro 5. Comparação dos valores médios das cultivares Anosta e Recent.

Anosta(A)	**Recent(R)**	**A-R**	**a'**	**LI**	**LS**
14,68	15,92	-1,24	[1 0 0]	-5,11	2,63
3,29	3,21	0,08	[0 1 0]	-1,45	1,61
0,60	0,83	-0,23	[0 0 1]	-0,75	0,29

Quadro 6. Comparação dos valores médios das cultivares Anosta e Bintje.

Anosta(A)	Bintje(B)	A-B	a'	LI	LS
14,68	14,07	0,61	[1 0 0]	-3,26	4,48
3,29	5,14	-1,85	[0 1 0]	-3,38	-0,32
0,60	0,54	0,06	[0 0 1]	-0,46	0,58

Quadro 7. Comparação dos valores médios das cultivares Anosta e Elvira.

Anosta(A)	Elvira(E)	A-E	a'	LI	LS
14,68	15,80	-1,12	[1 0 0]	-4,99	2,75
3,29	2,41	0,88	[0 1 0]	-0,65	2,41
0,60	0,44	0,16	[0 0 1]	-0,36	0,68

Quadro 8. Comparação dos valores médios das cultivares Gelda e Recent.

Gelda(G)	Recent(R)	G-R	a'	LI	LS
13,43	15,92	-2,49	[1 0 0]	-6,36	1,38
3,24	3,21	0,03	[0 1 0]	-1,50	1,56
0,51	0,83	-0,32	[0 0 1]	-0,84	0,20

Quadro 9. Comparação dos valores médios das cultivares Gelda e Bintje.

Gelda(G)	Bintje(B)	G-B	a'	LI	LS
13,43	14,07	-0,64	[1 0 0]	-4,51	3,23
3,24	5,14	-1,90	[0 1 0]	-3,43	-0,37
0,51	0,54	-0,03	[0 0 1]	-0,55	0,49

Quadro 10. Comparação dos valores médios das cultivares Gelda e Elvira.

Gelda(G)	Elvira(E)	G-E	a'	LI	LS
13,43	15,80	-2,37	[1 0 0]	-6,24	1,50
3,24	2,41	0,83	[0 1 0]	-0,70	2,36
0,51	0,44	0,07	[0 0 1]	-0,45	0,59

Quadro 11. Comparação dos valores médios das cultivares Recent e Bintje.

Recent(R)	Bintje(B)	R-B	a'	LI	LS
15,92	14,07	1,85	[1 0 0]	-2,02	5,72
3,21	5,14	-1,93	[0 1 0]	-3,46	-0,40
0,83	0,54	0,29	[0 0 1]	-0,23	0,81

Quadro 12. Comparação dos valores médios das cultivares Recent e Elvira.

Recent(R)	Elvira(E)	R-E	a'	LI	LS
15,92	15,80	0,12	[1 0 0]	-3,75	3,99
3,21	2,41	0,80	[0 1 0]	-0,73	2,33
0,83	0,44	0,39	[0 0 1]	-0,13	0,91

Quadro 13. Comparação dos valores médios das cultivares Bintje e Elvira.

Bintje(B)	Elvira(E)	B-E	a'	LI	LS
14,07	15,80	-1,73	[1 0 0]	-5,60	2,14
5,14	2,41	2,73	[0 1 0]	1,20	4,26
0,54	0,44	0,10	[0 0 1]	-0,42	0,62

A seguir apresenta-se a Tabela 12 com os resultados das comparações entre as diversas cultivares e as classes Graúda, Média e Miúda.

Tabela 12. Comparações entre as diversas cultivares segundo as classes Graúda, Média e Miúda.

Comparação	Classes		
	Graúda	**Média**	**Miúda**
Anosta x Gelda	P>0,05	P>0,05	P>0,05
Anosta x Recent	P>0,05	P>0,05	P>0,05
Anosta x Bintje	P>0,05	**P<0,05**	P>0,05
Anosta x Elvira	P>0,05	P>0,05	P>0,05
Gelda x Recent	P>0,05	P>0,05	P>0,05
Gelda x Bintje	P>0,05	**P<0,05**	P>0,05
Gelda x Elvira	P>0,05	P>0,05	P>0,05
Recent x Bintje	P>0,05	**P<0,05**	P>0,05
Recent x Elvira	P>0,05	P>0,05	P>0,05
Bintje x Elvira	P>0,05	**P<0,05**	P>0,05

A Tabela 13 mostra a média e o desvio-padrão das produções de batatinha por cultivares de classificação em Graúda, Média e Miúda.

Tabela 13. Média e desvio-padrão das produções de batatinha segundo cultivares nas diferentes classificações.

Classificação	Cultivares				
	Anosta	**Gelda**	**Recent**	**Bintje**	**Elvira**
Graúda	14,68 ± 1,96 a (*)	13,43 ± 1,49 a	15,92 ± 2,72 a	14,07 ± 2,92 a	15,80 ± 1,33 a
Média	3,29 ± 1,14 a	3,24 ± 0,86 a	3,21 ± 1,04 a	5,14 ± 1,47 b	2,41 ± 0,70 a
Miúda	0,60 ± 0,20 a	0,51 ± 0,11 a	0,83 ± 0,52 a	0,54 ± 0,27 a	0,44 ± 0,15 a

(*) Médias seguidas da mesma letra não diferem estatisticamente entre si.

A seguir será considerada uma resolução alternativa à técnica da análise de variância multivariada (MANOVA) do modelo em blocos completos, que consiste em

utilizar a técnica da análise de variância univariada (ANOVA) nos dados transformados pelo primeiro componente principal.

Para isto, considera-se o primeiro autovalor e seu correspondente autovetor, obtido do sistema linear homogêneo $|E^{-1}H - \lambda I|\ \mathbf{a} = \mathbf{0}$ ou, equivalentemente, $|H - \lambda E|\ \mathbf{a} = \mathbf{0}$.

Tem-se:

$$H = \begin{bmatrix} 55{,}8414 & -26{,}7047 & 3{,}2492 \\ -26{,}7047 & 48{,}7032 & 0{,}2858 \\ 3{,}2492 & 0{,}2858 & 1{,}0647 \end{bmatrix} \quad \text{e} \quad E = \begin{bmatrix} 214{,}4060 & -32{,}6613 & -11{,}5724 \\ -32{,}6613 & 33{,}6998 & 5{,}4890 \\ -11{,}5724 & 5{,}4890 & 3{,}9065 \end{bmatrix}$$

O polinômio característico de $(H-\lambda E)$ é dado por:

$P(H-\lambda E) = 17235{,}2729\ \lambda^3 - 43987{,}1014\ \lambda^2 + 21380{,}1576\ \lambda - 1568{,}0027$, cujas raízes são:

$\lambda_1 = 1{,}935539347$; $\lambda_2 = 0{,}527513242$ e $\lambda_3 = 0{,}089103192$.

A Tabela 14 mostra os autovalores e sua explicação individual e acumulada da variação total.

Tabela 14. Autovalores e explicação da variância total.

Componente Principal	**Autovalor**	**Explicação (%)**	**Explicação Acumulada (%)**
1	1,935539347	75,84	75,84
2	0,527513242	20,67	96,51
3	0,089103192	3,49	100,00

Observa-se a contribuição do primeiro autovalor expressa por:

$$100\ \frac{\lambda_1}{\lambda_1 + \lambda_2 + \lambda_3}\ \% = 100\ \frac{1{,}935539347}{2{,}552155781}\% = 75{,}84\%$$

Para $\lambda_1 = 1{,}935539347$ a solução normalizada (μ_1^*) para o sistema linear homogêneo será:

$$u_1^* = \begin{bmatrix} -0{,}008620433 \\ 0{,}516619682 \\ -0{,}85617159 \end{bmatrix}$$

Portanto, o primeiro componente principal pode ser escrito como:

Y_1 = - 0,008620433 G + 0,516619682 ME - 0,85617159 UM

Conforme item 5.7 da página 33, verificou-se que as correlações entre o primeiro componente principal e as classes de classificação da batatinha, são da ordem de -0,005500277, 0,668306640 e -4,1057946433 para graúda, média e miúda, respectivamente. Nota-se que as associações mais expressivas ocorreram com as classes média e miúda.

Substitui-se o valor de cada classe dentro de cada um dos blocos no primeiro componente principal obtido, e têm-se os resultados constantes no Quadro 14.

Quadro 14. Valores observados transformados a partir do primeiro componente principal segundo tratamento e bloco.

Bloco	**Tratamento**				
	Anosta	**Gelda**	**Recent**	**Bintje**	**Elvira**
B_1	0,8103	0,4857	1,0448	1,4240	0,9373
B_2	0,4747	0,9841	0,8447	2,6783	0,9875
B_3	1,0434	0,7692	1,0284	1,5461	0,7864
B_4	0,3311	0,8219	0,8964	1,3080	0,8587
B_5	1,5125	1,7070	0,7880	2,4082	1,2953
B_6	0,7780	1,9535	1,1117	3,6022	1,1074
B_7	1,4600	1,4494	0,6182	2,5756	0,7587
B_8	1,1041	1,5699	1,0913	2,3835	0,7527
B_9	1,4402	1,0439	0,6817	2,0684	0,4538
B_{10}	1,9924	1,0875	0,8141	2,2272	0,3831
B_{11}	1,1978	1,0287	0,1664	1,2815	0,4349
B_{12}	0,6257	0,5783	0,6388	1,3632	-0,0249

O quadro abaixo apresenta o resultado da análise de variância dos dados transformados por Y_1 = - 0,008620433 G + 0,516619682 ME - 0,85617159 MU

Quadro 15. Quadro da análise de variância univariada (ANOVA).

Causa de variação	**GL**	**SQ**	**QM**	***F***
Tratamento	4	13,82212	3,45553	*P*<0,05
Bloco	11	5,58638	0,50785	
Resíduo	44	7,13830	0,16223	
Total	59	26,54680		

Os resultados do Quadro 15 mostram que existe diferença significativa (P<0,05) entre as cultivares de batatinha.

Em continuidade ao estudo, procedeu-se a comparação de pares de médias pelo teste de Tukey.

Ou seja,

DHS (α) = $q_{(\alpha;k;\rho_{res})}\sqrt{\frac{QM\,\mathrm{Re}s}{b}}$ com $q_{(5\%;5;44)}$ = 4,0280,

resultando,

DHS (α) = 0,4683

As médias dos dados transformados ($\overline{T_1},\overline{T_2},\overline{T_3},\overline{T_4},\overline{T_5}$) com os desvios-padrão respectivos, expressos na seqüência, com as respectivas comparações entre todos os possíveis pares, estão estabelecidos a seguir:

$\overline{T_5}(Y_1)$ = 0,7276 ± 0,36 a

$\overline{T_3}(Y_1)$ = 0,8104 ± 0,27 a

$\overline{T_1}(Y_1)$ = 1,0642 ± 0,49 a

$\overline{T_2}(Y_1)$ = 1,1233 ± 0,46 a

$\overline{T_4}(Y_1)$ = 2,0722 ± 0,71 b

Pode-se concluir que no nível de 5% de significância existe diferença entre as cultivares estudadas quanto à produtividade, em kg por parcela, sendo a cultivar Bintje a mais produtiva entre as analisadas.

Portanto, os resultados obtidos em relação às comparações entre as produções das cultivares de batatinhas, segundo a classificação de tubérculos da EMBRAPA, mostraram-se em perfeita concordância nos métodos utilizados.

7. CONCLUSÕES

Concluiu-se a partir deste estudo que:

Um dos fatores da pouca utilização dos procedimentos multivariados na área agronômica, pode ser devido à carência de programas computacionais de fácil manipulação.

Outro fator deve-se à dificuldade de utilização de conhecimento estatístico-computacional apropriado para as estruturas de inter-relação existente nas variáveis biológicas, fatos que tornam complexos os procedimentos analíticos para a acurada e eficiente interpretação dos dados experimentais.

O sistema ACOMP vem oferecer aos pesquisadores, possibilidades da utilização de técnicas multivariadas, mais especificamente a análise dos componentes principais, na experimentação agronômica ou em outras áreas afins. O sistema foi construído objetivando oferecer ao usuário, facilidade na manipulação do "software" e entendimento de sua operacionalização, pois foi desenvolvido de maneira auto-explicativa.

Com a finalidade de esclarecer possíveis dificuldades encontradas na utilização de procedimentos multivariados, foram apresentados dois exemplos práticos, sendo um sobre o uso da técnica dos componentes principais como um processo gráfico de comparação de tratamentos (isolados) e outro, envolvendo delineamento experimental multivariado analisado pelas técnicas da MANOVA e ANOVA, sendo a última pelos dados transformados a partir do primeiro componente principal.

8. REFERÊNCIAS BIBLIOGRÁFICAS

ABELL, M. L.; BRASELTON, J. P. **The maple V handbook**, Boston, USA: Academic Press Limited, 1994. 730 p.

ALVES, R.M. et al. Seleção de descritores botânico-agronômicos para caracterização de germoplasma de cupuaçuzeiro. **Pesquisa Agropecuária Brasileira**, Brasília, DF, v. 38, n. 7, p. 807-818, jul. 2003.

ANDERSON, T.W. **An introduction to multivariate statistical analysis**. John Wiley, New York, 1958. 374 p.

ANDRADE, J. B. et al. Seleção de 39 variedades de cana-de-açúcar para a alimentação animal. **Brazilian Journal of Veterinary Research and Animal Science**, São Paulo, SP, v. 40, n. 4, p. 287-296, 2003.

ARTES, R. Aspectos estatísticos da análise fatorial de escalas de avaliação. **Revista de Psiquiatria Clínica**, São Paulo, SP, v. 25, n. 5, p. 223-228, 1998.

AZEVEDO, J. A.G. et al. Avaliação da divergência nutricional de variedades de cana-de-açúcar (Saccharum spp.). **Revista Brasileira de Zootecnia**, Viçosa, MG, v. 32, n. 6, p. 1431-1442, nov./dez. 2003.

BEHRENS, J. H.; SILVA, M. A. A. P. Perfil sensorial de vinhos brancos varietais brasileiros através de análise descritiva quantitativa. **Ciência e Tecnologia de Alimentos**, Campinas, SP, v. 20, n. 1, p. 60-67, abr. 2000.

BOZA, Y.; HORII, J. Influência da destilação sobre a composição e a qualidade sensorial da aguardente de cana-de-açúcar. **Ciência e Tecnologia de Alimentos**, Campinas, SP, v. 18, n. 4, p. 391-396, Out./Dez. 1998.

BRUNE, S.; MELO, P. E. Método rápido de avaliação do esverdeamento em tubérculos de batata. **Pesquisa Agropecuária Brasileira**, Brasília, DF, v. 36, n. 5, p. 809-814, maio 2001.

BUSSAB, W. O.; MORETTIN, P. A. **Estatística básica**. 5. ed. São Paulo: Saraiva, 2003. 526 p.

CARDELLO, H. M. A. B.; SILVA, M. A. A. P.; DAMÁSIO, M. H. Análise descritiva quantitativa de edulcorantes em diferentes concentrações. **Ciência e Tecnologia de Alimentos**, Campinas, SP, v. 20, n. 3, p. 318-328, set./dez. 2000.

CARDOSO, M. et al.**Temas em métodos quantitativos.** Lisboa: Edições Sílabo, 2004. 274 p.

CAVALCANTI, J. J. V. et al. Repetibilidade de caracteres de produção e porte da planta em clones de cajueiro-anão precoce. **Pesquisa Agropecuária Brasileira**, Brasília, DF, v. 35, n. 4, p. 773-777, abr. 2000.

CURI, P. R.; PADOVANI, C. R. Função discriminante linear: aplicação em experimentos biológicos. **Naturalia**, Rio Claro, SP, v. 5, p. 59-77, 1980.

CURI, P. R. **Análise de agrupamento complementada com ordenação pelos componentes principais e análise de variância multivariada. Aplicação em experimentos biológicos**. 1982. 107 f. Tese de Livre-Docência/ Departamento de Bioestatística – IBBMA-Universidade Estadual Paulista, Botucatu, SP, 1982.

DEMEY. J. R.; ADAMS, M.; FREITES, H. Uso del método de analisis de componentes principales para la caracterizacion de fincas agropecuarias. **Agronomia Tropical**, Maracay, Venezuela, v. 44, n. 3, p. 475-497, 1994.

EIRA, A. F.; SOUZA, N. L. Aparato de precisão para medição do volume de órgãos vegetais ou objetos de geometria irregular. **Summa Phytopathologica**, São Paulo, SP, v. 10, p. 34-35, 1984.

FARIAS NETO, J. T.; LINS, P. M. P.; MÜLLER, A. A. Estimativa dos coeficientes de repetibilidade para produção de fruto e albúmem sólido em coqueiro híbrido. **Pesquisa Agropecuária Brasileira**, Brasília, DF, v. 38, n. 10, p. 1237-1241, out. 2003.

FERREIRA, E. C. et al. Análise exploratória dos teores de constituintes inorgânicos em sucos e refrigerantes de uva. **Ecletica Quimica**, Araraquara, SP, v. 27, n. spe, p. 77-90, 2002.

FLECK, M.P.A.; BOURDEL, M.C. Método de simulação e escolha de fatores na análise dos principais. **Revista de Saúde Pública**, São Paulo, SP, v. 32, n. 3, p. 267-272, Jun.1998.

FONSECA, R. C. B.; FONSECA, I.C. B. Utilização de métodos estatísticos na caracterização do mosaico sucessional em floresta semidecídual. **Revista Árvore**, Viçosa, MG, v. 28, n. 3, p. 351-359, maio/jun. 2004.

FREIRE FILHO, F.R. et al. Adaptabilidade e estabilidade da produtividade de grãos de genótipos de caupi enramador de tegumento mulato. **Pesquisa Agropecuária Brasileira**, Brasília, DF, v. 38, n. 5, p. 591-598, maio 2003.

GOLFETO, J. H. et al. Propriedades psicométricas do inventário da depressão infantil (CDI) aplicado em uma amostra de escolares de Ribeirão Preto. **Revista de Psiquiatria Clínica**, São Paulo, SP, v. 29, n. 2, p. 66-73, 2002.

GOMES, F. P.; CONAGIN, A. Experimentos de adubação: planejamento e análise estatística. In: EMBRAPA. **Métodos de pesquisa em fertilidade do solo.** Londrina, PR: Departamento de Matemática Aplicada-UEL e Região Brasileira da Sociedade Internacional de Biometria. 1987. p. 82-87.

GOMES, J. B. V. et al. Análise de componentes principais de atributos físicos, químicos e mineralógicos de solos do bioma cerrado. **Revista Brasileira da Ciência do Solo**, Campinas, SP, v. 28, n. 1, p. 137-153, jan./fev. 2004.

GORESTEIN, C.; ANDRADE, L. Inventário de depressão de Beck: propriedades psicométricas da versão em português. **Revista de Psiquiatria Clínica**, São Paulo, SP, v. 25, n. 5, edição especial, p. 245-250, 1998.

GRAYBILL, F. A. **Introduction to matrices with applications in statistics**. Belmont: Wadsworth, 1969. 372 p.

GUARIENTI, E. M. et al. Avaliação do efeito de variáveis meteorológicas na qualidade industrial e no rendimento de grãos de trigo pelo emprego de análise de componentes principais. **Ciência e Tecnologia de Alimentos**, Campinas, SP, v. 23, n. 3, p. 500-510, set./dez. 2003.

HAIR, J. F. JR. et al. **Análise multivariada de dados**. 5. ed. Porto Alegre: Bookman, 2005. 593 p.

HORN, J. L.; ENGSTROM, R. Cattell's scree test in relation to Bartlett's chi square test and other observations on the number of factors problem. **Multivariate Behavioral Research**, Fort Worth, Tex.,US, v. 14, p. 283-300, 1979.

HOTTELING, H. Analysis of a complex of statistical variables into principal components. **The Journal Educational Psychology**, Cambridge, USA, v. 24, p. 417-41, 498-520, 1933.

HOTTELING, H. Simplifield calculation of principal components. **Psychometrika**, Williamsburg, Va, US, v. 1, p. 27-35, 1936.

JENSEN, A.; MARK, W. Henry Felix Kaiser (1927-1992): Obituary. **American-Psychologist**. Washington, USA, v. 49, n. 12, p. 1085, dez. 1994.

JOHNSON, R.A.; WICHERN, D.W. **Applied multivariate statistical analysis**. 3*rd* ed. Upper Saddle River, New Jersey: Prentice Hall, 1992. 642 p.

JOLICOUER, P.; MOSIMANN, J.E. Size and shape variation in the painted turtle. A principal component analysis. **Growth**, Bar Harbor, USA, v. 24, p. 339-354, 1960.

KENDALL, M.G.Factor Analysis. **Journal of the Royal Statistical Society:** Série B, London, GB, v. 12, p. 60-94, 1950.

LLORÀ, E. B. **Técnicas Computacionales: Métodos Numéricos y Programación**. Desenvolvido pela Universitat de Girona, 2004. Disponível em : http://www.stark.udg.es/~emili/docent/qtc/pdf/06_diag_pca.pdf . Acesso em: 12 dez. 2004.

LOPES, R. et al. Repetibilidade de características do fruto de aceroleira. **Pesquisa Agropecuária Brasileira**, Brasília, DF, v. 36, n. 3, p. 507-513, mar. 2001.

LOPES, F.C.F. et al. Relação entre variáveis na regulação do consumo de vacas Holandês × Zebu em lactação sob pastejo em capim-elefante. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, Belo Horizonte, MG, v. 56, n. 1, p. 52-60, fev. 2004.

MACHADO, C. F. et al. Divergência genética entre genótipos de feijoeiro a partir de técnicas multivariadas. **Revista Ciência Rural**, Santa Maria, RS, v. 32, n. 2, p. 251-258, mar./abr. 2002.

MALLANTS, D.; FEYEN, J. Defining homogeneous precipitation regions by means of principal components analysis – **Journal of Applied Meteorology**, v. 29, n. 9, p. 892-901, 1990.

MARQUES, A.; COELHO, A. A aplicação das estratégias relacionais em Portugal. In: JORNADAS HISPANO-LUSAS DE GESTIÓN CIENTÍFICA, 13., 2003, Lugo. **Anais eletrônicos**... Lugos: USC.ES, 2003. Disponível em : <http:// www.ti.usc.es/lugo-xiii-hispano-lusas/04_programa.htm>. Acesso em: 20 dez.2004.

MARTEL, J.H.I. et al. Estatística multivariada na discriminação de raças amazônicas de pupunheiras (Bactris gasipaes Kunth) em Manaus (Brasil). **Revista Brasileira de Fruticultura**, Cruz das Almas, BA, v. 25, n. 1, p. 115-118, abr. 2003.

MARTINELLO, G. E. et al. Divergência genética em acessos de quiabeiro com base em marcadores morfológicos. **Revista Horticultura Brasileira**, Brasília, DF, v. 20, n. 1, p. 52-58, mar. 2002.

MASCIOLI, A.S. et al. Estimativas de parâmetros genéticos e fenotípicos e análise de componentes principais para características de crescimento na raça Canchim. **Revista Brasileira de Zootecnia**, Viçosa, MG, v. 29, n. 6, p. 1654-1660, nov./dez. 2000.

MEDEIROS, C. A. F. **Comprometimento organizacional: um estudo de suas relações com características organizacionais e desempenho nas empresas hoteleiras.** 2003. 166 f. Tese (Doutorado em Administração de Empresas)-Faculdade de Economia, Administração e Contabilidade, Universidade de São Paulo, São Paulo, 2003.

MESSETTI, A. V. L.; PADOVANI, C.R. Utilização da análise de agrupamento no estudo da divergência genotípica de girassol (*Helianthus annuus L.*). **Energia na Agricultura**, Botucatu, SP, v. 15, n. 4, p. 26-35, 2000.

MONTEIRO, V. P.; PINHEIRO, J. C. V. Critério para implantação de tecnologias de suprimentos de água potável em municípios cearenses afetados pelo alto teor de sal. **Revista de Economia e Sociologia Rural**, Brasília, DF, v. 42, n. 2, p. 365-387, abr./jun. 2004.

MORGANO, M. A.; QUEIROZ, S. C. N.; FERREIRA, M. M. C. Determinação dos teores de minerais em sucos de frutas por espectrometria de emissão óptica em plasma indutivamente acoplado (ICP-OES). **Ciência e Tecnologia de Alimentos**, Campinas, SP, v. 19, n. 3, p. 344-348, set./dez. 1999.

MORRISON, D. F. **Multivariate statistical methods**. 4th ed. New York: Duxbury Press, 2003. 415 p.

MOTTA, P. E.F. et al. Ocorrência da macaúba em Minas Gerais: relação com atributos climáticos, pedológicos e vegetacionais. **Pesquisa Agropecuária Brasileira**, Brasília, DF, v. 37, n. 7, p. 1023-1031, jul. 2002.

NORMAN, G. R. e STREINER, D. L. **Biostatistics:** the bare essentials. St. Louis: Mosby Year Book, 1994. 260 p.

OLIVEIRA, L. G. F.; PADOVANI, C. R. Aplicação da técnica dos componentes principais, em experimentos em blocos completamente casualizados multivariados. **Energia na Agricultura**, Botucatu, SP, v. 16, n. 1, p. 33-40, 2001.

PADOVANI, C.R.P. **Desenvolvimento de aplicativo para o método de discriminação de Fisher e seu uso na experimentação agronômica**. 2004. 99 f. Dissertação (Mestrado em Agronomia/ Energia na Agricultura) – Faculdade de Ciências Agronômicas, Universidade Estadual Paulista, Botucatu, SP, 2004.

PEARSON, K. On lines and planes of closest fitto systems of points in space.**Philosophical Magazine,** London, GB**,** v. 2, p. 559-72, 1901.

PERONI, N; MARTINS, P. S.; ANDO, A. Diversidade inter- e intra-específica e uso de análise multivariada para morfologia da mandioca (Manihot esculenta Crantz): um estudo de caso. **Science Agricola**, Piracicaba, SP, v. 56, n. 3, p. 587, Jul. 1999.

PINTO, G. S.; ÁVILA-PIRES, T.C.S. Crescimento alométrico, morfologia e uso do habitat em cinco espécies de Mabuya Fitzinger (Reptilia, Scincidae). **Revista Brasileira de Zoologia**, Curitiba, PR, v. 21, n. 2, p. 161-168, jun. 2004.

RAO, A. et al. Genetic divergence among some brown plant hopper resistant rice varieties. **Indian Journal of Genetics & Plant Breeding**, New Delhi, India, v. 41, p. 179-185, 1981.

REIS, E. **Estatística Multivariada Aplicada**. Lisboa. Edições Sílabo, 1997. 343 p.

RODRIGUES, L. S. et al. Divergência genética entre cultivares locais e cultivares melhoradas de feijão. **Pesquisa Agropecuária Brasileira**, Brasília, DF, v. 37, n. 9, p. 1285-1294, set. 2002.

SHIMOYA, A. et al. Repetibilidade de características forrageiras do capim-elefante. **Science Agrícola**, Piracicaba, SP, v. 59, n. 2, p. 227-234, abr./jun. 2002.

SILVA, F.V. et al. Discriminação geográfica de águas minerais do Estado de São Paulo através da análise exploratória**. Ecletica Quimica**, Araraquara, SP, v. 27, n. spe, p. 91-102, 2002.

SILVA, J.A. et al. Uso da terra como determinante da distribuição da raiva bovina em Minas Gerais, Brasil. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, Belo Horizonte, MG, v. 53, n. 3, p. 273-283, jun. 2001.

SILVAROLLA, M. B. et al. Avaliação de progênies derivadas do híbrido de timor com resistência ao agente da ferrugem. **Bragantia**, Campinas, SP, v. 56, n. 1, p. 47-58, 1997.

SILVEIRA, D. X.; JORGE, M. R. Propriedades psicométricas da escala de rastreamento populacional para depressão CES-D em populações clínica e não-clínica de adolescentes e adultos jovens. **Revista de Psiquiatria Clínica**, São Paulo, SP, v. 25, n. 5, edição especial, p. 251-261, 1998.

SOARES, P.L. M.; SANTOS, J.M.; FERRAUDO, A.S. Estudo morfométrico comparativo de 58 populações brasileiras de Rotylenchulus reniformis (Nemata: Rotylenchulinae). **Fitopatologia Brasileira**, Brasília, DF, v. 29, n. 4, p. 419-424, jul./ago. 2004.

SOUZA, N. L. **Podridão de raízes de milho causada por isolados do complexo** *Pythium graminicolum – Pythium arrhenomanes* **sensu Watanabe, 1975**. 1985.129 f. Dissertação (Doutorado em Agronomia/ Fitopatologia) – Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 1985.

SOUZA-SOBRINHO, F. et al. Estimativas de repetibilidade para produção de matéria seca em alfafa. **Revista Ciência Rural**, Santa Maria, RS, v. 34, n. 2, p. 531-537, mar./abr. 2004.

STRAPASSON, E.; VENCOVSKY, R.; BATISTA, L. A. R. Seleção de descritores na caracterização de germoplasma de Paspalum sp. por meio de componentes principais. **Revista Brasileira de Zootecnia**, Viçosa, MG, v. 29, n. 2, p. 373-381, mar./abr. 2000.

STUKER, H. **Análise multivariada para dados onde a característica observada é subdividida em k classes**. 1986. 92 f. Dissertação (Mestrado em Agronomia/ Estatística e Experimentação Agronômica) Escola Superior de Agronomia "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 1986.

TEIXEIRA, D. L.; SPEHAR, C. R.; SOUZA, L. A. C. Caracterização agronômica de amaranto para cultivo na entressafra no Cerrado. **Pesquisa Agropecuária Brasileira**, Brasília, DF, v. 38, n. 1, p. 45-51, jan. 2003.

VIDA, G. et al. Use of principal components analysis in the selection of optimum agronomics treatments for winter wheat (Triticum aestivum L.) varieties with various baking qualities. **Acta Agronomica Hungarica**, Budapest, Hungary, v. 44, n. 1, p. 1-9, 1996.

VIDAL, A. G. R. **Clipper**: versão summer 87. Rio de Janeiro: Livros Técnicos e Cientificos Editora Ltda, 1989. 855p.

ZAMBRANO, A.Y.; DEMEY, J. R.; GONZÁLEZ, V. Grupos homogeneos de crecimiento y manipulación *in vitro* de seis cultivares comerciales de caña de azucar en Venezuela. **Agronomia Tropical**, Maracay, Venezuela, v. 45, n. 1, p. 51-72, 1995.

APÊNDICE

```
> with(linalg):
> with(stats):
> with(plots):
> ###############################################################################
> ###                  SISTEMA DE ANALISE DOS COMPONENTES PRINCIPAIS
> ###############################################################################
> ###    Variáveis de inicialização
> ###           tcond    =>    número de  Variáveis pesquisadas
> ###           Telm       =>    matriz com número de  Unidades Experimentais
> ###           A             =>     matriz com os dados apurados
> tcond:=06:
> Telm:=matrix(1,001,[011]):
> Ret_Clip := EXEC01O :
>
A:=matrix(00011,06,[11.030000,10.960000,10.510000,11.020000,0.830000,0.920000,10.250000,10.
47000
>
0,9.650000,12.230000,0.950000,0.980000,12.230000,6.640000,8.750000,11.450000,0.740000,1.200
000,1
>
4.060000,9.970000,9.630000,13.490000,1.120000,1.260000,11.830000,8.360000,10.020000,12.6800
00,0.
>
980000,1.120000,12.850000,6.680000,8.070000,13.080000,0.710000,1.040000,12.200000,6.400000,
9.130
>
000,10.510000,0.920000,0.810000,12.400000,7.340000,9.090000,11.580000,1.060000,1.820000,7.2
00000
>
,4.710000,3.330000,4.680000,0.920000,0.680000,11.850000,8.470000,7.180000,8.800000,1.060000
,1.740
> 000,16.600000,18.820000,18.320000,20.650000,1.840000,2.060000]):
> readlib(write):
> open(Ret_Clip):
> ###############################################################################
> ###       calcula media das variáveis   => medcol[i]
> ###############################################################################
> tgp:=1:
> medcol := array(1..tgp,1..tcond):
> writeln(MediaColunas):
> deslin := 0:
> for l from 1 to tgp do:
>         for j from 1 to tcond   do   medcol[l,j] := 0:
>                 for i from 1 to Telm[1,l] do nlin := i + deslin:
>                         medcol[l,j] := medcol[l,j] + A[nlin,j] od:
>                 medcol[l,j] := evalf(medcol[l,j] / Telm[1,l],6) :
>                 writeln(l,j,medcol[l,j]) od:
>         deslin := deslin + Telm[1,l]:
> od:
> totelem := deslin:
> # -----------------------------------
> #print(medcol);
> # -----------------------------------
> ###############################################################################
```

```
> ###    calc matriz de variância/covariância  => matval[j,j]    (matriz S)
> ##########################################################################################
> matval := array(1..tgp,1..tcond,1..tcond):
> deslin := 0:
> for l from 1 to tgp do:
>       for j from 1 to tcond   do   matval[l,j,j]:=0:
>             for i from 1 to Telm[1,l] do nlin := i + deslin:
>                   matval[l,j,j] := matval[l,j,j] +  A[nlin,j]^2  od :
>             matval[l,j,j] := ( matval[l,j,j] - Telm[1,l] * medcol[l,j]^2) / (Telm[1,l]-1)
od:
>       for i from 2 to tcond     do:
>             for j from 1 to i-1    do   matval[l,i,j] := 0:
>                   for k from 1 to Telm[1,l]  do nlin := k + deslin:
>                         matval[l,i,j] := matval[l,i,j] + A[nlin,i] * A[nlin,j] od:
>                   matval[l,i,j] := evalf(( matval[l,i,j] - Telm[1,l] * medcol[l,i] *
medcol[l,j] ) / (Telm[1,l]-1),6):
>                   matval[l,j,i] := matval[l,i,j] od od:
>       deslin := deslin + Telm[1,l] od:
> writeln(MatrizCovariancia):
> smatriz := array(1..tcond,1..tcond):
> for l from 1 to tgp do:
>       for i from 1 to tcond do:
>             for j from 1 to tcond   do  writeln(l,i,j,matval[l,i,j]):
>                      smatriz[i,j] := matval[l,i,j]:
> od  od  od:
> #
> ##########################################################################################
> ###    calc matriz de correlação  => matcorr[j,j]     (matriz R)
> ##########################################################################################
> matcorr := array(1..tgp,1..tcond,1..tcond):
> deslin := 0:
> for l from 1 to tgp do:
>       for j from 1 to tcond   do:
>                matcorr[l,j,j]:= matval[l,j,j] / evalf(sqrt( matval[l,j,j] ^ 2 ))      od:
>       for i from 2 to tcond     do:
>             for j from 1 to i-1      do:
>                   matcorr[l,i,j] := matval[l,i,j] / evalf(sqrt( matval[l,i,i] *
matval[l,j,j] ))  :
>                   matcorr[l,j,i] := matcorr[l,i,j]   od od :
>       deslin := deslin + Telm[1,l] od:
> writeln(MatrizCorrelacao):
> for l from 1 to tgp do:
>       for i from 1 to tcond do:
>             for j from 1 to tcond   do  writeln(l,i,j,matcorr[l,i,j])  od  od  od:
> # -----------------------------------
> #print(matcorr);
> # -----------------------------------
> ##########################################################################################
> ###  Raizes características (Autovalores) e Autovetores
> ###                         ( raizes1   -> autovalores de S    //    raizes2   ->
Autovetores de S
> ###                         ( raizes11 -> autovalores de R    //    raizes12 ->
Autovetores de R
> ##########################################################################################
> XX := matrix(tcond,tcond):
```

```
> for i from 1 to tcond  do          for j from 1 to tcond  do       XX[i,j] :=
matval[1,i,j]      od  od:
> raizes1:=array(1..tcond):
> raizes1:=eigenvals(XX):
> raizes2:=array(1..tcond):
> raizes2:=evalf(eigenvects(XX, 'radical'),6):
> # -----------------------------------
> #print(raizes1);
> #print(raizes2);
> # -----------------------------------
> YY := matrix(tcond,tcond):
> for i from 1 to tcond  do          for j from 1 to tcond  do       YY[i,j] :=
matcorr[1,i,j]      od  od:
> raizes11:=array(1..tcond):
> raizes11:=eigenvals(YY):
> raizes12:=array(1..tcond):
> raizes12:=evalf(eigenvects(YY, 'radical'),6):
> # -----------------------------------
> #print(raizes11);
> #print(raizes12);
> # -----------------------------------
> ###
> ##########################################################################################
> ###  Normalização dos Autovetores de S
> ##########################################################################################
> ###
> snorm := array(1..tcond,1..tcond):
> snorm2 := array(1..tcond,1..tcond):
> xvalor := array(1..tcond):
> for i from 1 to tcond do:
>        lixo := raizes2[i]:
>        xvalor[i] := lixo[1]:
>        lixo1 := lixo[3];
>        lixo2 := lixo1[1];
>        for j from 1 to tcond do:
>              snorm[i,j] := lixo2[j]:
>              snorm2[i,j] := 0:
>        od:
> od:
> #
> for i from 1 to tcond do:
>         divix := 0:
>          for j from 1 to tcond do:
>                divix := evalf(divix + ( snorm[i,j] ^ 2 )):
>          od:
>          divinorm := evalf(sqrt( divix )):
>          for j from 1 to tcond do:
>                   snorm2[i,j] := evalf( snorm[i,j] / divinorm):
>          od:
> od:
> ##### testando a normalização
> #for i from 1 to tcond do:
> #        divix := 0:
> #         for j from 1 to tcond do:
> #               divix := evalf(divix + ( snorm2[i,j] ^ 2 )):
> #         od:
```

```
> #         print(i,divix);
> #od:
> #
> #
> ###############################################################################
> ###  Normalização dos Autovetores de R
> ###############################################################################
> rnorm := array(1..tcond,1..tcond):
> rnorm2 := array(1..tcond,1..tcond):
> yvalor := array(1..tcond):
> for i from 1 to tcond do:
>        lixo := raizes12[i]:
>        yvalor[i] := lixo[1]:
>        lixo1 := lixo[3];
>        lixo2 := lixo1[1];
>        for j from 1 to tcond do:
>              rnorm[i,j] := lixo2[j]:
>              rnorm2[i,j] := 0:
>        od:
> od:
> #
> for i from 1 to tcond do:
>         divix := 0:
>          for j from 1 to tcond do:
>                divix := evalf(divix + ( rnorm[i,j] ^ 2 )):
>          od:
>          divinorm := evalf(sqrt( divix )):
>          for j from 1 to tcond do:
>                   rnorm2[i,j] := evalf( rnorm[i,j] / divinorm):
>          od:
> od:
>
> ###############################################################################
> ###  Classificando os Autovetores de S e R
> ###############################################################################
> svetor := array(1..tcond,1..tcond):
> rvetor := array(1..tcond,1..tcond):
> svalor := array(1..tcond):
> rvalor:= array(1..tcond):
> xclass := array(1..tcond):
> yclass := array(1..tcond):
> #
> for i from 1 to tcond do xclass[i] := 0:
>                                              yclass[i] := 0 od:
> #
> for i from 1 to tcond do:
>          maior := -999999999999:
>          indmor := 0:
>          for j from 1 to tcond do:
>                if xclass[j] = 0 then:
>                    if xvalor[j] > maior then maior := xvalor[j]:
>                                                                  indmor := j fi fi:
>           od:
>           svalor[i] := maior:
>           xclass[indmor] := i:
>           for j from 1 to tcond do:
```

```
>                svetor[i,j] := snorm2[indmor,j]:
>            od:
> od:
> #
> for i from 1 to tcond do:
>          maior := -999999999999:
>          indmor := 0:
>          for j from 1 to tcond do:
>               if yclass[j] = 0 then:
>                   if yvalor[j] > maior then maior := yvalor[j]:
>                                                              indmor := j fi fi:
>           od:
>           rvalor[i] := maior:
>           yclass[indmor] := i:
>           for j from 1 to tcond do:
>                 rvetor[i,j] := rnorm2[indmor,j]:
>            od:
> od:
> #
> writeln(AutoVetor_S):
> for i from 1 to tcond do:
>       writeln(VALOR,i,svalor[i]):
>       for j from 1 to tcond do:
>             writeln(VETOR,i,j,svetor[i,j]):
>        od:
> od:
> #
> writeln(AutoVetor_R):
> for i from 1 to tcond do:
>       writeln(VALOR,i,rvalor[i]):
>       for j from 1 to tcond do:
>             writeln(VETOR,i,j,rvetor[i,j]):
>        od:
> od:
> #
> if tcond = 2 then:
>    desenho := [[1,svalor[1]],[2,svalor[2]]]:
> fi:
> if tcond = 3 then:
>    desenho := [[1,svalor[1]],[2,svalor[2]],[3,svalor[3]]]:
> fi:
> if tcond = 4 then:
>    desenho := [[1,svalor[1]],[2,svalor[2]],[3,svalor[3]],[4,svalor[4]]]:
> fi:
> if tcond = 5 then:
>    desenho := [[1,svalor[1]],[2,svalor[2]],[3,svalor[3]],[4,svalor[4]],[5,svalor[5]]]:
> fi:
> if tcond = 6 then:
>    desenho :=
[[1,svalor[1]],[2,svalor[2]],[3,svalor[3]],[4,svalor[4]],[5,svalor[5]],[6,svalor[6]]]:
> fi:
> if tcond = 7 then:
>    desenho :=
[[1,svalor[1]],[2,svalor[2]],[3,svalor[3]],[4,svalor[4]],[5,svalor[5]],[6,svalor[6]],[7,sva
lor[7]]]:
> fi:
```

```
> if tcond = 8 then:
>    desenho :=
[[1,svalor[1]],[2,svalor[2]],[3,svalor[3]],[4,svalor[4]],[5,svalor[5]],[6,svalor[6]],[7,sva
lor[7]],[
> 8,svalor[8]]]:
> fi:
> if tcond = 9 then:
>    desenho :=
[[1,svalor[1]],[2,svalor[2]],[3,svalor[3]],[4,svalor[4]],[5,svalor[5]],[6,svalor[6]],[7,sva
lor[7]],[
> 8,svalor[8]],[9,svalor[9]]]:
> fi:
> if tcond = 10 then:
>    desenho :=
[[1,svalor[1]],[2,svalor[2]],[3,svalor[3]],[4,svalor[4]],[5,svalor[5]],[6,svalor[6]],[7,sva
lor[7]],[
> 8,svalor[8]],[9,svalor[9]],[10,svalor[10]]]:
> fi:
> if tcond = 11 then:
>    desenho :=
[[1,svalor[1]],[2,svalor[2]],[3,svalor[3]],[4,svalor[4]],[5,svalor[5]],[6,svalor[6]],[7,sva
lor[7]],[
> 8,svalor[8]],[9,svalor[9]],[10,svalor[10]],[11,svalor[11]]]:
> fi:
> if tcond = 12 then:
>    desenho :=
[[1,svalor[1]],[2,svalor[2]],[3,svalor[3]],[4,svalor[4]],[5,svalor[5]],[6,svalor[6]],[7,sva
lor[7]],[
> 8,svalor[8]],[9,svalor[9]],[10,svalor[10]],[11,svalor[11]],[12,svalor[12]]]:
> fi:
> if tcond = 13 then:
>    desenho :=
[[1,svalor[1]],[2,svalor[2]],[3,svalor[3]],[4,svalor[4]],[5,svalor[5]],[6,svalor[6]],[7,sva
lor[7]],[
>
8,svalor[8]],[9,svalor[9]],[10,svalor[10]],[11,svalor[11]],[12,svalor[12]],[13,svalor[13]]]
:
> fi:
> if tcond = 14 then:
>    desenho :=
[[1,svalor[1]],[2,svalor[2]],[3,svalor[3]],[4,svalor[4]],[5,svalor[5]],[6,svalor[6]],[7,sva
lor[7]],[
>
8,svalor[8]],[9,svalor[9]],[10,svalor[10]],[11,svalor[11]],[12,svalor[12]],[13,svalor[13]],
[14,svalor[14]]]:
> fi:
> if tcond = 15 then:
>    desenho :=
[[1,svalor[1]],[2,svalor[2]],[3,svalor[3]],[4,svalor[4]],[5,svalor[5]],[6,svalor[6]],[7,sva
lor[7]],[
>
8,svalor[8]],[9,svalor[9]],[10,svalor[10]],[11,svalor[11]],[12,svalor[12]],[13,svalor[13]],
[14,svalor[14]],[15,
> svalor[15]]]:
> fi:
> plot(desenho,style=line,title=`Auto Valores de S`);
```

```
> #
> if tcond = 2 then:
>    desenhor := [[1,rvalor[1]],[2,rvalor[2]]]:
> fi:
> if tcond = 3 then:
>    desenhor := [[1,rvalor[1]],[2,rvalor[2]],[3,rvalor[3]]]:
> fi:
> if tcond = 4 then:
>    desenhor := [[1,rvalor[1]],[2,rvalor[2]],[3,rvalor[3]],[4,rvalor[4]]]:
> fi:
> if tcond = 5 then:
>    desenhor := [[1,rvalor[1]],[2,rvalor[2]],[3,rvalor[3]],[4,rvalor[4]],[5,rvalor[5]]]:
> fi:
> if tcond = 6 then:
>    desenhor :=
[[1,rvalor[1]],[2,rvalor[2]],[3,rvalor[3]],[4,rvalor[4]],[5,rvalor[5]],[6,rvalor[6]]]:
> fi:
> if tcond = 7 then:
>    desenhor :=
[[1,rvalor[1]],[2,rvalor[2]],[3,rvalor[3]],[4,rvalor[4]],[5,rvalor[5]],[6,rvalor[6]],[7,rva
lor[7]]]:
> fi:
> if tcond = 8 then:
>    desenhor :=
[[1,rvalor[1]],[2,rvalor[2]],[3,rvalor[3]],[4,rvalor[4]],[5,rvalor[5]],[6,rvalor[6]],[7,rva
lor[7]],[8,r
> valor[8]]]:
> fi:
> if tcond = 9 then:
>    desenhor :=
[[1,rvalor[1]],[2,rvalor[2]],[3,rvalor[3]],[4,rvalor[4]],[5,rvalor[5]],[6,rvalor[6]],[7,rva
lor[7]],[8,r
> valor[8]],[9,rvalor[9]]]:
> fi:
> if tcond = 10 then:
>    desenhor :=
[[1,rvalor[1]],[2,rvalor[2]],[3,rvalor[3]],[4,rvalor[4]],[5,rvalor[5]],[6,rvalor[6]],[7,rva
lor[7]],[8,r
> valor[8]],[9,rvalor[9]],[10,rvalor[10]]]:
> fi:
> if tcond = 11 then:
>    desenhor :=
[[1,rvalor[1]],[2,rvalor[2]],[3,rvalor[3]],[4,rvalor[4]],[5,rvalor[5]],[6,rvalor[6]],[7,rva
lor[7]],[8,r
> valor[8]],[9,rvalor[9]],[10,rvalor[10]],[11,rvalor[11]]]:
> fi:
> if tcond = 12 then:
>    desenhor :=
[[1,rvalor[1]],[2,rvalor[2]],[3,rvalor[3]],[4,rvalor[4]],[5,rvalor[5]],[6,rvalor[6]],[7,rva
lor[7]],[8,r
> valor[8]],[9,rvalor[9]],[10,rvalor[10]],[11,rvalor[11]],[12,rvalor[12]]]:
> fi:
> if tcond = 13 then:
>    desenhor :=
[[1,rvalor[1]],[2,rvalor[2]],[3,rvalor[3]],[4,rvalor[4]],[5,rvalor[5]],[6,rvalor[6]],[7,rva
lor[7]],[8,r
```

```
> valor[8]],[9,rvalor[9]],[10,rvalor[10]],[11,rvalor[11]],[12,rvalor[12]],[13,rvalor[13]]]:
> fi:
> if tcond = 14 then:
>    desenhor :=
[[1,rvalor[1]],[2,rvalor[2]],[3,rvalor[3]],[4,rvalor[4]],[5,rvalor[5]],[6,rvalor[6]],[7,rva
lor[7]],[8,r
>
valor[8]],[9,rvalor[9]],[10,rvalor[10]],[11,rvalor[11]],[12,rvalor[12]],[13,rvalor[13]],[14
,rvalor[14]]]:
> fi:
> if tcond = 15 then:
>    desenhor :=
[[1,rvalor[1]],[2,rvalor[2]],[3,rvalor[3]],[4,rvalor[4]],[5,rvalor[5]],[6,rvalor[6]],[7,rva
lor[7]],[8,r
>
valor[8]],[9,rvalor[9]],[10,rvalor[10]],[11,rvalor[11]],[12,rvalor[12]],[13,rvalor[13]],[14
,rvalor[14]],[15,rvalor
> [15]]]:
> fi:
> plot(desenhor,style=line,title=`Auto Valores de R`);
> #
> print(smatriz);
> print(snorm);
> print(snorm2);
> print(raizes2);
> print(xclass);
> print(xvalor);
> print(svalor);
> print(svetor);
> print(rnorm);
> print(rnorm2);
> print(raizes12);
> print(yclass);
> print(yvalor);
> print(rvalor);
> print(rvetor);
> #
> ssoma := 0:
> rsoma := 0:
> for i from 1 to tcond do
>       ssoma := ssoma + svalor[i]:
>       rsoma := rsoma + rvalor[i]:
> od:
> #
> ##############################################################################
> ###  Critérios de KAISER
> ##############################################################################
> smedia := evalf(ssoma / tcond):
> rmedia := evalf(rsoma / tcond):
> writeln(KAISER_Media_S):
> for i from 1 to tcond do:
>       if svalor[i] > smedia then:
>           writeln(VALOR,i,svalor[i]):
>           for j from 1 to tcond do:
>                 writeln(VETOR,i,j,svetor[i,j]):
>            od:
```

```
>         fi:
> od:
> #
> writeln(KAISER_Media_R):
> for i from 1 to tcond do:
>       if rvalor[i] > rmedia then:
>           writeln(VALOR,i,rvalor[i]):
>           for j from 1 to tcond do:
>                 writeln(VETOR,i,j,rvetor[i,j]):
>            od:
>         fi:
> od:
> #
> writeln(KAISER_MaiorUM_S):
> for i from 1 to tcond do:
>       if svalor[i] > 1 then:
>           writeln(VALOR,i,svalor[i]):
>           for j from 1 to tcond do:
>                 writeln(VETOR,i,j,svetor[i,j]):
>            od:
>         fi:
> od:
> #
> writeln(KAISER_MaiorUM_R):
> for i from 1 to tcond do:
>       if rvalor[i] > 1 then:
>           writeln(VALOR,i,rvalor[i]):
>           for j from 1 to tcond do:
>                 writeln(VETOR,i,j,rvetor[i,j]):
>            od:
>         fi:
> od:
> #
> ##############################################################################
> ###  Relaciona os Fatores Interpretaveis p/ S e R        (% de explicação)
> ##############################################################################
> sexpl := array(1..tcond):
> rexpl := array(1..tcond):
> sexplac := array(1..tcond):
> rexplac := array(1..tcond):
> sacum := 0:
> racum := 0:
> for i from 1 to tcond do:
>        sexpl[i] := svalor[i] * 100 / ssoma:
>        sacum := sacum + svalor[i]:
>        sexplac[i] := sacum * 100 / ssoma:
>        rexpl[i] := rvalor[i] * 100 / rsoma:
>        racum := racum + rvalor[i]:
>        rexplac[i] := racum * 100 / rsoma:
> od:
> #
> writeln(INTERPRETAVEIS_S):
> for i from 1 to tcond do writeln(i,sexpl[i])  od:
> writeln(INTERPRETAVEIS_S_ACUM):
> for i from 1 to tcond do writeln(i,sexplac[i])  od:
> writeln(INTERPRETAVEIS_R):
```

```
> for i from 1 to tcond do writeln(i,rexpl[i])  od:
> writeln(INTERPRETAVEIS_R_ACUM):
> for i from 1 to tcond do writeln(i,rexplac[i])  od:
> #
> ##############################################################################
> ###  Correlação entre Variáveis Originais e Componentes Principais  de    S  e  R
> ##############################################################################
> scorrel := array(1..tcond,1..tcond):
> rcorrel := array(1..tcond,1..tcond):
> for i from 1 to tcond do:
>        for j from 1 to tcond do:
>              sraiz := evalf(sqrt( svalor[j] )):
>              scorrel[i,j] := ( svetor[j,i] * sraiz )  / sqrt(matval[1,i,i]):
>              rraiz := evalf(sqrt( rvalor[j] )):
>              rcorrel[i,j] := ( rvetor[j,i] * rraiz ):
>        od:
> od:
> #
> print(scorrel);
> print(rcorrel);
> #
> writeln(CORRELACAO_S):
> for i from 1 to tcond do for j from 1 to tcond do writeln(i,j,scorrel[i,j])  od od:
> writeln(CORRELACAO_R):
> for i from 1 to tcond do for j from 1 to tcond do writeln(i,j,rcorrel[i,j])  od od:
> #
> close();
> ####################   >>>>>>>>>>>>>>   FIM    <<<<<<<<<<<<<<<<<<<<
```

Analise dos Componentes Principais

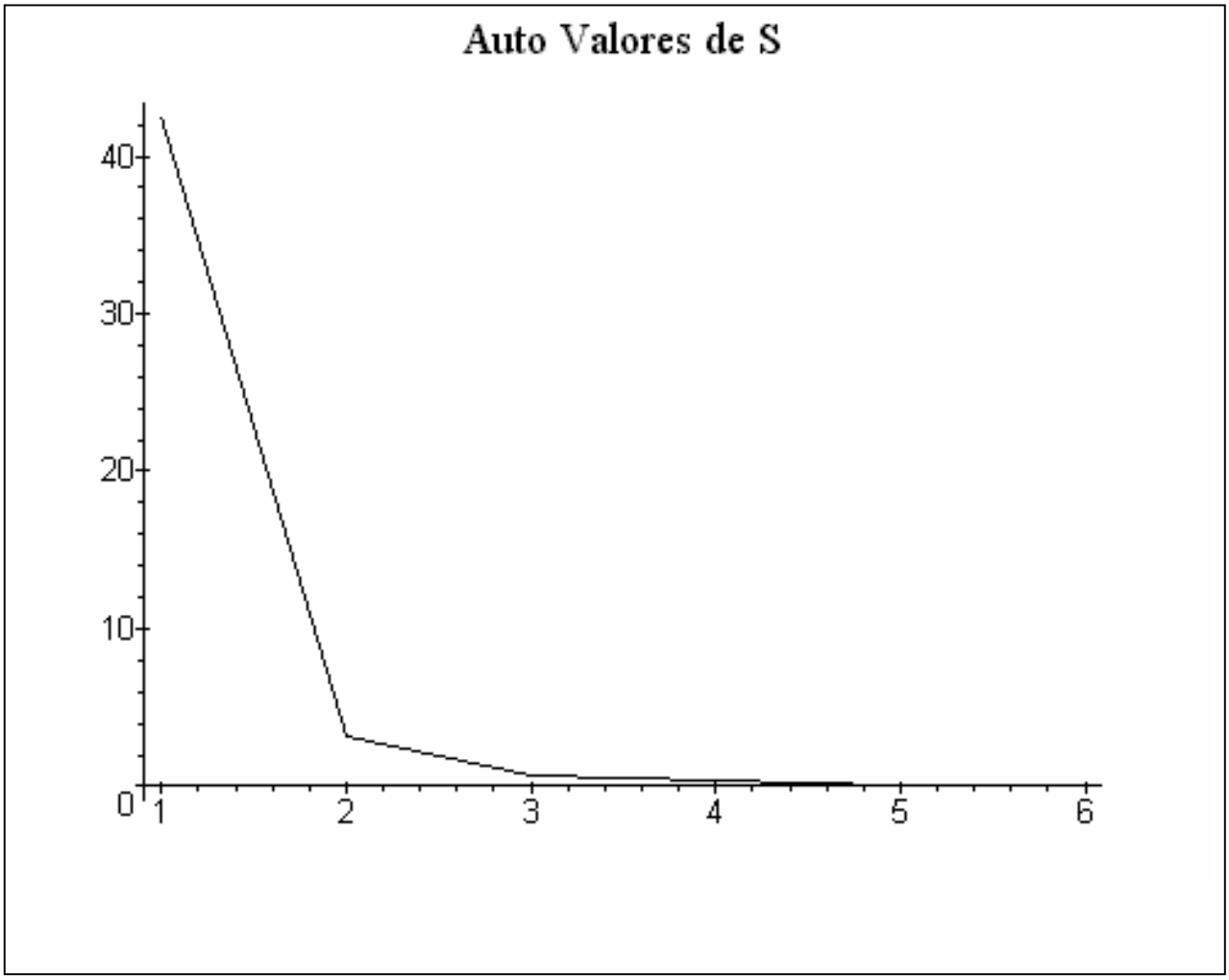
Experimento: EXEC01 - EFEITO INFEST. C/ ISOLADOS DE PYTHIUM
Numero de Variaveis: 6 18/08/05-10:04

Un Exp	ALT.AEREA-CM PESO S.RAIZ	VOL.RAIZ-CM3 PESO S.AEREA	PESO F.RAIZ	PESO F.AEREA
A	11.030000 0.830000	10.960000 0.920000	10.510000	11.020000
B	10.250000 0.950000	10.470000 0.980000	9.650000	12.230000
D	12.230000 0.740000	6.640000 1.200000	8.750000	11.450000
F	14.060000 1.120000	9.970000 1.260000	9.630000	13.490000
1G	11.830000 0.980000	8.360000 1.120000	10.020000	12.680000
1P	12.850000 0.710000	6.680000 1.040000	8.070000	13.080000
2P	12.200000 0.920000	6.400000 0.810000	9.130000	10.510000
3G	12.400000 1.060000	7.340000 1.820000	9.090000	11.580000
4P	7.200000 0.920000	4.710000 0.680000	3.330000	4.680000
5G	11.850000 1.060000	8.470000 1.740000	7.180000	8.800000
Testemunha	16.600000 1.840000	18.820000 2.060000	18.320000	20.650000

```
Analise dos Componentes Principais
-------------------------------------------------------------------------------
Experimento: EXEC01   - EFEITO INFEST. C/ ISOLADOS DE PYTHIUM
Numero de Variaveis: 6                                        18/08/05-10:04
-------------------------------------------------------------------------------
[ Vetor de Media ]
               12.045500          8.983640          9.425460         11.833600
                1.011820          1.239090

[ Matriz de Variancia e Covariancia ]
  L=1           5.343903          5.978000          6.788000          7.931000
                0.443800          0.717900
  L=2           5.978000         14.293434         12.301600         12.080000
                0.968010          0.999000
  L=3           6.788000         12.301600         12.542374         12.793000
                0.821400          0.934200
  L=4           7.931000         12.080000         12.793000         14.591112
                0.792600          0.999600
  L=5           0.443800          0.968010          0.821400          0.792600
                0.092192          0.097350
  L=6           0.717900          0.999000          0.934200          0.999600
                0.097350          0.198372


[ Matriz de Correlacao ]
  L=1           1.000000          0.684004          0.829130          0.898162
                0.632282          0.697260
  L=2           0.684004          1.000000          0.918763          0.836478
                0.843265          0.593277
  L=3           0.829130          0.918763          1.000000          0.945667
                0.763866          0.592257
  L=4           0.898162          0.836478          0.945667          1.000000
                0.683380          0.587547
  L=5           0.632282          0.843265          0.763866          0.683380
                1.000000          0.719861
  L=6           0.697260          0.593277          0.592257          0.587547
                0.719861          1.000000


[ Auto Valores de S e Auto Vetores Associados ]
  Auto Valor 1 =          42.601700
  Vetor 1       0.306627          0.542331          0.535138          0.567696
                0.036580          0.043224
  Auto Valor 2 =           3.235210
  Vetor 2       0.542763         -0.715659         -0.045132          0.434420
               -0.045395          0.020542
  Auto Valor 3 =           0.697836
  Vetor 3      -0.728753         -0.326529          0.341249          0.410080
               -0.098383         -0.260773
  Auto Valor 4 =           0.428904
  Vetor 4       0.154943         -0.268525          0.770211         -0.556336
                0.010687          0.032229
  Auto Valor 5 =           0.080343
  Vetor 5      -0.237249         -0.109500          0.043065          0.104862
                0.266932          0.920662
```

```
  Auto Valor 6 =           0.017558
  Vetor 6        0.003552         -0.054708         -0.008130          0.018036
                 0.956846         -0.284689

Analise dos Componentes Principais
-------------------------------------------------------------------------------
Experimento: EXEC01   - EFEITO INFEST. C/ ISOLADOS DE PYTHIUM
Numero de Variaveis: 6                                        18/08/05-10:04
-------------------------------------------------------------------------------
[ Auto Valores de R e Auto Vetores Associados ]
  Auto Valor 1 =           4.760560
  Vetor 1        0.407456          0.420900          0.436713          0.428186
                 0.397106          0.353650
  Auto Valor 2 =           0.575263
  Vetor 2       -0.107621         -0.152693         -0.325890         -0.377388
                 0.352065          0.769758
  Auto Valor 3 =           0.465449
  Vetor 3       -0.602053          0.475827          0.081096         -0.213574
                 0.519796         -0.297897
  Auto Valor 4 =           0.121122
  Vetor 4       -0.424083          0.412155          0.219595          0.025467
                -0.648877          0.424701
  Auto Valor 5 =           0.045083
  Vetor 5       -0.520164         -0.575709          0.278277          0.529343
                 0.164597          0.115129
  Auto Valor 6 =           0.032585
  Vetor 6       -0.097535          0.267969         -0.755539          0.589721
                 0.005500          0.006255



[ Kaiser - Auto Valores de  S  e Auto Vetores Associados ]
  Auto Valor 1 =          42.601700
  Vetor 1        0.306627          0.542331          0.535138          0.567696
                 0.036580          0.043224

[ Kaiser - Auto Valores de  R  e Auto Vetores Associados ]
  Auto Valor 1 =           4.760560
  Vetor 1        0.407456          0.420900          0.436713          0.428186
                 0.397106          0.353650



[ Fatores Interpretaveis - Matriz S ]
[ Contribuicao Individual ]
  % E(1) =        90.523366 %
  % E(2) =         6.874423 %
  % E(3) =         1.482816 %
  % E(4) =         0.911368 %
  % E(5) =         0.170718 %
  % E(6) =         0.037309 %

[ Contribuicao Acumulada ]
  % E(1) =        90.523366 %
  % E(2) =        97.397789 %
  % E(3) =        98.880605 %
```

```
  % E(4) =        99.791973 %
  % E(5) =        99.962691 %
  % E(6) =       100.000000 %

Analise dos Componentes Principais
-------------------------------------------------------------------------------
Experimento: EXEC01   - EFEITO INFEST. C/ ISOLADOS DE PYTHIUM
Numero de Variaveis: 6                                        18/08/05-10:04
-------------------------------------------------------------------------------
[ Fatores Interpretaveis - Matriz R ]
[ Contribuicao Individual ]
  % E(1) =        79.341855 %
  % E(2) =         9.587619 %
  % E(3) =         7.757404 %
  % E(4) =         2.018679 %
  % E(5) =         0.751371 %
  % E(6) =         0.543073 %

[ Contribuicao Acumulada ]
  % E(1) =        79.341855 %
  % E(2) =        88.929473 %
  % E(3) =        96.686877 %
  % E(4) =        98.705557 %
  % E(5) =        99.456927 %
  % E(6) =       100.000000 %



[ Correlacao Entre Variaveis Originais e Componentes Principais - S]
  L=1            0.865755          0.422311         -0.263347          0.043896
                -0.029090          0.000204
  L=2            0.936288         -0.340478         -0.072149         -0.046515
                -0.008210         -0.001917
  L=3            0.986255         -0.022922          0.080493          0.142429
                 0.003447         -0.000304
  L=4            0.970029          0.204558          0.089681         -0.095383
                 0.007781          0.000626
  L=5            0.786331         -0.268912         -0.270676          0.023051
                 0.249188          0.417574
  L=6            0.633437          0.082956         -0.489102          0.047390
                 0.585914         -0.084697

[ Correlacao Entre Variaveis Originais e Componentes Principais - R ]
  L=1            0.889017         -0.081626         -0.410744         -0.147592
                -0.110445         -0.017606
  L=2            0.918348         -0.115812          0.324628          0.143440
                -0.122238          0.048372
  L=3            0.952850         -0.247175          0.055327          0.076425
                 0.059086         -0.136384
  L=4            0.934246         -0.286234         -0.145708          0.008863
                 0.112394          0.106452
  L=5            0.866435          0.267027          0.354624         -0.225826
                 0.034948          0.000993
  L=6            0.771619          0.583831         -0.203236          0.147807
                 0.024445          0.001129
```


Auto Valores de S
40
30
20
10
0
1
2
3
4
5
6

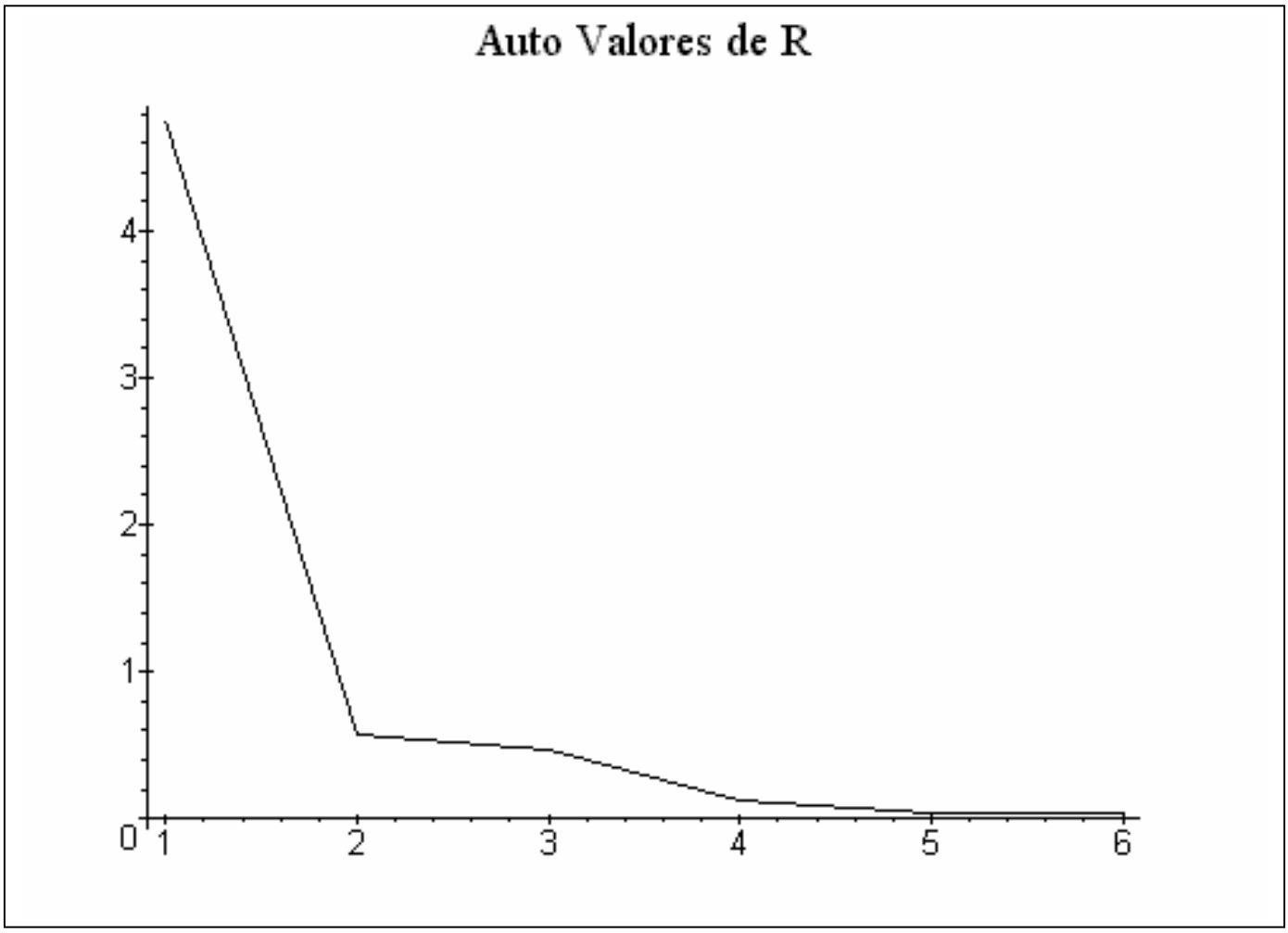

Auto Valores de R
4
3
2
1
0
1
2
3
4
5
6

Tabela 1. Valores observados nos perfis das variáveis quantificadas nos isolados de *Pythium* no cultivar de milho inoculado e não inoculado.

	Isolado										
Variável	A	B	D	F	1G	1P	2P	3G	4P	5G	TEST
	10,81	9,72	14,12	10,93	12,11	13,6	13,18	14,39	8,45	12,93	18,57
X_1	12,47	10,21	10,81	15,84	14,28	15,43	11,98	15,12	6,94	14,35	13,54
	9,83	9,87	11,23	13,26	10,32	12,94	10,43	11,28	7,02	10,09	16,48
	11,01	11,2	12,76	16,21	10,61	9,43	13,21	8,81	6,39	10,03	17,81
	13,29	10,05	7,29	10,21	7,93	7,33	7,09	8,43	3,09	9,1	21,07
X_2	9,43	11,23	8,33	8,69	9,21	9,43	5,39	9,05	5,34	10,48	19,34
	12,25	12,30	6,98	9,31	8,34	6,08	4,33	5,93	3,87	8,39	18,69
	8,87	8,30	3,96	11,67	7,96	3,88	8,79	5,95	6,54	5,91	16,18
	10,92	7,98	9,12	11,01	10,53	5,29	11,63	8,93	2,94	6,36	17,21
X_3	9,85	8,69	10,02	10,4	12,49	6,48	8,49	7,68	3,01	6,95	18,39
	10,47	10,00	7,90	9,05	9,56	9,70	10,50	9,72	4,84	8,41	18,25
	10,80	11,93	7,96	8,06	7,50	10,81	5,90	10,03	2,53	7,00	19,43
	11,00	13,40	9,18	12,03	13,25	13,75	8,90	9,89	3,99	9,30	23,42
X_4	12,45	10,23	12,01	15,14	14,38	12,54	10,07	12,04	5,38	9,00	18,90
	10,93	9,98	12,78	13,53	11,05	11,83	9,80	13,42	2,69	8,65	18,43
	9,70	15,31	11,83	13,26	12,04	14,20	13,27	10,97	6,66	8,25	21,85
	0,79	1,00	0,70	0,98	0,85	0,76	0,89	0,83	0,90	0,89	1,33
X_5	0,93	1,21	0,59	1,00	0,63	0,71	1,01	1,58	0,93	0,94	2,01
	0,80	0,83	0,83	1,01	1,51	0,73	0,85	0,99	0,85	1,38	2,23
	0,80	0,76	0,84	1,49	0,93	0,64	0,93	0,84	1,00	1,03	1,79
	0,78	1,05	1,01	1,40	1,09	0,89	0,81	1,76	0,54	1,65	2,23
X_6	0,98	1,03	1,53	0,99	1,31	0,98	0,90	1,71	0,73	1,83	1,93
	1,03	0,95	1,30	1,10	0,93	1,19	0,77	2,01	0,80	1,58	2,01
	0,89	0,89	0,96	1,55	1,15	1,10	0,76	1,80	0,65	1,90	2,07