

JOSÉ RICARDO LIBERATO

APLICAÇÕES DE TÉCNICAS DE ANÁLISE MULTIVARIADA EM
FITOPATOLOGIA

Tese Apresentada a Universidade Federal de Viçosa, como
Parte das Exigências do Curso de Fitopatologia, para a
Obtenção do Título de "Magister Scientiae".

VIÇOSA
MINAS GERAIS - BRASIL
FEVEREIRO - 1995

"A PERSISTÊNCIA É O CAMINHO DO ÊXITO"

"Cada competência que se rejeita, cada merecimento que se desdenha, cada gênio, cada talento. cada saber, que se recusa, que se desgosta, que se persegue, negando-se-lhe honras, prêmios e cargos, para se distribuírem como librés, a válidos e ociosos, a ignorantes e nulos, a comensais e parasitos, é um valor da cultura, um valor de produção. um valor de riqueza que se subtrai a fortuna do país, e de que se priva o tesouro geral da humanidade. São atos de desperdício, dilapidação e loucura, com cada qual ninguém sabe quanto vai perder a nação e o gênero humano".

Rui Barbosa

A DEUS .

A minha mãe, Olgany.

A minha esposa, Selma.

A minha filha. Marina.

AGRADECIMENTOS

À **Universidade** Federal de Viçosa, pela oportunidade de realização do curso.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (**CNPq**) e a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (**CAPES**), pela concessão da bolsa de estudos.

A minha mãe e aos meus irmãos. pelo apoio e incentivo constantes.

A minha esposa Selma, pelo apoio e incentivo constantes.

Ao professor Cláudio Horst **Bruckner**, pela orientação na iniciação a pesquisa durante o curso de graduação e pela amizade.

Ao professor **Luiz Antônio** Maffia, pela amizade, por ter me ensinado os primeiros caminhos da fitopatologia durante a iniciação científica, pelo grande incentivo a realização do curso, apoio constante, pelas críticas e sugestões.

Ao professor **Cosme Damião** Cruz, pelo apoio, incentivo, pelas críticas, sugestões e pela amizade, dedicação e boa vontade com que me ensinou as técnicas de análise **multivariada**.

Ao professor **Francisco Xavier** Ribeiro do Vale, pela amizade, orientação. pelos **ensinamentos**, pelo apoio constante e grande incentivo ao desenvolvimento do tema.

Ao professor **Laércio Zambolim**, pela amizade, orientação, pelos **ensinamentos**, pelo incentivo ao desenvolvimento do tema e apoio constante em minha vida acadêmica e profissional.

Ao professor **Geraldo Martins** Chaves, pelos **ensinamentos**, pelas críticas e sugestões.

Aos demais professores do Departamento de Fitopatologia, pelos ensinamentos e pelo apoio: **Acelino C. Alfnas**, **Eduardo Mizubuti**, **Francisco, Ferreira**, **Kyoshi Matsuoka**, **Onkar D. Dhingra**, **Murilo G. Carvalho**, **Reginaldo Romeiro**, **Sérgio H. Brommonschenkel** e **Silamar Ferraz**.

Ao professor **Adair Regazzi**, por ter me ensinado o teste de identidade de modelos de regressão.

Aos funcionários do Departamento de Fitopatologia: **Bruno, Camilo, Carminha, Cecília, Jésus, Joaquim, José Cláudio, José Orlando, Macabeu, Regina, Renato, Sérgio** e em especial a **José Carlos, Geraldo e Vicente**.

Ao amigo, colega de curso e pesquisador da EPAMIG, **Antônio A. Pereira**, pelo grande incentivo e pela ajuda prestada.

Aos amigos e colegas de curso, pelo apoio e pela agradável convivência: **Andrea, Arildo, Félix, Gil, Gláucia, Jader, José Clério, Marcelo, Marcos Cunha, Maria Amélia, Míria, Nicinha, Paulo Parizzi, Ramon, Rosalee, Silvaldo, Thor, Trazilbo, e Valdir**.

Aos **meus** sogros, **José Fernando Coelho** e **Maria do Socorro**, pelo apoio e pela ajuda na edição da **tese**.

A EMCAPA. Empresa Capixaba de Pesquisa Agropecuária, pelo apoio à **conclusão** do curso.

Aos pesquisadores da EMCAPA. amigos e companheiros de curso e de trabalho. **Hélcio Costa** e **José A. Ventura**. pelo exemplo de trabalho. enorme apoio e incentivo constantes durante o curso e no início de minha vida profissional.

Aos colegas pesquisadores da EMCAPA, pelo apoio: **Adelaide e Aureliano Costa. Danilo Milanez, João A. Molino, José A. Lani, Luciano F. Mendonça, Maria Amélia Ferrão, Romário G. Ferrão e Thales Matos**.

Aos funcionários do Setor de Fitopatologia da EMCAPA, Estação Experimental de Linhares (ES), pela ajuda constante: **Antônio Clemente, Auro C. Caon e Erlânia Cossuol**.

A todos que, direta ou indiretamente, contribuíram para a realização deste trabalho.

Aqueles professores, funcionários e colegas que, pelos **seus** exemplos de humildade e dedicação ao trabalho, dignificaram este período de **convivência** e representaram uma **lição** de vida.

BIOGRAFIA

JOSÉ RICARDO LIBERATO, filho de Dirceu Liberato e Olgany Helena Liberato, nasceu em Colatina (ES), em 14 de novembro de 1967.

Em março de 1990, graduou-se engenheiro-agrônomo pela Universidade Federal de Viçosa (UFV). Em abril de 1990, iniciou o curso de Mestrado em Fitopatologia na UFV.

Em dezembro de 1991, foi contratado pelo convênio FCAA (Fundação Ceciliano Abel de Almeida) - SEAMA (Secretaria Estadual para Assuntos do Meio Ambiente-ES) para atuar como pesquisador na EMCAPA (Empresa Capixaba de Pesquisa Agropecuária).

CONTEÚDO

	Página
EXTRATO	x
ABSTRACT	xii
1.INTRODUÇÃO	1
2.REVISÃO DE LITERATURA	3
2.1. Tipos de Variáveis	4
2.2. Análise de Componentes Principais	5
2.2.1. Aplicações da Análise de Componentes Principais em Fitopatologia	11
2.3. Análise Canônica	12
2.3.1. Aplicações da Análise Canônica em Fitopatologia	13
2.4. "Cluster Análise"	13
2.4.1. Medidas de Similaridade	15
2.4.1.1. Medidas de Distância	16
2.4.1.1.1. Distância Euclidiana	17
2.4.1.1.2. Distância de Mahalanobis	17
2.4.1.1.3. "Distância de Canberra"	19
2.4.1.2. Coeficientes de Similaridade	20
2.4.1.2.1. Coeficientes de Associação	21
2.4.1.2.2. Coeficiente de Correlação	23
2.4.1.2.3. Coeficientes de Similaridade Probabilística	24
2.4.1.3. Medidas de Similaridade Utilizadas para Dados de Eletroforese de Isoenzimas e Proteínas Solúveis e de DNA	24

2.4.2. Métodos de Agrupamento	25
2.4.3. Aplicações de "Cluster Análise" em Fitopatologia	30
2.5. Outras Técnicas de Análise Multivariada	37
2.5.1. "Path Análise" ou Análise de Trilha	37
2.5.2. Análise de Correlação Canônica	39
2.5.3. Análise Fatorial ("Factor Analysis")	41
2.5.4. Análise Discriminante	42
2.5.5. MANOVA	45
2.5.6. Análise de Coordenada Principal	46
2.6. Escolha da Técnica de Análise Multivaria	47
2.7. Influência do Tipo de Dados na Escolha da Técnica de Análise Multivariada	48
2.7.1. Variáveis Quantitativas Contínuas	46
2.7.1.1. Originadas de Delineamento Experimental	48
2.7.1.2. Não-Originadas de Delineamento Experimental	49
2.7.2. Variáveis Quantitativas Discretas	49
2.7.3. Variáveis Qualitativas Binárias	50
2.7.4. Variáveis Qualitativas Multicategóricas	51
2.7.5. Variáveis Ordinais	52
2.7.6. Mistura de Variáveis Quantitativas e Qualitativas	52
2.8. padronização e "Peso" de Variáveis	52
2.9. Descarte de Variáveis	53
3. MATERIAL E METODOS	58
3.1. Aplicação de "Cluster Análise" no Estudo do Progresso da Ferrugem (Tranzschelia sp.) em Cultivares de Pessegueiro (Prunus persica)	58
3.1.1. Material Experimental	58
3.1.2. Delineamento Experimental	58
3.1.3. Avaliação de Doença	59
3.1.4. Análise de Dados	59
3.2. Aplicação da Análise Canônica e "Cluster Análise" no Estudo da Resistência Horizontal de Genótipos de Coffea arabica a Ferrugem (Hemileia vastatrix) ...	61
3.2.1. Material Experimental	61
3.2.2. Técnica de Inoculação	61
3.2.3. Raça de H. vastatrix Utilizada	63
3.2.4. Delineamento Experimental	63
3.2.5. Avaliação dos Componentes de Resistência	63
3.2.5.1. Período de Incubação	63
3.2.5.2. Período Latente	63
3.2.5.3. Esporulação	64
3.2.5.4. Severidade	64
3.2.6. Avaliação do Grau de Doença	65
3.2.7. Análise de Dados	65
3.3. Avaliação da Resistência de Progênies de 'Catimor' a Oito Raças Fisiológicas de Hemileia vastatrix Utilizando "Cluster Análise"	66

3.4. Utilização da Análise de Componentes Principais e "Cluster Análise" no Estudo da Resistência Horizontal em Progênes de 'Catimor' a <i>Hemileia vastatrix</i>	66
3.5. Análises Multivariadas	67
3.5.1. Análise de Componentes Principais	67
3.5.2. Análise Canônica	70
3.5.3. "Cluster Análise"	73
3.5.3.1. Medidas de Similaridade	73
3.5.3.1.1. Distância Euclidiana	73
3.5.3.1.2. Distância de Mahalanobis	73
3.5.3.1.3. "Distância de Canberra"	75
3.5.3.2. Métodos de Agrupamento	75
3.5.3.2.1. Método do Vizinho Mais Próximo	75
3.5.3.2.2. Método de Tocher	76
4. RESULTADOS E DISCUSSÃO	78
4.1. Aplicação de "Cluster Análise" no Estudo do Progresso da Ferrugem (<i>Tranzschelia</i> sp.) em Cultivares de Pessegueiro (<i>Prunus persica</i>)	78
4.2. Aplicação da Análise Canônica e "Cluster Análise" no Estudo da Resistência Horizontal de Genótipos de <i>Coffea arabica</i> a Ferrugem (<i>Hemileia vastatrix</i>)	92
4.3. Avaliação da Resistência de Progênes de 'Catimor' a Oito Raças Fisiológicas de <i>Hemileia vastatrix</i> , Utilizando "Cluster Análise"	102
4.4. Utilização da Análise de Componentes Principais e "Cluster Análise" no Estudo da Resistência Horizontal em Progênes de 'Catimor' a <i>Hemileia vastatrix</i>	108
5. RESUMO E CONCLUSÕES	113
BIBLIOGRAFIA	116
APÊNDICES	128

EXTRATO

LIBERATO, José Ricardo, MS., Universidade Federal de Viçosa, fevereiro de 1995. **Aplicações de Técnicas de Análise Multivariada em Fitopatologia.** Professor Orientador: Francisco Xavier Ribeiro do Vale. Professores Conselheiros: Laércio Zambolim e Cosme Damião Cruz,

Este trabalho teve como objetivo avaliar técnicas estatísticas de análise multivariada ("cluster análise" - CA; análise de componentes principais - CP; e análise canônica - AC) no estudo da resistência do pessegueiro (*Prunus persica*) e do cafeeiro (*Coffea arabica*) as respectivas ferrugens (*Tranzschelia* sp. e *Hemileia vastatrix*).

O estudo da similaridade entre 11 cultivares de pessegueiro quanto ao progresso da ferrugem, empregando CA, apresentou concordância com os resultados de AACPD (área abaixo da curva de progresso de doença) e com os testes de identidade de modelos de progresso de doença.

No estudo da similaridade entre oito genótipos de *C. arabica* quanto aos componentes de resistência, período de incubação (PI), período latente (PL), severidade (SEV) e esporulação (ESP), obteve-se concordância entre os resultados obtidos por AC, CA e o grau de doença. PI foi o componente de resistência que menos contribuiu para a divergência entre os genótipos de café (8,73%). PL, SEV e ESP contribuíram com 34,49; 25,93 e 30.84%, respectivamente. As análises

de **variância** univariada não revelaram diferenças significativas entre **os** genótipos quanto aos componentes de resistência, ao contrário da análise **multivariada**.

No estudo da resistência de 49 progênies de 'Catimor' a oito raças fisiológicas de **H. vastatrix**, CA revelou a existência de similaridade entre as progênies e grande diferença entre estas e a testemunha, 'Catuaí'.

Estudou-se a resistência horizontal de 11 progênies de 'Catimor', considerando quatro componentes de resistência; PI, PL, SEV e NPF (número de pústulas por folha). CA e CP separaram as progênies em três grupos; e a testemunha 'Catuaí' constituiu um quarto grupo (o mais susceptível). Segundo a análise de CP, PI e NPF podem ser descartados por serem redundantes. PI e PL foram altamente correlacionados ($r = 0.95$). bem como NPF e SEV ($r = 0,99$).

Adicionalmente, descreveram-se a **importância**, os objetivos e as aplicações de outras técnicas em fitopatologia: análise de trilha, análise de correlação canônica, análise fatorial. análise discriminante, MANOVA e análise de coordenada principal

ABSTRACT

LIBERATO, José Ricardo, M.S., Universidade Federal de Viçosa, february of **1995**. Application of Multivariate Statistical Analysis in Plant Pathology. Adviser: Francisco Xavier Ribeiro do Vale. Committee members: Laércio Zambolim and Cosme Damião Cruz.

The importance, principles, applications on plant pathology and restrictions of the procedures of statistical multivariate analysis: cluster analysis (**CA**), canonical variables analysis (**CVA**) and principal components analysis (**PCA**), were described on evaluation of resistance of peach (**Prunus persica**) and coffee (**Coffea arabica**) to rusts **Tranzschelia** sp. and **Hemileia vastatrix** , respectively.

The study of similarity between 11 cultivars of peach regarding the rust progress, using CA showed agreement with the results of AUDPC (area under the disease progress curve) and with the results of **tests** of equality of disease progress models.

The similarity between eight genotypes of C. arabica were evaluated regarding to components of resistance: incubation (**IP**) and latent (**LP**) periods, severity (**SEV**) and sporulation (**SP**). The results of CA and CVA showed agreement with the reaction type, **wich** was evaluated by 1 to 6 scale. The component of resistance IP showed the minor **contribution to** divergence between genotypes of coffee (**8.73%**). LP, SEV and SP showed **34.49**, **25.93** and **30.84%**. respectively. The univariate analysis of variance did not show **significatives** differences between

genotypes, regarding the components of resistance individually, unlike of multivariate analysis of variance (MANOVA).

In the study of resistance of **49** progenies of coffee 'Catimor' to eight physiological races of *H. vastatrix*, CA showed the existence of similarity between the progenies and a difference between the progenies and the control 'Catuaí'.

In the study of horizontal resistance of 11 progenies of coffee 'Catimor', regarding to the components of resistance, IP, LP, SEV and NPL (number of pustules / leaf), CA and PCA clustered the progenies on three groups, and the control 'Catuaí' in a another group (The most susceptible). According to the PCA, IP and NPL can be **descarted**, why they were **redundants**. IP and LP were highly **correlated** ($r = 0.95$), as well as NPL and SEV ($r = 0.99$).

Additionally, the importance. **objetives** and applications of another procedures on plant pathology were related: path analysis, correlation canonical, factor analysis, discriminant analysis, MANOVA and principal coordinated analysis.

LINTRODUÇÃO

As técnicas de análise **multivariada** são métodos estatísticos apropriados para estudos em que **várias** variáveis devam ser consideradas simultaneamente. Para muitos tipos de dados biológicos, a **correlação** entre variáveis é comum. Assim, as informações providas por análises univariadas isoladas podem ser redundantes em **se** tratando de um complexo de variáveis. **As** técnicas de análise **multivariada** combinam, simultaneamente, as múltiplas informações **provenientes** de uma unidade experimental, podendo obter-se informações que **não** são possíveis com o **uso** da **estatística** univariada.

É crescente o número de exemplos de aplicações de análises **multivariadas** em fitopatologia. Porém, no Brasil, a escassez de recursos **computacionais**, os quais **são** **fundamentais** para a **utilização** destas técnicas, limitou o uso e inibiu o repasse desse conhecimento entre os pesquisadores. Atualmente, com a maior disponibilidade de recursos **computacionais**, o emprego de tais técnicas torna-se potencialmente grande e o **seu** conhecimento indispensável ao fitopatologista.

Há um interesse crescente nas técnicas **multivariadas**. Porém, para quem inicia **seu** aprendizado, as oportunidades para o uso inadequado destas técnicas **são** grandes, e na literatura há vários exemplos de **seu** uso e de interpretações **inapropriados**. Por essas razões, o presente **trabalho** teve como objetivo **ênfatizar** a importância. **os** princípios. as aplicações e as limitações das seguintes técnicas **multivariadas**: análise de componentes principais. análise **canônica** e

cluster análise: exemplificando o seu uso em dois estudos: a) avaliação da resistência de cultivares de pessegueiro [*Prunus persica* (L.) Batsch] a ferrugem (*Tranzschelia* sp.) e b) avaliação da resistência do cafeeiro (*Coffea arabica* L.) a ferrugem (*Hemileia vastatrix* Berk. e Br.).

Adicionalmente, *são* descritos a importância, os objetivos e exemplos da aplicação, em fitopatologia, de outras técnicas multivariadas: path análise" (análise de trilha), análise de correlação canônica, análise fatorial, análise discriminante. **MANOVA** e análise de coordenada principal.

2. REVISÃO DE LITERATURA

Quando, em um experimento, cada unidade experimental é avaliada sob diferentes aspectos, a utilização de técnicas de análise **multivariada** é recomendada, pois constituem-se métodos estatísticos apropriados para o estudo das relações entre variáveis que contêm **intercorrelações**. Em virtude de essas variáveis serem consideradas simultaneamente, são obtidas interpretações que não seriam possíveis com o uso da estatística univariada (JAMES e McCULLOCH, 1990).

A oportunidade de resumir, sucintamente, grandes grupos de dados, especialmente em estádios **exploratórios** de uma investigação, tem contribuído para crescente interesse em métodos **multivariados**.

Em fitopatologia, as análises **multivariadas** podem ser de grande utilidade em **epidemiologia**, na caracterização de isolados e espécies **fitopatogênicas** e no estudo de resistência de plantas **as** doenças.

Existem diferentes técnicas de análise **multivariada**, cada qual adequada a diferentes objetivos e tipos de dados.

2.1. Tipos de Variáveis

Todas as **características observáveis** de um indivíduo ou de uma unidade experimental que apresentam variabilidade **são** denominadas variáveis. Uma variável pode ser quantitativa, qualitativa ou ordinal.

As variáveis **são** denominadas quantitativas quando podem ser **mensuradas**, isto é, representadas pelo **seu** valor numérico. Podem ser contínuas ou discretas (**descontínuas**). **As** variáveis quantitativas contínuas **são** aquelas que assumem quaisquer valores entre certos limites, como temperatura, umidade relativa, peso, altura, produtividade etc. **As** variáveis quantitativas discretas assumem somente valores específicos, ou seja, **os possíveis** valores não **são** observados em uma escala contínua por causa da existência de lacunas (**descontinuidades**) entre **eles**, Na maioria das vezes **são** números inteiros, pois originam-se de contagem, por exemplo, **o número** de plantas mortas por uma doença em uma unidade experimental.

As variáveis **são** denominadas qualitativas quando podem ser distribuídas em um determinado número de categorias mutuamente exclusivas. **A mensuração** numérica não é **possível**. Uma **observação** é feita quando um indivíduo é **alocado** para uma entre duas ou mais categorias mutuamente exclusivas. **As** observações não podem ser significativamente ordenadas nem medidas, somente classificadas e então enumeradas. Podem ser binárias ou **multicategóricas**. **As** variáveis qualitativas binárias compreendem apenas duas categorias, como a presença ou ausência de uma dada **característica**; as variáveis qualitativas **são multicategóricas** quando existem mais de duas categorias mutuamente exclusivas e estas não podem ser dispostas em alguma ordem **óbvia**, por exemplo, o padrão de cor.

Os conceitos anteriormente citados foram adaptados de **SNEATH e SOKAL (1973)**, **LITTLE e HILLS (1978)**, **STEEL e TORRIE (1980)** e **VIEIRA (1988)**.

Uma variável é denominada ordinal quando pode ser distribuída em determinado número de categorias mutuamente exclusivas, mas tais categorias possuem ordenação natural. Por exemplo: grau de instrução, estágio de uma doença etc.

Para variáveis quantitativas, quando a magnitude de uma depende da magnitude da outra, **os** estatísticos referem-se a estas como variável dependente e variável independente, respectivamente.

2.2. Análise de Componentes Principais.

A técnica de componentes principais foi originalmente descrita por Karl Pearson, em 1901, para o ajustamento de mínimos quadrados de um subespaço a uma nuvem de pontos (SOUZA, 1988a) e, posteriormente, consolidada por Hotelling em 1933 e 1936, para o propósito particular de analisar estruturas de correlações (MORRISON, 1976; MARDIA et alii, 1979 e MANLY, 1986). Naquela época, os cálculos da análise eram extremamente desanimadores para mais de umas poucas variáveis, pois eram realizados manualmente. O uso da análise só foi difundido após os computadores eletrônicos tomarem-se amplamente disponíveis (MANLY, 1986).

Segundo Jeffers (1972), citado por CRUZ (1990), a técnica de componentes principais possui os seguintes propósitos: a) examinar as correlações entre caracteres estudados; b) resumir um grande conjunto de caracteres em um outro menor e de sentido biológico; c) avaliar a importância de cada caracter e promover a eliminação daqueles que contribuem pouco, em termos de variação, no grupo de indivíduos avaliados; e d) permitir o agrupamento de indivíduos com o mais alto grau de similaridade, mediante exames visuais em dispersões gráficas no espaço bi ou tridimensional.

A utilização da análise de componentes principais tem por finalidade proporcionar simplificação estrutural dos dados, de modo que a diversidade, influenciada a princípio por um conjunto p -dimensional (p = número de caracteres considerados no estudo), possa ser avaliada por um complexo bi ou tridimensional de fácil interpretação geométrica.

Esta técnica é aplicada a um conjunto de variáveis relativas a n indivíduos. O método consiste na transformação do conjunto original de variáveis em outro conjunto, os componentes principais, de dimensões equivalentes, porém com propriedades importantes, as quais serão descritas a seguir. O objetivo é resumir a informação contida no complexo de variáveis originais, eliminando as informações redundantes existentes em decorrência da correlação entre variáveis. Na realidade, a análise é uma troca de variáveis. As variáveis observadas são trocadas por novas variáveis abstratas, os componentes principais, esperando que os primeiros componentes guardem em si quase toda a informação que possuam as variáveis originais (PEREZ-GONZALES, 1992).

Cada componente principal é uma combinação linear de todas as variáveis originais. Portanto, segundo **IEZZONI e PRITTS (1991)**, a análise é *sensível* somente a relações lineares entre as variáveis. Relações de ordem superior podem não ser detectadas. **Os** componentes principais *são* independentes entre *si* e estimados com o propósito de reter, em ordem de estimação, o máximo de informação, em termos de variação **total**, contida nos dados iniciais (**CRUZ,1990**). O coeficiente de uma variável individual representa a contribuição desta variável para o componente principal. Este valor depende das outras variáveis que **são** incluídas no estudo. **Se** diferentes conjuntos de variáveis **são** incluídos, os coeficientes serão diferentes (**JAMES e McCULLOCH, 1990**). **Os** coeficientes dos componentes principais **são** também denominados elementos de autovetores ("eigenvectors") e a **variância** associada a cada componente principal é denominada autovalor ou "eigenvalue". A correlação entre a variável original padronizada e o componente principal é denominada "loading" e indica o quanto uma variável poderia substituir o componente principal (**JAMES e McCULLOCH, 1990**).

Os princípios básicos desta técnica **são** descritos por **MORRISON (1976)**. **MARDIA et alii (1979)**, **DUNTEMAN (1984)**, **MANLY (1986)** e **JOHNSON e WICHERN (1988)**, dentre outros. **Esses** autores demonstram que cada componente principal retém uma percentagem da **variância** original, e as **variâncias são** decrescentes do primeiro ao último componente principal, ou seja, o primeiro componente possui a maior **variância**, o segundo possui a segunda maior e assim sucessivamente. A importância de um componente **se** avalia por meio da percentagem de **variância** que o mesmo retém. Assim, o primeiro componente é o mais importante, pois retém a maior parte da variação total encontrada nos dados originais. A independência ou **ortogonalidade** (ausência de correlação) entre os componentes é propriedade útil porque significa que **esses** componentes **estão** medindo diferentes "**dimensões**" dos dados (**MARDIA et alii, 1979 e MANLY, 1986**). A consequência é **que** o somatório das **variâncias** dos componentes principais (autovalores) corresponde a soma das **variâncias** das variáveis originais (**MORRISON, 1976**). Não há, portanto, perda de informação na transformação. Assim, **se** um dos componentes principais possui grande **variância** em relação a **variância** total, então aquele componente principal captura a maioria da informação existente no grupo de variáveis originais.

Na maioria das situações, o número de componentes principais **necessário** para extrair completamente a informação contida em um grupo de **variáveis**, é igual ao número de variáveis

sob consideração. Isso é verdadeiro se a matriz de **covariância** (ou correlação) correspondente as variáveis for não-singular (isto é, sua inversa existe). Se o rank, número de autovalores **não-nulos**, da matriz de **covariância** (ou correlação) for menor que a dimensão da matriz de **covariância** (isto é, o número de variáveis), então o número de componentes principais é igual ao rank da matriz de **covariância** (DUNTEMAN, 1984). Mas, os primeiros componentes poderão conter a maioria da variação das variáveis originais, reduzindo, assim, a **dimensionalidade** do grupo de dados. Esta simplificação proporciona melhor **compreensão** da natureza das variáveis originais.

Uma questão é **quantos** componentes reter na análise. Existe um teste estatístico para decidir **quantos** componentes principais devem ser retidos, o qual é relatado por MARDIA et alii (1979) e MANLY (1986). Porém, o problema de **inferência estatística** para componentes principais extraídos de uma matriz de correlação **amostral** é extremamente complexo, segundo MARDIA et alii (1979) e DUNTEMAN (1984). Esses autores relatam que, na maioria dos casos, a seleção sobre **quantos** componentes reter é baseada em métodos simples e práticos: reter componentes que expliquem, pode-se dizer, 80% ou mais da variação total; **plotar os autovalores** (λ_i versus i) e determinar onde grandes autovalores cessam e pequenos autovalores iniciam: e excluir componentes cujo autovalor seja menor que a média. Para uma matriz de correlação, autovalores **menores** que a unidade devem ser descartados. Isso significa que um componente principal deverá ser descartado se possuir menor **variância** que uma das variáveis originais padronizadas cujas **variâncias são iguais a unidade**.

Um dos objetivos da análise consiste em resumir o complexo de informações das variáveis originais nos dois primeiros componentes principais. Quando estes dois componentes principais acumulam percentagem relativamente alta da variação total, geralmente referida como acima de 80%, considera-se a análise **satisfatória**. Como os componentes principais são ortogonais entre si, representando contribuições independentes para a **variância** (ADAMS e WIERSMA, 1978), os indivíduos avaliados podem ser representados em um gráfico **cartesiano bidimensional**, sendo os eixos constituídos pelos dois primeiros componentes principais. Neste gráfico, a similaridade entre os indivíduos será avaliada de forma subjetiva, entretanto, o fenômeno poderá ser interpretado com considerável simplificação. Em algumas **situações**, quando os dois primeiros componentes não envolvem percentagem de **variância** considerada **satisfatória**, a análise é **complementada** pela dispersão gráfica em relação ao terceiro e quarto componentes,

como foi realizado por SMITH (1984) e SMITH et alii (1985). Alguns autores utilizam gráficos constituídos pelos três primeiros componentes para ilustrar em um espaço tridimensional a diversidade entre indivíduos, por exemplo SNEATH e SOKAL (1973). AKORODA (1983), SILVEIRA NETO (1986), HILLIG e IEZZONI (1988) e PEREZ-GONZALES (1992).

A viabilidade de utilização dos componentes principais dependerá da possibilidade de resumir o conjunto de variáveis originais em apenas poucos componentes, o que significará ter uma aproximação do comportamento dos indivíduos oriundos de um espaço multidimensional em um espaço bi ou tridimensional (CRUZ, 1990). A variância, antes dispersa entre as variáveis correlacionadas, é agora atribuída, em sua quase totalidade, a um número muito menor de componentes (ADAMS e WIERSMA, 1978). Desse modo, a análise proporcionará simplificação considerável nos cálculos estatísticos e na interpretação dos resultados, principalmente quando o número de indivíduos avaliados for relativamente grande (CRUZ, 1990).

Os melhores resultados são obtidos quando as variáveis originais são altamente correlacionadas, positiva ou negativamente. Segundo MANLY (1986), nesse caso, então, é completamente concebível que 20 ou 30 variáveis originais possam ser adequadamente representadas por dois ou três componentes principais. Se um ou poucos componentes principais explicam a maior parte da variação em um grande grupo de variáveis, então tem-se, efetivamente, redução da dimensionalidade do grupo de dados.

É fato questionável, na análise de componentes principais, a avaliação da divergência em análise gráfica e o estabelecimento de grupos de similaridade de maneira subjetiva, com base na simples inspeção visual da dispersão (CRUZ, 1990). Há casos em que a visualização do grau de similaridade entre indivíduos estudados não é muito clara. Em outros casos, a análise não consegue resumir o complexo de informações das variáveis originais, ou seja, os dois ou três primeiros componentes contêm uma percentagem da variância total insuficiente para uma avaliação segura da similaridade (JEGER et alii, 1983; AKORODA, 1983; SMITH, 1984; SMITH et alii, 1985; MURPHY et alii, 1986; ZHU et alii, 1988; HILL et alii, 1989; PEREIRA, 1989; ANDERSON et alii, 1990 e CHEN et alii, 1991). Nessas situações, algumas vezes é utilizado o recurso de “cluster análise” conjugada a análise de componentes principais. Com os indivíduos simulando pontos em um espaço multidimensional, como determinado por seus escores em relação aos componentes principais, a distância geométrica entre quaisquer dois indivíduos pode

ser calculada. Esta distância é denominada euclidiana e representa a similaridade ou dissimilaridade entre um par de indivíduos. ADAMS e WIERSMA (1978) apresentaram discussão sobre a utilização conjugada dessas duas técnicas em estudos sobre divergência genética. Por sua vez, certos autores, como MIRANDA et alii (1988), têm utilizado apenas os escores dos dois primeiros componentes para o cálculo da distância euclidiana, valendo-se, para esse fim, da propriedade de independência entre tais componentes. Tal procedimento é muitas vezes utilizado para complementar as informações da dispersão gráfica, em virtude de permitir o estabelecimento de grupos de maneira menos subjetiva da que se verifica em exames visuais (CRUZ, 1990).

A técnica de componentes principais também se caracteriza por trabalhar com a média amostral ou ser usada nas situações em que não há repetições de dados. Os componentes principais dependem somente da matriz de variância e covariância entre variáveis, podendo ser muito sensíveis a observações atípicas (estranhas) ou indicar associação quando, de fato, pouca existe. Observações atípicas podem ter considerável efeito nas correlações e, conseqüentemente, nos componentes principais (JOHNSON e WICHERN, 1988). Em virtude dos coeficientes dos componentes principais serem altamente sujeitos a variabilidade amostral não se deveria dar muita ênfase aos exatos valores dos coeficientes (JAMES e McCULLOCH, 1990).

A análise de componentes principais é influenciada pela escala das variáveis, sendo os coeficientes dos componentes principais altamente dependentes das unidades de medida associadas com cada variável. Esse aspecto é discutido por MARDIA et alii (1979) e DUNTEMAN (1984). É aconselhável que os caracteres estudados possuam a mesma unidade e dimensões não muito discrepantes (ou seja, variância com semelhantes ordens de magnitude). Como na maioria das vezes isso não ocorre, recomenda-se utilizar as variáveis reduzidas, ou seja, variáveis padronizadas com média igual a zero e variância igual a unidade (ou variáveis padronizadas apenas com a variância igual a unidade). Isto significa que a matriz de covariância é transformada em uma matriz de correlações antes que a análise de componentes principais seja conduzida. Caso contrário, a variável com a maior variância possuiria maior peso na estimação dos componentes, sendo este peso mais em função das unidades de medida do que da estrutura básica das observações. Assim, as combinações lineares das quantidades originais poderiam ter pouco significado e as variáveis padronizadas e matriz de correlações deveriam ser utilizadas (MORRISON, 1976). A padronização das variáveis assegura que todas elas possuam,

inicialmente, igual peso na análise (MANLY, 1986). Segundo IEZZONI e PRITTS (1991), se os caracteres avaliados têm diferentes dimensões, então, geralmente os dados não possuem distribuição multivariada normal. Assim, a extrapolação de resultados para uma população maior não deverá ser realizada. Uma solução é a padronização das variáveis. Entretanto, segundo DUNTEMAN (1984) e MORRISON (1976) a distribuição multivariada normal não é necessária para que a análise de componentes principais seja aplicada.

Embora a análise, formalmente, não requeira a distribuição normal multivariada, ela é mais apropriada para variáveis quantitativas contínuas. Quando a distribuição de razões e proporções é razoavelmente próxima a normal, a análise pode ser útil, mas pode não capturar relações não-lineares. Para dados constituídos de contagens, razões, proporções ou percentagens, pode-se realizar transformações para tornar sua distribuição mais apropriada, previamente a análise de componentes principais, ou então utilizar um método não-linear (JAMES e McCULLOCH, 1990). STAUFFER et alii (1985) sujeitaram os dados de percentagem a uma transformação arco seno da raiz quadrada e os dados de contagem a uma transformação de raiz quadrada (PIMENTEL GOMES, 1984), antes de realizarem a análise de componentes principais.

CAMPBELL e ATCHLEY (1981) e SOUZA (1988a) apresentam essa técnica sob uma visão geométrica, assim como MARDIA et alii (1979) e DUNTEMAN (1984). Mensurações quantitativas de um número p de características podem ser representadas pelas posições que elas ocupam em um espaço p -dimensional. Desse modo, um diagrama de dispersão é composto de eixos ortogonais individuais para cada característica, no qual os valores padronizados das características para cada observação são plotados. Se algumas das características são correlacionadas, então o grupo de pontos se assemelhará a uma nuvem elipsoidal p -dimensional, pois se os caracteres não são independentes entre si (correlação não-nula), os eixos tomam-se oblíquos. Pela análise de componentes principais, o principal eixo (primeiro componente) e subsequentes eixos ortogonais desta elipsóide são encontrados. O primeiro componente principal conterá a máxima variância, entre todos os valores das características, que pode ser atribuída a um único eixo. Cada componente principal seguinte conterá uma percentagem progressivamente menor da variância remanescente (HILLIG e IEZZONI, 1988). Geometricamente, os componentes principais representam a seleção de novo sistema de coordenadas, obtido pela rotação do sistema original com as variáveis originais como eixos coordenados. Estes novos eixos representam as

direções com a máxima variabilidade e provêm descrição mais simples e mais **parcimoniosa** da estrutura de covariância. Os autovetores determinam a direção de máxima variabilidade e os autovalores especificam as variâncias (JOHNSON e WICHERN, 1988).

Componentes matematicamente ortogonais (independentes) não necessariamente representam padrões independentes na natureza. Assim, interpretações biológicas devem ser feitas com cuidado (JAMES e McCULLOCH, 1990).

Para a análise, os dados de todas as variáveis observadas devem originar do mesmo indivíduo. TAI e YOUNG (1989), avaliando 29 genótipos de batata (*Solanum tuberosum*) pela análise de componentes principais, consideraram como nove variáveis as médias de produção dos genótipos em cada um de nove experimentos distintos. O uso da análise é inadequado para esse tipo de dado.

2.2.1. Aplicações da Análise de Componentes Principais em Fitopatologia

A utilização da técnica de componentes principais é recente em fitopatologia, havendo poucos exemplos de sua aplicação.

A técnica é utilizada em estudos de resistência de plantas a doenças, por causa da possibilidade de caracterizar genótipos considerando, simultaneamente, todos os componentes de resistência, além de avaliar a importância dos mesmos.

JEGER et alii (1983) utilizaram essa técnica para caracterizar 41 genótipos de trigo quanto à resistência parcial a *Septoria nodorum*, avaliando 11 componentes de resistência.

LALANCETTE et alii (1987) utilizaram componentes principais para determinar a similaridade entre isolados de *Venturia inaequalis* resistentes e sensíveis a benomil, quanto a quatro componentes de resistência: período de incubação, colonização, esporulação e frequência de infecção. Isolados resistentes e sensíveis a benomil tenderam a formar grupos distintos.

ANDERSON et alii (1990), utilizando a análise de componentes principais e "cluster análise", caracterizaram genótipos de amendoim quanto a resistência a *Cercospora arachidicola*, *Cercosporidium personatum* e *Puccinia arachidis*.

Em estudos de variabilidade genética de microrganismos, baseados em eletroforese de proteínas e isoenzimas, a análise de componentes principais pode ser mais precisa que a

observação visual na comparação de isolados, possibilitando que a relação entre estes seja mais facilmente compreendida.

ZHU et alii (1988). pelas análises de componentes principais e "cluster análise" para dados de isoenzimas, demonstraram a existência de variabilidade genética **intraespecífica** do fungo ectomicorrízico **Suillus tomentosus** em regiões florestais no Canadá. A **variação** foi maior entre do que dentro de regiões florestais.

2.3. Análise Canônica

Este procedimento foi descrito por R.A. Fisher em 1936 e, posteriormente, desenvolvido por M.S. Bartlett, P.C. Mahalanobis e C.R. Rao, citados por CAMPBELL e ATCHLEY (1981). para examinar alguns problemas **significantes** relativos a sistemática **biológica**. É também conhecida como análise de variáveis **canônicas**.

As variáveis **canônicas** são combinações lineares das variáveis originais, sendo determinadas **de** tal modo que a variação entre grupos é **maximizada** em relação a variação dentro de grupos. Espera-se que a configuração do grupo possa ser adequadamente representada em um **subespaço** bi ou tridimensional (a semelhança da análise de componentes principais) pelos primeiros dois ou três vetores **canônicos**. O primeiro vetor **canônico** é fornecido pelos coeficientes da combinação linear que **maximiza** a razão da **variância** entre e dentro **de** grupos. A correspondente razão é referida como raiz **canônica**. Sucessivas combinações lineares das variáveis originais **são** escolhidas para serem não-correlacionadas tanto dentro quanto entre grupos (CAMPBELL e ATCHLEY, 1981).

Segundo CAMUSSI et alii (1985), as transformações para variáveis **canônicas** permitem a **visualização** ótima de diferenças entre populações, pela redução de dimensões que preserve a maioria das **informações** biológicas.

É um método de ordenação, cujo objetivo é avaliar o grau de similaridade entre materiais experimentais, considerando tanto a matriz de **variâncias** e **covariâncias** residuais quanto a matriz de **variâncias** e **covariâncias** entre médias fenotípicas dos caracteres avaliados. Portanto, a análise é empregada em dados provenientes de delineamentos experimentais. **Esse** método, diferentemente da análise de componentes principais, considera as possíveis diferenças

na dispersão sobre as médias. O método apresenta a vantagem adicional de manter o princípio de "cluster análise", utilizando a distância de Mahalanobis, qual seja o de considerar as correlações residuais existentes entre as médias dos tratamentos. Essa técnica possui estreita relação com a análise de função discriminante linear e com a distância de Mahalanobis.

Assim como a distância de Mahalanobis, a análise canônica assume que as correlações entre quaisquer dois caracteres sejam aproximadamente a mesma em cada tratamento. Embora não haja critérios para testar a igualdade de matrizes de correlação, a hipótese de igualdade entre as matrizes de covariância de cada tratamento pode ser testada com relativa facilidade e acuracidade, se as amostras possuírem distribuição multivariada normal. SEAL (1964), MORRISON (1976) e ARNOLD (1981) descrevem esses testes.

As variáveis podem possuir diferentes escalas, pois estas não afetam os resultados. É comum a transformação das variáveis originais em variáveis padronizadas e não-correlacionadas, de modo que a matriz de dispersão residual se iguale a matriz identidade. Para essa transformação, tem sido utilizado o processo de condensação pivotal descrito por RAO (1952) e exemplificado por SINGH e CHAUDHARY (1979). Após a transformação, o processo de estimação das variáveis canônicas equivale ao descrito para os componentes principais.

2.3.1. Aplicações da Análise Canônica em Fitopatologia

HANSEN e HAMM (1983) utilizaram a análise canônica, análise discriminante e "cluster análise" na avaliação de similaridade baseada em características morfológicas e de crescimento de 54 isolados de *Phytophthora megasperma*, de 14 hospedeiros. O objetivo era determinar a extensão na qual diferenças morfológicas acompanham especialização na patogenicidade. Grupos de isolados definidos inicialmente por patogenicidade apresentaram similaridade com relação a características morfológicas e de crescimento.

2.4. "Cluster Analise"

"Cluster análise" é também denominada de análise de agrupamento ou análise de conglomeração. Decidiu-se denominar esta análise aqui, de "cluster análise" e não de análise de

agrupamento, que seria a sua tradução na língua portuguesa, para que não haja confusão entre análise de agrupamento ("cluster análise") e método de agrupamento, que é uma etapa de "cluster análise". Porém, é necessária precaução com essa denominação. pois existem na literatura técnicas de análise univariada também denominadas "cluster análise" (SCOTT e KNOTT, 1974 e GATES e BILBRO, 1978). Como exemplo da aplicação de tais técnicas citam-se MADDEN et alii (1982), BAGNALL e TAI (1986) e SEALY et alii (1988).

"Cluster análise" tem por finalidade reunir, por algum critério de classificação, os indivíduos (ou objetos, locais, populações, amostras etc.) em vários grupos, de tal forma que exista homogeneidade dentro e heterogeneidade entre grupos (SNEATH e SOKAL, 1973 e MARDIA et alii, 1979). A análise estuda as semelhanças ou diferenças entre indivíduos independentemente de suas causas, baseando-se em suas características. São produzidos grupos, existindo ou não grupos naturais (JAMES e McCULLOCH, 1990). A análise pode gerar grupos não esperados, sugerindo relações a serem investigadas.

A análise normalmente é utilizada na seguinte situação: dado um conjunto de n indivíduos ou unidades amostrais, avaliado em relação a p variáveis. determina-se um esquema para agrupar os indivíduos, segundo algum critério de similaridade ou dissimilaridade, de modo que o conjunto inicial de observações seja dividido em vários grupos homogêneos (MANLY, 1986 e CRUZ, 1990). Ao contrário da análise discriminante (ver seção 2.5.4.), não são feitas considerações prévias sobre o número de grupos e a introdução de um novo indivíduo pode alterar os grupos formados.

"Cluster análise" envolve basicamente duas etapas. A primeira relaciona-se com a estimação de uma medida de similaridade (ou dissimilaridade) entre os indivíduos ou populações a serem agrupados e a segunda, com a adoção de um método de agrupamento para formação dos conglomerados (CRUZ, 1990). O resultado depende da medida de similaridade e do método de agrupamento escolhidos. A análise é essencialmente descritiva, não sendo utilizado teste estatístico de hipóteses.

Há basicamente dois tipos de objetivos no uso de "cluster análise": verificar o relacionamento entre indivíduos (técnica Q) e, ou, verificar o relacionamento entre caracteres (técnica R) (SNEATH e SOKAL, 1973 e SILVEIRA NETO, 1986).

"Cluster análise" é uma técnica, potencialmente, muito útil, mas requer cuidados na sua utilização, caso contrário não serão obtidos resultados válidos. Normalmente surgem várias questões sobre a análise, como o tipo de medida de similaridade e o método de agrupamento a serem utilizados, a adequação da partição obtida e o número final de grupos desejáveis. Estas questões serão abordadas adiante.

2.4.1. Medidas de Similaridade

O primeiro estágio de "cluster análise" é a conversão da matriz ($n \times p$) de dados em matriz quadrada (da ordem $n \times n$, sendo n o número de indivíduos) de similaridade ou dissimilaridade individuais, que são medidas da relação entre pares de indivíduos, dado o valor de um conjunto de p variáveis. Em cada interseção da i -ésima fila e da k -ésima coluna desta matriz, coloca-se a medida de similaridade (ou dissimilaridade) entre o i -ésimo e o k -ésimo indivíduo. Com quase todas as medidas de similaridade a matriz é simétrica. Assim sendo, somente a sua metade inferior ou superior é necessária para representá-la. A alta similaridade indica que os dois indivíduos são comuns em relação ao conjunto de variáveis, enquanto a alta dissimilaridade indica o contrário (MARDIA et alii, 1979).

Ao se obter a matriz de similaridade (ou dissimilaridade), perde-se informação. Deve-se eleger a medida de similaridade que minimize essa perda. Além disso, os agrupamentos baseados em diferentes medidas de similaridade podem levar a diferentes resultados. Mas qual medida de similaridade (ou dissimilaridade) escolher? Um grande número destas medidas tem sido proposto e utilizado em análise multivariada (MARDIA et alii, 1979). Embora, segundo SNEATH e SOKAL (1973), vários estudos comparativos tenham sido feitos, ainda não há solução para esta questão. A escolha da medida é freqüentemente baseada no tipo de dados disponíveis (ver seção 2.7.). Quando o tipo de dados é tal que algumas possíveis medidas de similaridade possam ser utilizadas, a escolha entre elas é sempre baseada na preferência do pesquisador (SNEATH e SOKAL, 1973).

Segundo SILVEIRA NETO (1986), é importante alertar que para qualquer medida de similaridade os valores só são comparáveis dentro do mesmo estudo, não tendo validade as

comparações de resultados de semelhança entre indivíduos ou amostras que não estiverem envolvidos na sua determinação.

SNEATH e SOKAL (1973) discutem a estimação da similaridade quando não há dados de algumas variáveis relativos a alguns indivíduos ou quando a comparação de algum carácter entre dois indivíduos é inaplicável.

Quando a magnitude da medida reflete a magnitude da dissimilaridade, ela é denominada medida de distância, e quando reflete a magnitude da similaridade, ela é denominada coeficiente de similaridade.

2.4.1.1. Medidas de Distância

O princípio das medidas de distancia considera que se dois indivíduos são similares eles estão próximos um do outro, ou seja, eles *são* comuns em relação ao conjunto de variáveis e vice-versa.

Segundo MARDIA *et alii* (1979) e JOHNSON e WICHERN (1988). qualquer medida de distância, $d(P,Q)$, entre dois pontos P e Q é válida, contanto que satisfaça as seguintes propriedades, sendo R um ponto intermediário:

$$d(P,Q) = d(Q,P) ;$$

$$d(P,Q) > 0 \text{ se } P \neq Q ;$$

$$d(P,Q) = 0 \text{ se e somente se } P = Q ;$$

$$d(P,Q) \leq d(P,R) + d(R,Q) \text{ (desigualdade triangular).}$$

Espera-se que $d(P,Q)$ aumente quando a dissimilaridade entre P e Q aumentar. Assim $d(P,Q)$ é também descrito como um coeficiente de dissimilaridade. Sempre que possível, é recomendável utilizar distancias verdadeiras, isto é, distâncias que satisfaçam as propriedades relatadas acima.

As medidas de distancia mais amplamente utilizadas *são* a distancia euclidiana e a distância de Mahalanobis.

2.4.1.1.1. Distância Euclidiana

Considerando o caso mais simples, em que existem n indivíduos, cada um dos quais possuindo valores para p variáveis, a distância euclidiana entre eles é obtida por analogia ao teorema de Pitágoras (MANLY, 1986), para um espaço multidimensional.

A distância euclidiana, quando estimada a partir das variáveis originais, apresenta a inconveniência de ser influenciada pela escala (MANLY, 1986), pelo número de variáveis e pela correlação existente entre as mesmas. Para se contornar a influência do número de variáveis sobre essa estimativa, recomenda-se a utilização da distância euclidiana média, dividindo-a pela raiz quadrada do número de variáveis. A influência das escalas é contornada pela padronização prévia das variáveis em estudos, para que possuam média igual a zero e variância igual a unidade. O maior inconveniente é a pressuposição da independência entre variáveis (CRUZ, 1990).

Esta distância é mais freqüentemente utilizada para variáveis quantitativas, sendo também utilizada quando as unidades para o cálculo são escores de componentes principais ou de variáveis canônicas.

2.4.1.1.2. Distância de Mahalanobis

Segundo ARUNACHALAM (1981), em estudos sobre divergência genética, é desaconselhável a quantificação da divergência pela distância euclidiana quando forem avaliados vários caracteres com graus significativos de correlação residual.

Na avaliação de um grupo de genótipos com relação a um conjunto de características correlacionadas, a distância entre qualquer par de genótipos deve considerar o grau de dependência entre as variáveis (CRUZ, 1987). Várias medidas têm sido propostas para a quantificação das distâncias entre duas populações, quando são disponíveis informações sobre média, variância e covariância residual das populações (MANLY, 1986), ou seja, quando existem repetições de dados. Uma das medidas mais utilizadas é a estatística D^2 (distância de Mahalanobis), proposta por Mahalanobis em 1936 (CRUZ, 1990), a qual considera a variabilidade de cada unidade amostral e não somente a medida de tendência amostral, sendo recomendada

para dados provenientes de delineamentos experimentais, e, principalmente, quando os caracteres são correlacionados. D^2 considera a correlação entre as características analisadas por meio da matriz de dispersão das variáveis. Quando as correlações entre as variáveis são nulas, considerando-se as variáveis padronizadas, a distância de Mahalanobis é equivalente a distância euclidiana

Em algumas situações, avalia-se a similaridade a partir de várias características, com base em delineamentos experimentais apropriados, envolvendo casualizações e repetições. Nesses casos, a utilização da matriz de dispersão residual comum a todos os tratamentos na estimação de D^2 é mais apropriada, uma vez que os quadrados médios residuais têm distribuição multivariada normal (GOODMAN, 1968 e ARUNACHALAM, 1981). A utilização desta matriz também é mais apropriada na análise canônica para a transformação das variáveis originais X em variáveis não-correlacionadas Z (ARUNACHALAM, 1981). portanto, D^2 e a análise canônica são baseadas em caracteres quantitativos contínuos, desde que somente estes caracteres podem, a priori, seguir distribuição multivariada normal. É necessário cuidado antes de incluir variáveis quantitativas discretas. Em alguns casos, uma apropriada transformação pode restaurar a distribuição normal, mas tais casos, segundo ARUNACHALAM (1981). deveriam cuidadosamente ser escrutinados antes da inclusão.

Porém, segundo Hotelling (1954), citado por GOODMAN (1967 e 1968), D^2 é equivalente ao teste entre dois vetores de médias de tratamentos em uma análise multivariada de variância de experimento baseado em delineamento inteiramente ao acaso. Assim, assume-se que a distribuição multinormal e a homogeneidade da matriz de covariância são o equivalente multivariado para a suposição de normalidade e homogeneidade de variância, no caso dos conhecidos testes t e F univariados. A não-normalidade geralmente possui pequeno efeito no teste F e a não-homogeneidade de variância geralmente também possui pequeno efeito, a menos que as diferenças sejam muito grandes ou o número de observações por tratamento diferirem grandemente. A multinormalidade permite a realização de certos testes de significância, mas não é necessária para a estimação de distâncias. O mais sério problema é a diferença entre as matrizes de covariância, sendo este assunto discutido por GOODMAN (1967). Segundo SEAL (1964). a hipótese de igualdade de matrizes de covariância pode ser testada com relativa facilidade e acuracidade se todas as amostras possuírem distribuição multivariada normal. SEAL

(1964). MORRISON (1976) e ARNOLD (1981) relatam testes de igualdade entre matrizes de covariância.

Raramente os parâmetros populacionais são conhecidos, sendo usual então estimar D^2 pelos valores amostrais correspondentes. O ideal para a estimativa de D^2 é que as covariâncias sejam acuradamente conhecidas. Quando as covariâncias podem somente ser estimadas com um pequeno número de graus de liberdade, provavelmente seja mais adequado utilizar uma distância que não considera a covariância ou correlação entre variáveis (MANLY, 1986). Porém, é difícil dizer precisamente o que é pequeno número de graus de liberdade nesse contexto (MANLY, 1986).

Segundo RAO (1952) e MANLY (1986). D^2 somente pode ser estimada se as matrizes de covariância das unidades amostrais forem homogêneas e se existir distribuição normal multidimensional, o que restringiria, portanto, o seu uso. No entanto, segundo SNEATH e SOKAL (1973), já foi demonstrada considerável robustez para a violação dessas hipóteses, o que faz da distância de Mahalanobis um instrumento muito útil.

Segundo SINGH e CHAUDHARY (1979) e MANLY (1986). os valores de D^2 obtidos para cada par de indivíduos têm distribuição qui-quadrado associados a p graus de liberdade, sendo p o número de caracteres avaliados. Assim, a significância de cada D^2 pode ser avaliada comparando a estimativa com o valor tabelado do qui-quadrado, no nível considerado.

CRUZ (1987) ilustra a obtenção da distância de Mahalanobis para dados obtidos por SINGH e CHAUDHARY (1979) referente a experimento envolvendo oito variedades de cevada avaliadas em delineamento em blocos ao acaso com quatro repetições em relação a quatro características.

Segundo SNEATH e SOKAL (1973). existem, além da distância de Mahalanobis, alguns coeficientes de similaridade que computam as variações dentro das amostras tanto quanto diferenças entre médias.

2.4.1.1.3. "Distância de Canberra"

THOMPSON e REES (1979) relatam a utilização da "distância de Canberra" (denominada pelos autores de "Canberra metric") como medida de similaridade em "cluster

análise", para avaliar a resistência de cultivares segundo as suas curvas de progresso de doença, em complemento aos procedimentos convencionais (MADDEN, 1986 e CAMPBELL e MADDEN, 1990). Esta medida de distância foi proposta por LANCE e WILLIAMS (1967).

Para a sua aplicação não é necessário assumir um modelo definido para a curva de progresso de doença. Esta medida depende apenas de um par de indivíduos ou grupos sendo comparados, não sendo afetada pelos valores dos outros indivíduos ou tratamentos. A medida é mais sensível a diferenças proporcionais que absolutas, isto é, a "distância de Canberra" entre 1 e 10% de doença é maior que entre 51 e 60% de doença. Segundo THOMPSON e REES (1979), a "distância de Canberra" parece ser uma transformação sensata, biologicamente, em que diferenças na severidade de doenças entre cultivares no início de uma epidemia são enfatizadas mais que similares diferenças absolutas no decorrer da epidemia. Convém acrescentar que medidas de similaridade são um tópico em que certa quantidade de arbitrariedade parece inevitável (MANLY, 1986).

Outras medidas de distâncias são discutidas por SNEATH e SOKAL (1973).

2.4.1.2. Coeficientes de Similaridade

Muitos dos coeficientes de similaridade foram desenvolvidos para a aplicação em taxonomia. Taxonomia numérica é o **agrupamento** por métodos numéricos de unidades taxonômicas, baseado em suas características, sendo utilizado na avaliação de relações taxonômicas e na construção de "taxa", na avaliação da evolução e análise filogenética. Os métodos requerem a conversão de informações sobre entidades taxonômicas em valores numéricos, tendo como principal vantagem a repetibilidade e objetividade.

O intenso desenvolvimento da taxonomia numérica desde meados da década de 1950 resultou em rápida elaboração de coeficientes de similaridade, freqüentemente com uma base empírica, sem justificação teórica adequada. SNEATH e SOKAL (1973) citam vários autores que tentaram sumarizar e comparar **estes** coeficientes.

Os coeficientes de similaridade podem ser divididos em três grupos (adaptado de SNEATH e SOKAL, 1973): coeficientes de associação, de correlação e de similaridade probabilística.

2.4.1.2.1. Coeficientes de Associação

Existem muitos coeficientes de associação na literatura biológica e não-biológica, muitos deles não sendo aplicado mais de uma vez, não podendo assim serem validados empiricamente. SNEATH e SOKAL (1973) discutem os coeficientes de associação que têm sido extensivamente utilizados em taxonomia numérica,

Os coeficientes de associação constituem-se de vários algoritmos envolvendo dados qualitativos. Também podem ser aplicados para variáveis quantitativas ou ordinais, porém, neste caso, parte da informação é perdida na recodificação dos caracteres.

Geralmente, os coeficientes de associação são calculados com caracteres binários, os quais são por conveniência codificados como 0 ou 1. Os códigos 0 e 1 podem representar a ausência ou a presença de uma característica ou o resultado, negativo ou positivo, de uma reação bioquímica; respectivamente; ou pode ser uma arbitrária designação, como no caso de uma estrutura possuindo somente duas formas, sendo o código 0 designado para uma forma e o 1 para outra.

Quando caracteres qualitativos binários são comparados em relação a pares de indivíduos em uma matriz de dados, o resultado pode ser sumariado em uma tabela convencional 2 x 2 de frequência:

		indivíduo	
		1	0
indivíduo k	1	a	b
	0	c	d

Sendo **a** o número de caracteres codificados com 1 ("concordância positiva") em ambos os indivíduos, **d** é o número de caracteres codificados com zero para ambos os indivíduos ("concordância negativa") e **b** e **c** registram o número de caracteres nos quais ambos os indivíduos discordam.

- Coeficiente de concordância simples

$$S_{sm} = \frac{a + d}{a + b + c + d}$$

É um dos mais antigos e mais simples coeficientes. Sua forma complementar corresponde ao quadrado da distância euclidiana média baseada em caracteres qualitativos binários não-padronizados, ou seja,

$$d = \sqrt{1 - S_{sm}}$$

- Coeficiente de Jaccard

$$S_j = \frac{a}{a + b + c}$$

Este coeficiente omite as concordâncias negativas. Em sua classe é o mais simples dos coeficientes.

Há dúvidas se as concordâncias negativas deveriam ser incorporadas no coeficiente de associação. Pode ser argumentado que a avaliação da similaridade entre dois indivíduos, baseada na ausência mútua de certas características, é imprópria. Citando um exemplo de JOHNSON e WICHERN (1988), em um agrupamento de pessoas é notório que duas delas lendo grego antigo é uma evidência mais forte de similaridade que a ausência dessa habilidade.

Muitas das aplicações de coeficientes de associação, principalmente no campo da microbiologia, têm incluído concordâncias negativas em seus coeficientes. O coeficiente de Jaccard é apropriado quando as concordâncias negativas devem ser excluídas.

SNEATH e SOKAL (1973) e JOHNSON e WICHERN (1988) citam diversos outros coeficientes, alguns dos quais fornecem pesos diferentes as concordâncias e discordâncias.

Estes coeficientes podem ser utilizados em quaisquer dos métodos de agrupamento hierárquicos, simplesmente pela alteração de seus sinais, ou subtraindo-os da unidade.

Muitos dos coeficientes de associação e de correlação podem ser relacionados com distâncias quando eles são expressos como complementos de seu valor máximo, ou em termos de uma medida especial. Assim, a distância é considerada como o complemento da similaridade.

Segundo JOHNSON e WICHERN (1988), é sempre possível a construção de similaridades a partir de distâncias, por exemplo:

$$S_{ik} = \frac{1}{1 + d_{ik}}$$

em que $0 < S_{ik} \leq 1$ é a similaridade entre os indivíduos i e k e d_{ik} é a correspondente distância.

Porém, nem sempre distâncias podem ser obtidas a partir de Similaridade Segundo GOWER (1966a), isso só pode ser feito se a matriz de similaridade for positiva semidefinida. Com essa condição e a similaridade máxima sendo S_{ii} igual a unidade,

$$d_{ik} = \sqrt{2(1 - S_{ik})}$$

por exemplo, possui a propriedade de distância (JOHNSON e WICHERN, 1988).

2.4.1.2.2. Coeficiente de Correlação

O mais comum é o coeficiente de correlação momento-produto de Pearson:

$$r_{ik} = \frac{\sum_{j=1}^p (X_{ji} - \bar{X}_i)(X_{jk} - \bar{X}_k)}{\sqrt{\sum_{j=1}^p (X_{ji} - \bar{X}_i)^2 (X_{jk} - \bar{X}_k)^2}}$$

em que X_{ji} e X_{jk} são os valores do carácter j para os indivíduos i e k , respectivamente; \bar{X}_i e \bar{X}_k são as médias de todos os valores para os indivíduos i e k , respectivamente; e p é o número de caracteres.

O coeficiente expressa a relação entre dois indivíduos, sendo análogo a eq. 3.2b (ver seção 3.5.). que, porém, expressa a correlação entre duas variáveis.

Segundo SNEATH e SOKAL (1973), não se aplicam testes de significância estatística aos coeficientes de correlação, pois estes não possuem distribuição normal bivariada para dados de taxonomia numérica, por causa da heterogeneidade dos vetores representativos dos indivíduos e dos caracteres não serem independentes para cada indivíduo.

Um dos inconvenientes da utilização dos coeficientes de correlação na análise de agrupamento r que poderia ocorrer uma perfeita correlação entre indivíduos não-idênticos, onde o vetor de um indivíduo seria o de outro multiplicado por um escalar. Essa é uma propriedade indesejável da correlação como medida de similaridade. Este coeficiente é raramente utilizado.

2.4.1.2.3. Coeficientes de Similaridade Probabilística

A filosofia destes coeficientes é que a concordância entre dois indivíduos quanto a caracteres raros é menos provável que quanto a caracteres comuns, e deveriam, portanto, ter **peso maior**. Esses coeficientes não serão relatados aqui. Veja SNEATH e SOKAL (1973).

2.4.1.3. Medidas de Similaridade Utilizadas para Dados de Eletroforese de Isoenzimas e Proteínas Solúveis e de DNA

É comum a utilização de "cluster análise" na interpretação dos resultados obtidos por eletroforese de isoenzimas ou proteínas solúveis, em estudos de taxonomia e variabilidade genética de microrganismos. Outras análises também podem ser utilizadas. BENIN et alii (1988) utilizaram a análise fatorial e SMITH (1984), ZHU et alii (1988) e CHEN et alii (1991). a análise de componentes principais

Como medidas de similaridade genética utilizadas para dados de eletroforese citam-se a distância euclidiana (ver **seção 2.4.1.1.1.**). os coeficientes de associação S_j e S_m (ver **seção 2.4.1.2.1.**) e as medidas de identidade genética propostas por NEI (1972 e 1973). Hedrick (1971) e Rogers (1972). citados por ALFENAS et alii (1991). os quais discutem a aplicação de várias dessas medidas. Além dessas, existem outras, como as utilizadas por OTROSINA e COBB (1987). LINDE et alii (1990). TUSKAN et alii (1990). HELLMANN e CHRIST (1991) e PETRUNAK e CHRIST (1992).

Segundo ALFENAS et alii (1991). a distância euclidiana e o coeficiente de associação S_j devem ser utilizados nos casos em que a análise genética de progênies é impraticável (por exemplo: certas espécies de fungo que não apresentam fase sexuada e segregação). dificultando a interpretação de zimogramas. Nesses casos, pode-se proceder a análise de bandas

isoenzimáticas individuais para diferenciar isolados ou populações. A distância euclidiana é estimada com base nas frequências das bandas isoenzimáticas. Os coeficientes de associação são apropriados para dados qualitativos binários, como presença ou ausência de bandas isoenzimáticas.

Para a avaliação da similaridade a partir de dados provenientes de análises de DNA, NEI e LI (1979) propuseram o coeficiente de similaridade $F_{ik} = 2n_{ik} / (n_i + n_k)$, em que n_{ik} é o número de fragmentos de DNA comum entre dois indivíduos e n_i e n_k são os números de fragmentos para os indivíduos i e k , respectivamente. Esse coeficiente tem sido comumente utilizado em fitopatologia (HARTUNG e CIVEROLO, 1989; GRAHAM et alii, 1990; CHEN et alii, 1992; LEE et alii, 1992 e WHISSON et alii, 1992). COOKSEY e GRAHAM (1989) e PRUVOST et alii (1992) utilizaram como medida de distância o complemento de F_{ik} , ou seja, $1 - F_{ik}$.

2.4.2. Métodos de Agrupamento

Como no processo de agrupamento é desejável ter informações relativas a cada par de indivíduos (ou populações), o número de estimativas de medidas de similaridade é relativamente grande (igual a $n(n - 1) / 2$, em que n é o número de indivíduos), o que torna impraticável o reconhecimento de grupos homogêneos pelo simples exame visual daquelas estimativas. Para realizar essa tarefa, faz-se uso dos métodos de agrupamento (CRUZ, 1990). Deve-se enfatizar que os métodos são basicamente descritivos e completamente numéricos, sendo o número de grupos formados desconhecido previamente. Alguns autores (BONDE et alii, 1984; HARTUNG e CIVEROLO, 1989; NYGAARD et alii, 1989 e CHEN et alii, 1992), porém, estudaram a variabilidade entre isolados de fungos ou bactérias fitopatogênicos, baseada apenas em uma medida de similaridade, não utilizando métodos de agrupamento.

De acordo com Everitt (1980), citado por RIBOLDI (1986), os métodos de agrupamento podem ser classificados em cinco tipos, dos quais os mais importantes são os métodos hierárquicos e os de otimização. Os métodos hierárquicos, citados por SNEATH e SOKAL (1973), são subdivididos em métodos hierárquicos divisivos, como o apresentado por EDWARDS e CAVALLI-SFORZA (1965) (o qual é também utilizado no teste univariado de SCOTT e KNOTT, 1974), e em métodos hierárquicos aglomerativos, como o método do vizinho mais próximo

("single linkage method" - SLM); o do vizinho mais distante ("complete linkage method" - CLM); o da ligação média ("average linkage"), ponderado (WPGMA - método ponderado de agrupamento aos pares utilizando médias aritméticas) ou não-ponderado (UPGMA - método não-ponderado de agrupamento aos pares utilizando médias aritméticas); o do centróide, ponderado (WPGMC) ou não-ponderado (UPGMC); e o método de WARD (1963). Como exemplo de método de otimização, tem-se o método de Tocher, citado por RAO (1952).

Nos métodos hierárquicos, os indivíduos ou populações *são* agrupados por um processo de aglomeração ou divisão (métodos hierárquicos aglomerativos ou divisivos, respectivamente), o qual se repete em vários níveis até que seja estabelecido um dendograma ou diagrama de árvore, no qual o eixo "x" representa os indivíduos ou as populações e o eixo "y", a similaridade (ou dissimilaridade), ou vice-versa. As ramificações da árvore originam $n - 1$ ligações. A primeira bifurcação representa a primeira ligação; a segunda, a segunda ligação; e assim sucessivamente até todos se unirem na "árvore", cujos ramos representam os grupos (MARDIA et alii, 1979). Os indivíduos *são* agrupados de acordo com a sua proximidade ou similaridade. O procedimento básico de todos os métodos aglomerativos é similar, iniciando com o cálculo de uma matriz de similaridade entre os indivíduos (ou populações) e finalizando com um dendograma, que representa a **síntese** dos resultados, o que ocasiona certa perda de informações. Ainda assim, no caso desta perda ser pequena, o resumo da informação é importante para a comparação, classificação e discussão dos fenômenos. Nesse caso, não há preocupação com o número ótimo de grupos, uma vez que o interesse maior **está** na árvore e nas ramificações que **são** obtidas.

O conceito de representação hierárquica dos dados foi desenvolvido primeiramente na biologia, pois para tais dados supõe-se existir uma estrutura hierárquica (MARDIA et alii, 1979). No entanto, os métodos hierárquicos também *são* usados em muitas outras áreas, nas quais a estrutura hierárquica pode não ser a mais apropriada e, obviamente, o seu uso necessitaria de avaliação mais cuidadosa. Embora os resultados sejam tradicionalmente representados por dendogramas que têm a vantagem de serem facilmente interpretáveis; SNEATH e SOKAL (1973) discutem outros modos de apresentar os resultados de "cluster análise".

Os métodos hierárquicos aglomerativos de agrupamento geralmente podem reproduzir distâncias entre grupos adjacentes fielmente, mas poderão apresentar distorções em distâncias entre membros de grupos maiores (SNEATH e SOKAL, 1973). A este respeito, a análise de

componentes principais caracteriza-se pela representação adequada de distância entre grupos principais, mas é capaz de distorcer a distância entre vizinhos mais próximos (ROHLF, 1968). Recomenda-se, então, sempre que possível, proceder as duas análises. De maneira geral, ao se fazer o agrupamento, as informações em nível de **indivíduo são** perdidas, restando apenas aquelas sobre médias dos grupos. Isto explica, segundo CRUZ (1990), a tendência de se proceder, simultaneamente, "cluster análise" e análise de componentes principais ou de variáveis canônicas. VARMA e GULATI (1982) e DAS e GUPTA (1984) obtiveram concordâncias de resultados, utilizando D^2 e análise canônica, com relação a identificação de genótipos com mais alta similaridade em seus estudos.

Nos métodos de otimização, **os** grupos **são** formados pela adequação de algum critério de agrupamento, ou seja, o objetivo é alcançar uma partição dos indivíduos que otimize (maximize ou minimize) alguma medida pré-definida (RIBOLDI, 1986). As diferenças entre técnicas são funções dos métodos utilizados para uma partição inicial e dos critérios de agrupamento utilizado para a otimização, alguns dos quais **são** citados por RIBOLDI (1986). No método proposto por Tocher, citado por RAO (1952), estabelece-se o critério de manter a distância média intragrupo sempre inferior a qualquer distância intergrupo. Esses métodos diferenciam-se dos hierárquicos por serem **os** grupos formados mutuamente exclusivos **ou**, sob o contexto de teoria de conjuntos, por subdividir o grupo original em subgrupos não-vazios, cuja interseção é nula e a união reconstitui o conjunto total (CRUZ, 1990). RIBOLDI (1985) discute **os** problemas no **uso** das técnicas de otimização.

Os métodos de agrupamento foram desenvolvidos por analistas preocupados com a resolução de problemas práticos específicos, o que ocasionou grande quantidade de técnicas propostas. **As** diferenças entre **os** métodos existem em função de diferentes formas de definir proximidade entre um indivíduo e um grupo contendo vários indivíduos ou entre grupos de indivíduos. JOHNSON e WICHERN (1988) ilustram a definição de proximidade de alguns métodos hierárquicos aglomerativos. Ao contrário de outros campos da análise multivariada, que se encontram teoricamente mais fundamentados, **os** desenvolvimentos teóricos com relação aos métodos de agrupamento não acompanharam as afirmativas empíricas. Muitas técnicas são propostas, mas não há, ainda, uma teoria generalizada e amplamente aceita (CURTI, 1983).

Não existe método aceito como o melhor. O pesquisador tem de decidir qual o método mais adequado ao seu trabalho, uma vez que **os** diferentes métodos podem levar a diferentes padrões de agrupamento, o que poderá por em dúvida a validade do agrupamento obtido (RIBOLDI, 1985; MANLY, 1986 e CRUZ, 1990).

Em virtude de **os** métodos hierárquicos aglomerativos serem mais simples computacionalmente (SNEATH e SOKAL, 1973), **eles** têm sido mais utilizados que **os** divisivos (MANLY, 1986). Dos métodos hierárquicos aglomerativos, o SLM foi o melhor em estudos empíricos desenvolvidos e é o único que satisfaz vários critérios matemáticos (Everitt, 1980, citado por RIBOLDI, 1985). Segundo Johnson (1967), citado por RIBOLDI (1986), a escolha do método SLM **se** justifica pela simplicidade, facilidade computacional e por possuir a propriedade de invariância **sob** transformações lineares dos dados.

Uma vez que o método SLM une grupos pela menor ligação entre **eles**, a técnica não pode discernir grupos pobremente separados, como ilustram JOHNSON e WICHERN (1988). Por outro lado, é um dos poucos métodos que pode delinear grupos não-elipsóides. A tendência para selecionar grupos como filamentos longos é conhecida como "encadeamento" ("chaining"), que pode ser enganoso, **se os** indivíduos em extremidades opostas da cadeia são, de fato, completamente dissimilares, sendo isto ilustrado por SNEATH e SOKAL (1973) e JOHNSON e WICHERN (1988). SNEATH e SOKAL (1973) discutem quatro modificações no método para evitar **este** problema.

O método CLM é similar ao SLM, constituindo-se na sua antítese direta, em razão de a distância entre dois grupos ser definida como a maior distância entre pares de indivíduos em cada grupo. Assim, **este** método tende a produzir grupos compactos sem efeito de encadeamento. Nos métodos SLM e CLM, a utilização de uma nova medida de similaridade que possua a mesma ordenação relativa como as medidas iniciais não alterará a configuração dos grupos. **Os** métodos de ligação média, ao contrário, consideram a distância entre dois grupos como a distância média entre todos **os** pares de indivíduos e quaisquer alterações na avaliação de distância podem afetar a configuração final dos grupos.

O método UPGMA é o mais freqüentemente utilizado em ecologia e sistemática (JAMES e McCULLOCH, 1990) e em taxonomia numérica (SNEATH e SOKAL, 1973). Entre **os**

taxonomistas existe grande concordância na utilização dos métodos de agrupamento aos pares, ponderados ou não. para evitar os extremos dos SLM e CLM (CURI, 1983).

LINDE et alii (1990) encontraram resultados similares no estudo de variabilidade de *Uromyces appendiculatus* baseado em isoenzimas e virulência, utilizando os métodos SLM, UPGMA e do centróide.

Um problema comum a todas as técnicas de agrupamento é a dificuldade de decidir quanto ao número de grupos presente nos dados. É necessária uma decisão quanto ao estágio do processo de agrupamento que pode ser considerado como ótimo (RIBOLDI, 1986). Não há critério definitivo para se determinar o melhor número de grupos, embora CURI (1983) cite diversos autores que o propuseram. Por outro lado, a interpretabilidade e simplicidade são importantes na análise de dados e qualquer inferência rígida sobre o número de grupos pode não ser produtiva. Para a solução desse problema, quando se utilizam métodos hierárquicos, as delimitações dos grupos podem ser estabelecidas por um exame visual do dendograma, em que se avaliam pontos de alta mudança de nível, tomando-os, em geral, como delimitadores do número de indivíduos (ou populações) para determinado grupo (CRUZ, 1990). Além disso, para verificar a adequação da partição obtida, uma vez fixado o número de grupos e conhecidos os seus componentes, é comum a utilização da análise discriminante. Outra opção é a aplicação de várias técnicas de agrupamento ao mesmo conjunto de dados e somente grupos formados por todos ou pela maioria dos métodos seriam aceitos (RIBOLDI, 1985). FERNANDES e LIMA (1991) utilizaram o teste t para médias a fim de verificarem se os grupos diferiam, estatisticamente, entre si.

Os resultados obtidos pelos métodos de agrupamento necessitam de um reexame da matriz de dados ou de ser complementados por outras técnicas multivariadas, como a análise de componentes principais ou a análise canônica. Para a avaliação da partição, pode-se utilizar a análise discriminante. HANSEN e HAMM (1983) confirmaram os resultados obtidos pelas análises canônica e "cluster análise" com a análise discriminante.

Segundo CURI (1983), independente do método utilizado, é importante que sejam efetuadas medidas do grau de ajuste entre a matriz de similaridade original e a matriz resultante do processo de agrupamento. Quanto maior o grau de ajuste, menor será a distorção ocasionada pelo método, ou seja, ocorrerá maior retenção da informação contida na matriz de similaridade. Não existe nenhuma concordância geral sobre classificação ótima, porém existem vários critérios

para avaliar a eficiência do agrupamento. O mais comumente aplicado é o coeficiente de correlações cofenéticas (CCC) desenvolvido por SOKAL e ROHLF (1962), cujo cálculo é também exemplificado por ROMESBURG (1984) e por SILVEIRA NETO (1986), que também relata outras medidas de ajuste. Segundo SNEATH e SOKAL (1973), um valor de CCC > 0,8 é comumente satisfatório, sendo **este** valor também considerado por GOTTWALD et alii (1991) na avaliação de similaridade de strains de *Xanthomonas campestris* pv. *citrumelo*. CHEN et alii (1991) também utilizaram este critério para avaliar agrupamento de isolados de *Pythium* pelo método UPGMA.

Outro aspecto importante a ser considerado em alguns estudos, não só com o uso de "cluster análise", como também com análise de componentes principais e análise canônica, é a consistência das estimativas e do padrão de agrupamento, quando a análise é repetida no tempo e, ou, espaço.

Mais detalhes e exemplos numéricos de métodos de agrupamento podem ser encontrados em MARDIA et alii (1979), CURI (1983), RIBOLDI (1985 e 1986). e MANLY (1986).

SNEATH e SOKAL (1973). SINGH e CHAUDHARY (1979) e JOHNSON e WICHERN (1988) relatam métodos alternativos não-numéricos de agrupamento. Algumas vezes **esses** métodos podem ser informativos, porém quase sempre não o são (JOHNSON e WICHERN, 1988).

2.4.3. Aplicações de "Cluster Analise" em Fitopatologia

A análise pode ser utilizada em estudos de compatibilidade vegetativa entre isolados fúngicos, facilitando a determinação dos VCGs (grupos de compatibilidade vegetativa).

A partir de dados de **testes** de compatibilidade vegetativa entre isolados de *Verticillium dahliae* de diferentes hospedeiros, utilizando mutantes *nit* (não-utilizadores de nitrato), STRAUSBAUGH et alii (1992) aplicaram "cluster análise" (baseada no coeficiente de associação S_j) para determinar os VCGs. A análise permitiu, também, evidenciar subgrupos dentro dos VCGs.

Em epidemiologia, "cluster análise" pode ser utilizada para avaliar a similaridade entre curvas de progresso de doença (CPDs). Como variáveis, consideram-se cada época de avaliação, ou então, elementos da CPD. HAU e KRANZ (1990) discutem a utilização de "cluster análise" e outras técnicas de análise multivariada em epidemiologia.

"Cluster análise" foi utilizada para descobrir padrões de epidemias. Kranz (1968), citado por HAU e KRANZ (1990), encontrou 12 grupos de CPDs na comparação de 40 patossistemas. CAMPBELL et alii (1980) analisaram 11 curvas de progresso da podridão do hipocótilo do feijoeiro, identificando, no mínimo, dois tipos de epidemias, que diferiram principalmente quanto a taxa de progresso de doença.

THOMPSON e REES (1979) agruparam 45 cultivares de trigo, segundo as suas respostas a epidemia da ferrugem do colmo (*Puccinia graminis* f.sp. *tritici*), utilizando "cluster análise" baseada em seis avaliações de doença durante o ciclo da cultura.

Recentemente, com o aumento dos estudos de variabilidade de microrganismos utilizando isoenzimas e análise de DNA, há um crescente aumento na utilização de "cluster análise" para avaliar a similaridade entre isolados de fungos, bactérias e MLO (microrganismos tipo micoplasmas) fitopatogênicos. "Cluster análise" é mais precisa que a observação visual dos resultados, simplificando a sua apresentação, possibilitando que a relação entre isolados seja mais facilmente compreendida. Além disso, a análise pode sugerir relações e hipóteses a serem investigadas.

COOKSEY e GRAHAM (1989) utilizaram como medida de distância o complemento do coeficiente de Similaridade F ($D = 1 - F$) e o método UPGMA de agrupamento em dados de análise de DNA de dez strains de *Pseudomonas syringae* pv. *tomato* e dez strains de *Xanthomonas campestris* pv. *vesicatoria*. Strains de *Pseudomonas* resistentes ao cobre formaram um grupo separado e uniforme, apesar de terem sido isolados de diferentes locais. Strains de *Xanthomonas* isolados de pimentão e tomate foram separados em grupos geográficos, mas não foi observada relação entre os grupos e os hospedeiros. Segundo os autores, os resultados sugerem que a especificidade em relação ao hospedeiro evoluiu separadamente após a distribuição geográfica.

HARTUNG e CIVEROLO (1989) utilizaram a medida de similaridade F (NEI e LI, 1979) no estudo da similaridade genética de vários patovares de *Xanthomonas campestris* baseado em dados de RFLP (restriction fragment length polymorphism), com o objetivo principal de avaliar a relação entre strains de *X. campestris* isolados de viveiros de citros na Flórida (strains FCN) e *X. campestris* pv. *citri*, agente etiológico do cancro cítrico, o qual possui quatro formas

previamente conhecidas (A, B, C e D). Foram estudados 21 strains de *X. campestris* pv. *citri*, 14 strains FCN e dez strains de cinco outros patovares. Verificaram-se similaridade alta entre strains de mesmo patovar e baixa entre strains de diferentes patovares. Os strains da forma A foram altamente similares entre si, sendo claramente distintos dos strains das demais formas (B, C e D) de *X. campestris* pv. *citri*, os quais formaram um segundo grupo distinto. Os strains FCN foram mais similares entre si que em relação aos outros patovares de *X. campestris*, a exceção de *X. campestris* pv. *alfalfae*. A grande dissimilaridade entre strains FCN e *X. campestris* pv. *citri* (sendo consistente com sintomatologia, epidemiologia e distribuição) permitiu aos autores concluir que a doença associada a viveiros de citros na Flórida não é uma forma de cancro cítrico, e que o patógeno não é um strain de *X. campestris* pv. *citri*.

GRAHAM et alii (1990) utilizaram "cluster análise" (F, UPGMA) em dados de RFLP e de análise de ácidos graxos no estudo da relação entre strains de *Xanthomonas campestris* provenientes de citros e outros hospedeiros.

PRUVOST et alii (1992) utilizaram "cluster análise" para avaliar a variação genética de strains de *X. campestris* pv. *citri*, associados com as formas A, B, C e D de cancro cítrico.

GOTTWALD et alii (1991) utilizaram "cluster análise" (UPGMA) na avaliação da similaridade de strains de *Xanthomonas campestris* pv. *citrumelo* baseada em dados de análise de DNA, sorologia e virulência em folhas de citros destacadas. Todos os strains classificados como fortemente virulentos (os autores usaram o termo agressividade) apresentaram alta similaridade com relação a análise de RFLP e, com exceção de um strain, com relação à reação com seis anticorpos monoclonais. Houve concordância entre os resultados da análise de DNA e sorologia. Ambas foram capazes de distinguir os strains mais virulentos dos menos virulentos.

CHEN et alii (1992) utilizaram o coeficiente de similaridade F (NEI e LI, 1979) no estudo da variação genética entre 24 strains de *Xylella fastidiosa*, isolados de vários hospedeiros, baseado em dados de DNA, obtidos por RFLP.

LEE et alii (1992) estudaram 15 strains de MLO (organismos tipo micoplasma), utilizando "cluster análise" (F, UPGMA) baseada em dados de DNA, obtidos por RFLP. A análise revelou a existência de três grupos, concordando com os resultados de sorologia, mas não com a classificação prévia baseada em gama de hospedeiros e, ou, sintomatologia.

WHISSON et alii (1992) utilizaram "cluster análise" (baseada em uma distância genética obtida a partir de F_e e no método de agrupamento UPGMA), para avaliar a relação entre dez isolados australianos e norte-americanos de *Phytophthora megasperma* f. sp. *glycinea*, baseada em dados de análise de DNA. Com a exceção de um isolado, os demais formaram dois outros grupos que coincidiram com a origem geográfica. A alta similaridade entre os grupos separados geograficamente sugere, segundo os autores, que os isolados australianos são derivados de ancestrais comuns aos isolados norte-americanos.

BONDE et alii (1984), utilizando o coeficiente de similaridade de Rogers (1972), citado por ALFENAS et alii (1991), aplicado a dados de isoenzimas, conseguiram distinguir espécies de *Peronosclerospora*. Os dados apresentaram evidências que certas espécies poderiam ter sido mal identificadas. Assim, os autores sugerem uma reavaliação taxonômica do gênero.

OTROSINA e COBB (1987) utilizaram "cluster análise" baseada em dados de nove isoenzimas, no estudo da variabilidade genética de *Verticicladiella wagneri*.

ZHU et alii (1988) utilizaram "cluster análise" (distância euclidiana e método de Ward) baseada em dados de presença ou ausência de 14 bandas isoenzimáticas, para resumir a variação genética intraespecífica do fungo ectomicorrízico *Suillus tomentosus*. Os padrões de variação genética sugeridos por "cluster análise" foram também demonstrados pela análise de componentes principais.

NYGAARD et alii (1989) estudaram a variabilidade entre 300 isolados de quatro espécies de *Phytophthora*, utilizando o coeficiente de similaridade S_{sm} baseado na presença ou ausência de bandas isoenzimáticas.

TUSKAN et alii (1990) estudaram 201 isolados *Peridermium harknessii* coletados de 13 regiões geográficas e três hospedeiros, agrupando por "cluster análise" [distância genética proposta por NEI (1972) e método SLM] as regiões em dois grupos principais, segundo dados de isoenzimas de seus isolados. Segundo os autores, esta diversidade tem implicações no estudo de interações patógeno-hospedeiro.

CHEN et alii (1991) compararam 38 isolados representando sete espécies de *Pythium* pela eletroforese de proteínas solúveis e isoenzimas, sendo assinalada a presença ou ausência (1 ou 0) de uma banda particular de proteína ou enzima. Na interpretação dos resultados foram utilizadas "cluster análise" (S_{sm} , UPGMA) e análise de componentes principais, as quais foram

concordantes entre si. Segundo **os** autores, a comparação de isolados por observação visual foi menos precisa em relação **as** análises multivariadas, que possibilitaram que a relação entre espécies fosse mais facilmente interpretada. A análise de componentes principais e "cluster análise" das proteínas solúveis ou das isoenzimas tenderam agrupar **os** isolados por espécie. **Os** resultados sugeriram a possibilidade de **P. arrhenomanes** e **P. graminicola** não representarem espécies distintas. A variação com relação a isoenzimas não foi associada ao hospedeiro de origem ou habitat.

BONDE et alii (1991) utilizaram o coeficiente de similaridade de Rogers (1972), citado por ALFENAS et alii (1991), no estudo de variabilidade, baseado em isoenzimas, de dez isolados de **Septoria** de citros dos Estados Unidos (referidos como **S. citri**) e 18 isolados da Austrália (referidos como **S. depressa**). Pôde-se concluir que **os** isolados, com uma exceção, pertencem a uma única espécie.

HELLMANN e CHRIST (1991) estudaram a variação isoenzimática de 63 isolados de *Ustilago hordei*, estimando distancias genéticas baseadas no número de alelos diferentes entre dois tipos eletroforéticos e em dados de virulência. **Os** resultados de "cluster análise" não apresentaram nenhuma relação entre grupos baseados em dados de isoenzimas e grupos baseados em dados de virulência.

PETRUNAK e CHRIST (1992) estudaram a variabilidade isoenzimática de 54 isolados de *Alternaria solani* e 96 isolados de *A. alternata* de vários hospedeiros. **Os** 150 isolados foram classificados em 35 tipos eletroforéticos. "Cluster análise", na qual a distancia genética foi considerada como o número de diferentes alelos entre dois tipos eletroforéticos, dividiu **os** isolados, com exceção de um, em grupos que corresponderam **as** espécies. Não foi encontrada associação significativa dos fenótipos isoenzimáticos com **os** hospedeiros ou com a origem geográfica dos isolados.

"Cluster análise" é utilizada em estudos de variabilidade de fitopatógenos, com relação a características culturais, morfológicas, **nutricionais**, bioquímicas e fisiológicas, dentre outras.

Isolados de *Sphaeropsis sapinea* podem ser distinguidos pelas características do micélio em meio de cultura, sendo designados isolados tipo A e **B**. Com o objetivo de determinar **se** esses tipos representam duas distintas formas do patógeno, PALMER et alii (1987) estudaram suas características culturais, virulência e isoenzimas. **Os** isolados tipo A e B apresentaram

diferenças entre si quanto ao efeito da luz na esporulação, padrões de isoenzimas e virulência. "Cluster análise" (distância euclídiana média e UPGMA) foi usada para determinar se cinco isolados tipo A e dois tipo B poderiam ser distinguidos com base em características de crescimento em seis meios de cultura, sob seis diferentes temperaturas e dimensões dos esporos. "Cluster análise" revelou dois grupos distintos, de acordo com o tipo de isolado.

ENGELBRECHT e HATTINGH (1989) examinaram os grupos fenotípicos de strains sul-africanos e de referências, de *Pseudomonas solanacearum* de fumo e outros hospedeiros. Foi utilizada "cluster análise" (S_{sm} , UPGMA) no estudo de 59 strains, baseado em 55 características. Com 85% de similaridade, três grupos e dois solitários strains foram distinguidos. Grupos 1, 2 e 3 corresponderam aos biovars III, I e II, respectivamente. Todos os strains dos grupos 1 e 2 pertenciam a raça 1 e os do grupo 3, a raça 3.

HANSEN e HAMM (1983) utilizaram "cluster análise" e análise canônica na avaliação da similaridade entre isolados de *Phytophthora megasperma* de diferentes hospedeiros quanto a caracteres morfológicos.

ROOS e HATTINGH (1987) utilizaram "cluster análise" (S_{sm} , UPGMA) baseada em 215 características culturais, bioquímicas, nutricionais e fisiológicas, com o objetivo de examinar o agrupamento fenotípico de 113 strains de *Pseudomonas syringae*, isolados de diferentes hospedeiros.

"Cluster análise" pode ser utilizada em estudos de variabilidade de virulência entre isolados fitopatogênicos. Em muitos estudos a reação de cada genótipo hospedeiro C considerada uma variável.

REGO (1993) utilizou o método de Tocher e a distancia euclidiana média baseada na reação de sete variedades diferenciadoras no estudo da variabilidade de 22 isolados de *Colletotrichum orbiculare*.

SCHILDER e BERGSTROM (1990), estudando a variação de virulência em 17 isolados de *Pyrenophora tritici-repentis*, demonstraram a existência de moderada especialização fisiológica. Embora todos os isolados tenham sido patogênicos a todos os 12 cultivares de trigo estudados, os resultados de avaliações de resistência em programas de melhoramento podem ser alterados consideravelmente de acordo com o isolado utilizado. Os autores utilizaram "cluster analise" (distancia euclidiana e método de agrupamento do centróide) para determinar a

similaridade entre **os** isolados, baseados em **seus** padrões de virulência em 12 cultivares de trigo e vice-versa. O objetivo do **uso** de "cluster análise" foi auxiliar a **seleção** de isolados mais apropriados para a avaliação de resistência em **genótipos** de trigo.

LINDE et alii (1990) utilizaram os métodos de agrupamento SLM, do centróide e UPGMA e o coeficiente de associação S_{sm} em "cluster análise" de 27 isolados de **Uromyces appendiculatus**, baseados em 13 isoenzimas e virulência em 18 linhagens diferenciadoras de **feijoeiro**. Foi encontrada maior diversidade para virulência em relação a isoenzimas. Não houve relação entre origem geográfica e similaridade dos isolados. Foram identificados três grupos fenotípicos, suportando a hipótese que **U. appendiculatus** não é uma espécie homogênea.

LEBEDA e JENDRULEK (1987b) demonstraram o potencial da aplicação de "cluster análise" no estudo da especificidade de interação hospedeiro-patógeno, utilizando o patossistema **Bremia lactucae-Lactuca sativa**.

LEBEDA e JENDRULEK (1987a) utilizaram "cluster análise" (S_{sm} , SLM) no estudo da variabilidade genética de fitopatógenos. Na avaliação da similaridade entre fenótipos de virulência de **Bremia lactucae**, foram agrupados **48** isolados do patógeno de acordo com a presença ou ausência de 11 fatores de virulência. **Os** resultados apresentaram relação entre a origem geográfica dos isolados e a sua constituição genética.

A análise pode ser utilizada em estudos de resistência de plantas a varias doenças, simultaneamente.

ANDERSON et alii (1990) utilizaram a análise de componentes principais e "cluster análise", com base na distância euclidiana, para avaliar **55** genótipos de amendoim quanto a resistência a **um** complexo de três doenças.

PATAKY et alii (1988) utilizaram "cluster análise" (método de agrupamento de Ward) no estudo da resistência de 66 híbridos de milho-doce a quatro doenças, avaliadas em três anos consecutivos, A análise revelou grupos de híbridos com similar reações **as** doenças.

Além dos já relatados anteriormente, existem outros exemplos de aplicações de "cluster análise", que **são** citados a seguir.

Existe um grande número de testes rápidos, desenvolvidos com a proposta de identificar a capacidade de produtos químicos produzirem efeitos mutagênicos ou carcinogênicos. Em decorrência da existência de varias teorias sobre a indução do câncer, o número de **testes**

propostos tem aumentado muito. Porém, ainda não existe concordância geral sobre os mesmos, o que gera situação confusa e complexa. Com o objetivo de saber se existem grupos homogêneos de testes com características distinguíveis a respeito das respostas dos produtos químicos, BENIGNI e GIULIANI (1985) utilizaram "cluster análise" no estudo da performance de 21 testes de identificação da habilidade de produtos químicos produzirem efeitos carcinogênicos ou mutagênicos, considerando 42 produtos químicos.

LEONARD e LEATH (1990) utilizaram a medida de identidade genética proposta por NEI (1972) e uma modificação desta e o método UPGMA, para estudar a divergência genética em populações de *Cochliobolus carbonum* em milho.

2.5. Outras Técnicas de Análise Multivariada

Existem outras técnicas multivariadas além das descritas anteriormente, algumas das quais serão aqui relatadas, embora superficialmente, por causa da sua importância e de seu grande potencial de utilização em fitopatologia.

2.5.1. "Path Análise" ou Análise de Trilha

Muitos estudos possuem como objetivo avaliar o efeito de uma variável ou de um conjunto de variáveis sobre outra. Nesses casos, muitas vezes é utilizada a análise de regressão, que é metodologia estatística utilizada para prever valores de uma variável resposta (dependente) a partir de um conjunto de valores de variáveis preditoras (independentes).

Em epidemiologia é comum o estudo da influência de variáveis climáticas na ocorrência e intensidade de doenças, utilizando-se, freqüentemente, para tanto, a análise de regressão. Porém, muitas variáveis meteorológicas e ambientais selecionadas para a análise são altamente correlacionadas. O problema é como determinar quais variáveis exercem influência relativamente grande na doença, simultaneamente reconhecendo os efeitos da correlação entre as variáveis climáticas (denominadas independentes). Quanto mais variáveis são incluídas em um estudo de correlações, mais complexas e importantes se tornam as associações indiretas. Em muitos casos, uma pequena correlação entre variáveis poderá ser resultante de um balanço de "paths" de

influência muito grande, porém **opostos**, induzidos por causas comuns, obscurecendo a **importância** de certos fatores ao longo daquele "path".

"Path análise" é um método baseado na decomposição do coeficiente de correlação em efeitos diretos e indiretos de uma variável pelas outras, ou seja, o efeito direto de uma variável independente em uma variável dependente pode ser separado dos efeitos indiretos da variável independente na variável dependente por virtude de ser altamente correlacionada com outra variável independente. A análise tem sido útil para encontrar causas diretas e indiretas de correlação e permitir um exame detalhado de forças específicas atuando para produzir uma correlação. Com o uso da "path análise" pode-se determinar quais variáveis necessitam ser medidas para capacitar a predição da doença.

Mais informações **são** fornecidas por LI (1975) e SINGH e CHAUDHARY (1979).

Muitos estatísticos não classificam "path análise" como uma técnica de análise **multivariada**, usando o termo apenas quando existe mais de uma variável "resposta" ("dependente") mensurada em uma unidade experimental.

Ainda é pequena a utilização desta técnica em fitopatologia.

COSTA et alii (1993) utilizaram "path análise" no estudo da influência de variáveis climáticas na intensidade da vassoura-de-bruxa (**Crinipellis** perniciosa) em cacauzeiros.

BOWERS et alii (1990) utilizaram "path análise" no estudo dos efeitos de variáveis climáticas no progresso de requeima (**Phytophthora capsici**) do pimentão. A técnica foi útil para determinar as variáveis que exerceram grande e direto efeito no desenvolvimento da doença e, portanto, as variáveis para incluir em estudos posteriores. A análise revelou relações que poderiam de outra forma permanecer desconhecidas, caso o estudo **fosse** baseado na análise de correlações somente. A quantidade cumulativa de chuva possuiu o maior efeito direto na intensidade e na taxa de progresso da doença.

"Path análise" foi utilizada por SNIJDERS e PERKOWSKI (1990) para estabelecer as relações entre a intensidade da doença causada por *Fusarium culmorum* e o conteúdo de **micotoxina** no grão, a redução do peso do grão e a redução da produção de trigo. "Path análise" revelou a relação entre a **micotoxina deoxynivalenol** e a redução do peso do grão.

VAN BRUGGEN e ARNESON (1986) avaliaram a importância relativa das variáveis que determinam a influência de *Rhizoctonia solani* no desenvolvimento e na produção do feijoeiro, pela "path análise".

HAMPTON (1975) utilizou "path análise" para estudar a redução da produtividade do feijoeiro causada pelos vírus do mosaico amarelo (BYMV) e mosaico comum (BCMV). O maior determinante da redução da produtividade, induzida por ambas viroses, isoladamente, foi o decréscimo do número de vagens por planta. Como a análise demonstrou que os efeitos indiretos dos componentes de produção (número de vagens por planta, número de sementes por vagem e peso da semente) foram pequenos para plantas saudáveis e infectadas, o autor pôde sugerir, mais apropriadamente para o BYMV, que nos programas de melhoramento, a tolerância ao vírus poderia ser avaliada em termos dos componentes de produção específicos, uma vez que a tolerância a este vírus é mais freqüentemente encontrada do que a resistência e imunidade.

2.5.2. Análise de Correlação Canônica

Em alguns conjuntos de dados multivariados as variáveis dividem-se naturalmente em dois grupos. A análise de correlações canônicas permite identificar e quantificar as associações entre os dois grupos de variáveis. É uma generalização da correlação e regressão que é aplicável quando os atributos de um único grupo de objetos podem ser divididos naturalmente em dois conjuntos (JAMES e McCULLOCH, 1990).

De um ponto de vista, a análise de correlação canônica é uma **extensão** da regressão múltipla. No caso da regressão múltipla, existe somente uma variável dependente. Por outro lado, a situação da correlação canônica envolve, no mínimo, duas variáveis do "lado dependente" e, no mínimo, duas variáveis do "lado independente". Embora alguns pesquisadores possam considerar um grupo de variáveis como variáveis independentes e outro grupo como variáveis dependentes, o método não faz tais distinções entre os dois grupos.

Na análise de regressão múltipla, as variáveis são separadas em um grupo **X**, contendo p variáveis, e um grupo **Y** contendo $q = 1$ variável. O método da regressão consiste em encontrar a combinação linear $a'X$ que seja a mais correlacionada com Y . Na análise de correlação canônica, o grupo **Y** contém $q \geq 2$ variáveis, sendo o seu principal objetivo resumir as

associações entre os grupos X e Y em termos de umas poucas covariâncias (ou correlações) escolhidas criteriosamente.

O método consiste em encontrar um vetor de coeficientes para cada um dos dois grupos de variáveis, tal que a correlação entre as duas combinações lineares seja máxima. Determina-se o primeiro par de combinações lineares que possuam a maior correlação. A seguir, determina-se outro par de combinações lineares que possua a segunda maior correlação entre todos os pares não-correlacionados com o par selecionado inicialmente. e assim sucessivamente. As combinações lineares são denominadas variáveis canônicas e suas correlações, correlações canônicas. O número de correlações canônicas é igual a dimensão do menor grupo de variáveis.

A análise de correlações canônicas possui certas propriedades similares as da análise de componentes principais, porém esta última considera as interrelações dentro de um grupo de variáveis, enquanto aquela considera a relação entre dois grupos de variáveis

dada par de variáveis canônicas fornece a correlação de uma combinação linear do primeiro grupo de variáveis com uma das variáveis do segundo grupo, refletindo, essencialmente, a predição dessa variável do segundo grupo, a partir do primeiro grupo de variáveis (DUNTEMAN, 1984).

Segundo DUNTEMAN (1984), o método é usualmente utilizado com variáveis quantitativas contínuas, sendo necessário assumir a existência de normalidade multivariada quando testes de significância estatística são utilizados. A análise também pode ser utilizada quando há uma mistura de variáveis contínuas e qualitativas (MARDIA et alii, 1979 e DUNTEMAN, 1984) ou se todas as variáveis são qualitativas (DUNTEMAN, 1984). Porém, segundo JAMES e McCULLOCH (1990), o procedimento não é ideal para variáveis qualitativas.

Em fitopatologia, cita-se a possibilidade de emprego desta metodologia nos casos em que se objetiva avaliar as relações entre, por exemplo, caracteres de resistencia versus caracteres morfológicos (ou fisiológicos) do hospedeiro, dados de isoenzimas versus virulência de isolados fitopatogênicos, dentre outras.

Mais informações sobre esta técnica são fornecidas por RAO (1952), MARDIA et alii (1979). ARNOLD (1981), o qual denomina-a de análise canônica, DUNTEMAN (1984). MANLY (1986) e JOHNSON e WICHERN (1988), dentre outros.

ENGLISH *et alii* (1989) utilizaram a análise de correlação canônica e a análise discriminante na caracterização de microclimas sob a copa de videiras, associados a remoção de folhas e controle de *Botrytis cinerea*. A análise revelou que nenhuma das variáveis climáticas estudadas discriminou isoladamente os microclimas.

2.5.3. Análise Fatorial ("Factor Analysis")

É um método estatístico utilizado para reduzir um grande número de variáveis correlacionadas a um pequeno número de variáveis não-correlacionadas, denominadas fatores, que são variáveis hipotéticas. O propósito essencial é descrever, se possível, as relações de covariância entre as variáveis em termos de fatores, agrupando as variáveis por suas correlações. Isto é, todas as variáveis, dentro de um grupo em particular, são altamente correlacionadas entre si, mas possuem correlações relativamente pequenas com variáveis de um outro grupo. Cada grupo de variáveis representa um fator que é responsável pelas correlações observadas. É possível, então, explicar qual fator ou grupo de Variáveis que mais contribui para a divergência entre os indivíduos.

A sua computação básica possui muita semelhança com a análise de componentes principais, mas os dois métodos possuem diferentes objetivos. Informações sobre análise fatorial são fornecidas por SEAL (1964), MORRISON (1976), FACHEL (1976), MARDIA *et alii* (1979), DUNTEMAN (1984), MANLY (1986), JOHNSON e WICHERN (1988) e SOUZA (1988b).

HEADRICK *et alii* (1990), pela análise fatorial, selecionaram quatro fatores principais que foram usados como variáveis independentes (em substituição as variáveis originais) em uma análise de regressão múltipla, para estudar as relações entre o conteúdo de diferentes tipos de carboidratos do grão e infecção de grãos de milho-doce por *Fusarium moniliforme*.

FRAVEL *et alii* (1983) avaliaram, pela análise fatorial, a função de fatores abióticos na contribuição para o declínio de *Juniperus conferta*, uma vez que não havia evidência de um agente biótico como causa primária do declínio. Seis entre 20 componentes edáficos foram relacionados com o declínio.

JEGER *et alii* (1983) utilizaram a análise fatorial para investigar a independência entre 11 componentes de resistência sobre uma extensa gama de genótipos hospedeiros, no

patossistema trigo - *Septoria nodorum*. A análise indicou que a necrose induzida pelo patógeno e a produção de esporos foram os mais importantes componentes de resistência parcial.

2.5.4. Análise Discriminante

A análise de função discriminante linear é uma técnica que visa a classificação de um objeto ou indivíduo em um dentre dois ou mais possíveis grupos ou populações, previamente definidos. Foi proposta por Fisher em 1936 (SINGH e CHAUDHARY, 1979), cuja idéia foi transformar as observações multivariadas X para observações univariadas Y , tal que os Y s obtidos de duas populações sejam tão separados quanto possível. Fisher sugeriu fazer combinações lineares de X para criar os Y s porque assim eles seriam funções simples de X e facilmente manuseáveis matematicamente (JOHNSON e WICHERN, 1988).

A partir das mensurações dos indivíduos obtém-se um conjunto de combinações lineares (Y_i) das variáveis, as quais são denominadas "funções discriminantes" (DUNTEMAN, 1984) ou "funções discriminantes canônicas" (MANLY, 1986).

A função discriminante "Y" é definida como

$$Y = b_1 X_1 + b_2 X_2 + \dots + b_n X_n ,$$

em que X_1, X_2, \dots, X_n são as variáveis mensuradas e b_1, b_2, \dots, b_n são os coeficientes de ponderação. Os valores b são estimados para que, baseada em valores Y , a razão da variância entre e dentro de populações seja maximizada. Assim, Y poderá ser usado para separar grupos se os valores médios variam, consideravelmente, de grupo para grupo, com os valores Y dentro de um grupo sendo regularmente constante. Assim, Y servirá como melhor discriminante para duas populações que quaisquer caracter isoladamente.

Na análise discriminante, é necessário definir previamente os grupos ou populações e, portanto, conhecer as características dos indivíduos pertencentes a cada grupo. Assim, com base no conhecimento prévio de um grande número de indivíduos, podem-se obter, com precisão, as estimativas das médias dos grupos para cada característica, além da estimativa da matriz de variância e covariância destes dados. Estas prévias informações devem ser confiáveis, caso contrário não é possível afirmar com segurança que a classificação posterior de um indivíduo é correta. Com base nessas informações prévias, é possível construir funções discriminantes que

permitam a máxima discriminação entre os grupos. Estas funções são úteis para alocar com um certo grau de confiança um indivíduo, ou indivíduos, para um dos grupos, dentre os quais eles são conhecidos por pertencer a um deles. O número de funções discriminantes, linearmente independentes, necessário para discriminar entre "k" grupos é $k - 1$.

Após estabelecidas as funções discriminantes e dadas as mensurações das variáveis X_1, X_2, \dots, X_k para cada indivíduo, calculam-se os escores discriminantes. Baseado nos valores dos escores de cada indivíduo em relação as funções discriminantes e comparando-os com os escores médios de cada grupo previamente definido em relação às funções discriminantes, este indivíduo é então classificado em um dos grupos.

A informação sobre a frequência relativa, nas quais os indivíduos de cada grupo são encontrados, pode ser útil na discriminação. Há situações em que a probabilidade de um indivíduo pertencer a um dado grupo é diferente para os diferentes grupos. Por exemplo, se existem dois grupos e se for conhecido que a maioria dos indivíduos, 90%, é alocada no grupo 1 e muito poucos indivíduos, 10%, no grupo 2, a análise poderá considerar essas probabilidades.

Os custos de uma má classificação também poderão ser incorporados na análise. Por exemplo, considerando a existência de dois grupos, se um indivíduo pertencente ao primeiro grupo for classificado incorretamente no segundo grupo e provocar efeito negativo muito maior na interpretação e utilização dos resultados que um indivíduo pertencente ao segundo grupo classificado incorretamente no primeiro, essa diferença deveria ser considerada pela análise discriminante.

A análise discriminante, usualmente, não pode prover um método de classificação livre de erros. Isso ocorre porque pode não haver uma clara distinção entre características mensuradas das populações. Isto é, os grupos podem se sobrepor. A análise permite derivar as proporções esperadas de classificação correta e incorreta para os grupos ou populações de interesse. Uma das limitações do uso da função discriminante para problemas de classificação é a redução de sua efetividade quando mais de dois grupos são considerados.

Segundo JOHNSON e WICHERN (1988). MANOVA pode ser utilizada para testar as diferenças entre os vetores de médias dos diferentes grupos. Embora diferenças aparentemente significantes não impliquem automaticamente em classificação eficiente. o teste é um primeiro

passo necessário, pois caso não sejam encontradas diferenças significantes, a elaboração das **funções** discriminantes poderá provavelmente ser uma perda de tempo.

Segundo MANLY (1986), a análise discriminante e os testes de significância consideram que a matriz de covariância dentro de grupos é a mesma para todos os grupos, ou seja, os grupos **são** amostras de **distribuição multinormal** com diferentes médias e a mesma matriz de covariância. Caso contrário, a análise pode não ser adequada e os testes de significância não serem confiáveis. Segundo MANLY (1986), é possível obter boas discriminações em populações ou grupos que não tenham distribuição **multinormal**. O único problema é o estabelecimento da **significância** dos resultados. Quando a matriz de covariância dentro de grupos não é a mesma para todos os grupos, pode ser mais apropriado utilizar funções discriminantes quadráticas. **Se** a distribuição dos dados difere muito da **distribuição** normal, então um procedimento diferente para discriminação pode ser requerido.

Quando a análise é utilizada para reduzir as dimensões dos dados, ou seja, quando não se conhecem previamente os grupos, ela denomina-se análise de variáveis canônicas (JAMES e McCULLOCH, 1990). Segundo ENGLISH et alii (1989), a análise canônica é uma descritiva forma da análise discriminante. Deve-se dizer que **existe** uma estreita relação entre a análise canônica, distância de Mahalanobis e função discriminante linear.

A análise utiliza variáveis quantitativas contínuas, porém, segundo JOHNSON e WICHERN (1988), variáveis qualitativas podem ser utilizadas. Segundo LACHENBRUCH e GOLDSTEIN (1979), pode-se utilizar uma mistura de tipos de variáveis na análise. O método não é afetado pela escala individual das variáveis (MANLY, 1986).

Mais informações sobre a análise discriminante são fornecidas por RAO (1952 e 1973), MORRISON (1976), MARDIA et alii (1979), SINGH e CHAUDHARY (1979), ARNOLD (1981), DUNTEMAN (1984), MANLY (1986) e JOHNSON e WICHERN (1988), dentre outros.

A aplicação dessa técnica em fitopatologia é incomum. HINDORF (1973), pela análise discriminante, diferenciou, significativamente, a população de *Colletotrichum* spp. em *Coffea* arabica no Quênia, com base em seus caracteres morfológicos, em *C. coffeanum*, *C. acutatum* e *C. gloeosporioides*.

NOE e BARKER (1985) relacionaram 26 componentes edáficos com o padrão de distribuição de nematóides no solo. Utilizaram a análise discriminante para determinar a

adequabilidade das variáveis edáficas para discriminar as parcelas amostrais de solo de acordo com a densidade de parcelas em categorias de alta, média e baixa densidade de nematóides. Discriminações ótimas foram obtidas com seis a oito caracteres do solo. Segundo a análise, o conteúdo de argila e as concentrações de cobre e sódio foram particularmente úteis em explicar a variação espacial entre populações de três espécies de fitonematóides.

BRUGNEROTTO et alii (1991) utilizaram a análise discriminante para a identificação de isolados de *Phytophthora* spp. do cacau, baseada em quatro características, conseguindo separar, com precisão, os isolados pertencentes a cada uma das espécies *P. palmivora* e *P. capsici*.

ENGLISH et alii (1989) caracterizaram microclimas sob a copa de videiras (*Vitis vinifera* L.) com e sem remoção de folhas, comparando-os entre si e o seu efeito no desenvolvimento do mofo cinzento do cacho causado por *Botrytis cinerea*. Foi utilizada a análise discriminante para distinção entre microclimas

2.5.5. MANOVA

O objetivo desta técnica é verificar a existência de diferenças entre tratamentos com relação a várias variáveis simultaneamente.

Nos experimentos mais comuns, cada parcela fornece o valor de uma só variável a ser analisada pelos métodos estatísticos tradicionais. Porém, há casos em que cada parcela fornece observações relativas a duas ou mais variáveis a serem analisadas. Se estas variáveis forem independentes entre si, as análises isoladas de variância de cada variável seriam suficientes. Todavia, segundo HARRIS (1975), a probabilidade de se cometer um erro tipo I (α) aumenta rapidamente com o aumento do número de testes, ou seja, com o aumento do número de variáveis.

Porém, o mais comum para muitos tipos de dados biológicos é que exista correlação (positiva ou negativa) entre as variáveis. Nesse caso, a análise multivariada de variância (MANOVA), isto é, a análise de variância aplicada simultaneamente a todas as variáveis que possam interessar, pode ser muito útil, pois considera a correlação entre variáveis.

MANOVA é um procedimento para testar diferenças entre grupos de acordo com as médias de todas as variáveis, sendo o procedimento aplicado a dados com distribuição normal multidimensional (JAMES e McCULLOCH, 1990). É possível obter resultados não-significativos em testes univariados e resultado significativo em teste multivariado e vice-versa (MANLY, 1986).

MADDEN (1986) e CAMPBELL e MADDEN (1990) sugerem a utilização de MANOVA para comparações de epidemias, utilizando caracteres como nível inicial e final de doença, taxa de progresso de doença e AACPD.

AMORIM e BERGAMIM FILHO (1992) utilizaram MANOVA para comparar variedades de cana-de-açúcar quanto ao progresso do carvão (*Ustilago scitaminea*), considerando, como variáveis, quatro elementos da curva de progresso de doença.

Mais informações sobre MANOVA podem ser obtidas em RAO (1952), SMITH et alii (1962), MORRISON (1976), MARDIA et alii (1979), PIMENTEL GOMES (1984, 1987), DEMÉTRIO (1985), MANLY (1986), JOHNSON e WICHERN (1988) e CRUZ (1990).

2.5.6. Análise de Coordenada Principal

GOWER (1966a) desenvolveu esta técnica na qual é possível calcular os componentes principais a partir de uma matriz de distância euclidiana sem possuir a matriz de dados originais ou a matriz de variância e covariância das características dos indivíduos. A análise de coordenada principal de uma matriz de distância de Mahalanobis é idêntica à análise canônica (GOWER, 1966b).

Este método é também aplicável a matrizes originadas de coeficientes de associação ou outras distâncias que não a euclidiana, porém, em tais casos, a solução somente se aproxima dos componentes principais.

Uma das vantagens desta técnica é a possibilidade de ordenar grupos de indivíduos encontrados na literatura, para os quais a matriz de dados originais não é disponível.

THOMPSON e REES (1979) utilizaram esta técnica a partir de uma matriz de similaridade (estimada com o uso da "distância de Canberra") para ordenarem 45 cultivares de trigo, baseados no progresso da ferrugem do colmo.

CHEN et alii (1992) utilizaram a análise a partir de uma matriz de similaridade, estimada pelo coeficiente de similaridade F (NEI e LI, 1979). no estudo da variação genética de *Xylella fastidiosa*.

2.6. Escolha da Técnica de Análise Multivariada

A técnica de análise multivariada a ser utilizada dependerá do tipo de dados obtidos e dos objetivos do pesquisador. A seguir, serão relacionados alguns possíveis objetivos e as técnicas adequadas aos mesmos.

Quando o objetivo é comparar tratamentos baseados em dados multivariados, deve-se utilizar a MANOVA. Para tanto, os dados devem ser originados de um delineamento experimental.

Quando existem dados adequados a uma análise de regressão múltipla e o objetivo é avaliar os efeitos diretos e indiretos das variáveis independentes na variável dependente, a análise adequada é a "path análise".

A análise de correlação canônica é utilizada para identificar e quantificar as associações entre dois grupos de variáveis, sem fazer distinções entre variáveis dependentes e independentes.

Quando os objetivos são reduzir as dimensões de um conjunto de dados, classificar, ordenar e agrupar indivíduos (ou populações), objetos ou tratamentos, podem-se utilizar a análise de componentes principais, análise canônica, "cluster análise", análise discriminante ou a análise fatorial.

A análise discriminante tem por objetivo classificar novos indivíduos em grupos previamente definidos, o que não faz as análises canônicas, de componentes principais e "cluster análise". A análise fatorial separa as variáveis por suas correlações, em grupos ou fatores, sendo possível então explicar qual o grupo de variáveis que mais contribui para a divergência entre os indivíduos.

Em muitos estudos fitopatológicos foram utilizadas duas ou mais técnicas conjuntamente: análise fatorial e de componentes principais (JEGER et alii, 1983 e JEGER, 1980); "cluster análise" e análise de coordenadas principais (THOMPSON e REES, 1979); "cluster análise" e componentes principais (LALANCETTE et alii, 1987; ZHU et alii, 1988 e CHEN et alii,

1991 e 1992); e análise canônica, "cluster análise" e análise discriminante (HANSEN e HAMM, 1983).

2.7. Influência do Tipo de Dados na Escolha da Técnica de Análise **Multivariada**

Quanto ao tipo de dados (variáveis), será relatado aqui apenas a sua influencia nas análises **canônica**, de componentes principais e "cluster análise".

Serão descritas, a seguir, algumas possíveis situações quanto ao tipo e a origem dos dados.

2.7.1. Variáveis Quantitativas Contínuas

2.7.1.1. Originadas de Delineamento Experimental

A análise de componentes principais, bem como a distância euclidiana, não deve ser utilizada nestas situações, como realizado por JEGER (1980). JEGER et alii (1983), ANDERSON et alii (1990) e SCHILDER e BERGSTROM (1990). pois confundiriam as fontes de variação dentro e entre amostras (JAMES e McCULLOCH, 1990).

É indicada a utilização das análises **canônica** e, ou, "cluster análise". utilizando a distância de **Mahalanobis**, pois ambas consideram a covariância residual. Para tanto, todas as variáveis devem ser mensuradas em cada parcela experimental.

Quando no estudo de progresso de doença, cada época de avaliação de doença é considerada como uma variável, **C** inadequado o **uso** das análises **que** consideram a covariância ou correlação, como a análise de componentes principais, ou que consideram a covariância residual entre variáveis, como a análise canônica e "cluster análise" utilizando a distância de **Mahalanobis**. Essa afirmação também pode ser aplicada para o estudo de variabilidade de isolados de **patógenos** em **que** o grau de reação de cada uma das variedades a um dado isolado **fitopatogênico** é considerado uma variável. Uma indicação é o **uso** de "cluster análise" utilizando a distância **euclidiana**, como realizado por REGO (1993) e SCHILDER e BERGSTROM (1990). e, ou, a "distância de Canberra".

Para dados constituídos de razões, proporções ou percentagens deveria ser considerada a transformação arco **seno** raiz quadrada (PIMENTEL GOMES, 1984), assim poder-se-ia tomar sua distribuição mais próxima da normal multivariada, para perfazer a análise canônica e, ou, estimar a distância de Mahalanobis, ou perfazer a análise de componentes principais, caso os dados não provenham de um delineamento experimental.

2.7.1.2. Não-Originadas de Delineamento Experimental

Nesta situação não há repetições, ou caso houver, trabalha-se com as médias. Utiliza-se a análise de componentes principais (**neste** caso, todas as variáveis devem ser mensuradas em cada parcela experimental) e, ou, "cluster análise" baseada na distância euclidiana, ou em uma distância similar (distância que não considera variâncias e covariâncias). Pode-se usar também, segundo SOKAL (1974), um coeficiente de correlação.

Em algumas situações os dados podem ser transformados para uma variável qualitativa binária (SNEATH e SOKAL, 1973), calculando-se então um coeficiente de similaridade que requer variáveis binárias e procedendo "cluster análise". SILVEIRA NETO (1986) relata fórmulas para essa transformação. Porém, nesse caso, perde-se informação.

Em alguns estudos, as variáveis necessitam ser agrupadas. Para avaliar a similaridade entre variáveis pode ser utilizada alguma medida de distância ou o coeficiente de correlação linear (eq. 3.2b - **seção** 3.5.), sendo considerado o valor absoluto das correlações no agrupamento. para variáveis quantitativas.

2.7.2. Variáveis Quantitativas Discretas

Para dados constituídos de contagens, poderia ser considerada a utilização da transformação raiz quadrada (PIMENTEL GOMES, 1984) para tornar a sua distribuição mais apropriada a uma análise de componentes principais, análise canônica ou "cluster análise", utilizando a distância de Mahalanobis.

2.7.3. Variáveis Qualitativas Binárias

Os caracteres binários por conveniência são codificados em 0 e 1, que podem representar a ausência ou presença de certa característica, respectivamente; uma reação bioquímica negativa ou positiva, respectivamente; etc. Os pares de indivíduos são comparados com base nos códigos 0 e 1. Indivíduos similares possuem mais características em comum que indivíduos dissimilares. Poderia ser considerada mais indicada neste caso a utilização de "cluster análise" baseada em um coeficiente de associação (ver seção 2.4.1.2.1.)

Segundo JOHNSON e WICHERN (1988), embora a distância euclidiana possa ser usada para avaliar similaridade baseada em variáveis binárias, ela fornece uma contagem do número de caracteres discordantes (0 - 1 e 1 - 0) entre dois indivíduos e atribui pesos iguais aos caracteres concordantes (1 - 1 e 0 - 0), sendo inadequada quando for necessário atribuir pesos diferentes aos caracteres concordantes "positivos" (1 - 1) e "negativos" (0 - 0). Para permitir um tratamento diferenciado das concordâncias 1 - 1 e 0 - 0, pode ser utilizado o coeficiente de Jaccard (ver seção 2.4.1.2.1.). JOHNSON e WICHERN (1988) citam diversas medidas de similaridades com este tratamento diferenciado.

Segundo JAMES e McCULLOCH (1990), os procedimentos que utilizam combinações lineares de variáveis em algum padrão, como as análises canônica, fatorial, de componentes principais e de função discriminante linear, são apropriados quando se objetiva interpretar combinações lineares ótimas de variáveis. Na aplicação de métodos lineares, assume-se que os valores das variáveis aumentam ou decrescem regularmente e que não existem interações. Se as variáveis não são linearmente relacionadas, as variâncias e covariâncias são pobres sumários estatísticos, pois não revelam as associações não-lineares entre variáveis. Portanto, pode não ser adequada a utilização dessas técnicas com variáveis qualitativas, como realizado por LEBEDA e JENDRULEK (1987b), HILL et alii (1989) e CHEN et alii (1991). As variáveis qualitativas e ordinais são manuseadas mais eficientemente com modelos não-lineares ou então, deveriam sofrer transformações para que a relação entre elas seja, no mínimo, aproximadamente linear.

Em alguns estudos, as variáveis necessitam ser agrupadas. Para avaliar a similaridade entre variáveis qualitativas binárias, pode-se utilizar algum coeficiente de associação. citado por SNEATH e SOKAL (1973) e JOHNSON e WICHERN (1988).

2.7.4. Variáveis Qualitativas Multicategóricas

Estes caracteres não podem ser ordenados, podendo ser codificados por diferentes símbolos ou letras, como A, B, C, ..., para cada categoria. Alguns autores, como REGO (1993), utilizam a distância euclidiana para avaliar dissimilaridade entre indivíduos baseados neste tipo de dado, codificando cada categoria com um número, por exemplo, 1, 2, 3, 4 etc. Esse procedimento é incorreto, pois, em decorrência do processo matemático, resulta em uma distância euclidiana maior do indivíduo classificado na categoria A (ou 1) em relação ao indivíduo classificado na categoria C (ou 3), que em relação ao indivíduo classificado na categoria B (ou 2), considerando apenas uma variável. Assim, considerando como variável a cor da colônia fúngica em um estudo de taxonomia, não existem condições de afirmar que a diferença entre a categoria com cor branca e a com cor verde é maior ou menor que a diferença entre a categoria com cor branca e a com cor cinza. Analogia pode ser feita para a análise de isoenzimas de isolados de fungos, quando para uma dada enzima (variável) existem mais de dois fenótipos isoenzimáticos. Nesse caso, a utilização da distância euclidiana pode gerar falsos resultados, quando se considera cada fenótipo como uma categoria.

Segundo SNEATH e SOKAL (1973), não é apropriado estimar quaisquer medidas de similaridade que assumam variáveis contínuas, como a distância de Mahalanobis, a partir de variáveis qualitativas multicategóricas.

Para o processamento, as variáveis qualitativas multicategóricas podem ser tratadas como binárias. Para uma dada característica (variável), quando um mesmo símbolo (categoria) coincide para dois indivíduos, uma concordância (1 - 1) é assinalada, caso contrário, uma discordância (1 - 0) é assinalada. Um outro método é converter o carácter qualitativo multicategórico em alguns novos caracteres binários, sendo o número destes novos caracteres igual ao número de categorias do carácter multicategórico original. SNEATH e SOKAL (1973) discutem a viabilidade desta aplicação, bem como outros modos de codificar e processar os caracteres qualitativos multicategóricos.

2.7.5. Variáveis Ordinais

Segundo SNEATH e SOKAL (1973), quando os caracteres multicategóricos são ordenados, eles podem ser processados como caracteres quantitativos.

2.7.6. Mistura de Variáveis Quantitativas e Qualitativas

Segundo JOHNSON e WICHERN (1988), existe muito pouca teoria disponível para o caso em que algumas variáveis são quantitativas contínuas e outras são qualitativas. No caso da função discriminante linear, a eficiência da análise depende da correlação entre variáveis contínuas e qualitativas; uma baixa correlação em um grupo de indivíduos, mas uma alta correlação em outro grupo ou uma alteração no sinal da correlação entre dois grupos poderão indicar condições desfavoráveis para a análise. Essa afirmação também pode ser aplicada a "cluster análise" baseada na distância de Mahalanobis e à análise canônica. IEZZONI e PRITTS (1991) afirmam que o uso simultâneo de variáveis qualitativas e quantitativas na análise de componentes principais não é apropriado.

Segundo SNEATH e SOKAL (1973), no caso de haver variáveis qualitativas binárias e quantitativas, pode-se codificar as variáveis quantitativas para 0 e 1 e utilizar uma medida de similaridade como previamente relatado.

GOWER (1971) propôs um coeficiente de similaridade para as situações em que há mistura de variáveis quantitativas e qualitativas.

2.8. Padronização e "Peso" de Variáveis

Para caracteres quantitativos com distribuição normal multidimensional, sua informação é proporcional à variância. Se uma variável mensurada possuir maior variação que outras, então esta poderá dominar o cálculo da medida de distância (ou dos componentes principais). Na prática é desejável que todas as variáveis tenham a mesma influência no cálculo da distância. Assim, se as variâncias são igualadas, por padronização, então cada caracter contribui com igual informação. Embora a padronização equalize o tamanho de diferentes caracteres, ela não remove

a diferença de tamanho entre indivíduos (SNEATH e SOKAL, 1973). Não é necessária a padronização de variáveis binárias, porém é recomendada para variáveis ordinais e quantitativas, quando se procede a análise de componentes principais ou quando se estima a distância euclidiana. Não é necessária a padronização de variáveis na estimativa da distância de Mahalanobis e na análise canônica.

Em uma classificação em que todas as variáveis possuam o mesmo peso, a importância da variável é baseada em variância e covariâncias. A distância de Mahalanobis e a análise canônica já impõem peso às variáveis durante o processo matemático. A variável com maior variabilidade amostral (quadrado médio do resíduo) possui menor peso nas análises.

Para taxonomistas, peso indica que uma característica taxonômica possui maior significado para os propósitos de sua classificação que outras, ou seja, o carácter é mais estável taxonomicamente (ROGERS, 1963). Em muitos estudos pode o pesquisador querer dar, artificialmente, pesos diferentes às variáveis. SNEATH e SOKAL (1973) discutem este assunto, alertando que se pesos diferenciados são admitidos, critérios exatos necessitam ser fornecidos para estimá-los, e se não for possível decidir quanto peso dar ao carácter, deve-se então dar pesos iguais, evitando alocar pesos irracionalmente.

PALMER et alii (1987), estudando isolados de *Sphaeropsis sapinea*, deram pesos aleatórios de 1/6, 1/6 e 2/3, respectivamente, para as características de crescimento em diferentes meios de cultura, em diferentes temperaturas e dimensões de esporos, no cálculo da distância euclidiana entre os isolados.

2.9. Descarte de Variáveis

Uma das questões básicas no estudo da divergência entre indivíduos ou populações diz respeito ao número e tipo de características a serem avaliadas, pois a maior ou menor extensão de similaridade depende não somente das características (variáveis) avaliadas, como também do seu número. Não existem bases teóricas para determinar o número de características a serem medidas. Tem sido relatado, no melhoramento vegetal, que os caracteres importantes para adaptação e seleção natural são mais apropriados e devem ser escolhidos para os estudos de divergência e agrupamento (CRUZ, 1990 e ARUNACHALAM, 1981). As variáveis escolhidas

devem ser relevantes para a classificação esperada (MANLY, 1986). Segundo ADAMS e WIERSMA (1978), os caracteres a serem preservados na análise de divergência deverão ser apenas aqueles que representam a estrutura fundamental do sistema biológico que está sendo estudado, devendo ainda serem suficientemente diversos para representar, no mínimo, as dimensões mais importantes do sistema.

Em muitas situações os pesquisadores tenderão, naturalmente, a avaliar maior número de caracteres, gerando, com isso, acréscimo considerável de trabalho. Porém, após a análise inicial, pode ser possível identificar somente os caracteres mais importantes, aumentando, assim, a eficiência do método em aplicações futuras. Quando o número de caracteres utilizados toma-se elevado, é possível que muitos deles pouco contribuam para a discriminação dos indivíduos avaliados, por serem relativamente invariantes entre estes ou por serem redundantes em virtude de serem altamente correlacionados com outros caracteres. Essa situação apresenta como consequência aumento no trabalho de caracterização, sem melhoria na precisão, além de tornar mais complexa a análise e interpretação dos dados. Portanto, a eliminação dos caracteres redundantes e de difícil mensuração torna-se desejável, a fim de facilitar o estudo de divergência, reduzindo tempo e custos da experimentação (PEREIRA, 1989). A redução no número de caracteres, com a eliminação daqueles que menos contribuem para a divergência, deve facilitar as interpretações sem causar perda considerável de informação.

Segundo CRUZ (1990), caracteres dispensáveis em estudos de divergência genética são aqueles relativamente invariantes entre os indivíduos estudados, são fortemente afetados pelo ambiente, apresentam instabilidade com a mudança ambiental ou são redundantes por estarem correlacionados com outros caracteres.

As variáveis selecionadas e descartadas devem apresentar correlações significativas entre si, ou seja, as variáveis descartadas devem ser redundantes (ser responsáveis pelo mesmo tipo de informação já contida nas variáveis selecionadas). Por outro lado, as variáveis selecionadas devem ter baixas correlações entre si. Deste modo, cada variável será responsável por um tipo de informação biológica exclusiva e a ação conjunta das mesmas será complementar para a descrição geral dos indivíduos ou populações estudados.

SNEATH e SOKAL (1973) discutem os problemas relativos ao descarte de variáveis (caracteres) com baixa estabilidade ambiental (baixo controle genético), a redundância de informação e a correlação entre caracteres.

Diversas proposições sobre o descarte de variáveis têm sido apresentadas por BEALE et alii (1967), JOLLIFFE (1972 e 1973) e HUSSAINI et alii (1977). Na análise de componentes principais, tem-se adotado o critério de avaliação da importância dos caracteres a partir dos coeficientes de ponderação (autovetores) das variáveis associados aos últimos componentes que, por estimação, retêm proporção mínima da variação total (MARDIA et alii, 1979). Baseado no princípio que a importância ou variância dos componentes principais decresce do primeiro para o último, tem-se que os últimos componentes são responsáveis pela explicação de uma fração muito pequena da variância total. Assim, a variável que domina (aquela que possui o maior coeficiente) o componente principal de menor autovalor (menor variância) deve ser a menos importante para explicar a variância total e, portanto, passível de descarte (PEREIRA, 1989). A razão é que variáveis altamente correlacionadas com os menores componentes principais representam variação praticamente insignificante (DUNTEMAN, 1984). A seguir, o próximo componente de menor autovalor é considerado, podendo se descartar a variável com o maior coeficiente (valor absoluto) no componente e a qual não tenha sido previamente descartada (MARDIA et alii, 1979). BEALE et alii (1967) propuseram em essência esse método de descarte, considerando que apenas seriam descartadas variáveis de componentes cujo autovalor fosse inferior ou igual a um determinado valor a ser definido. JOLLIFFE (1972 e 1973) avaliou, em dados simulados e reais, oito métodos de descarte de variáveis, em que dois estavam relacionados com regressão múltipla, quatro com componentes principais e dois com técnicas de conglomeramento, concluindo que os métodos possuíam eficiência equivalente em rejeitar variáveis. No caso de se utilizarem componentes principais, os resultados serão mais satisfatórios se o número de variáveis rejeitadas for igual ao número de componentes cuja variância é inferior a 0.7. Este critério é estabelecido para os casos em que os dados são padronizados e assim sendo, estimam-se os autovalores a partir da matriz de correlações, e sua soma corresponde ao total de variáveis analisadas.

Outra maneira, largamente utilizada pelos melhoristas, para avaliar a importância de um grupo de características incluídas na avaliação da divergência genética é a frequência de "rank 1",

que cada característica recebe quando proporciona a maior magnitude do quadrado da diferença entre médias das características não-correlacionadas e padronizadas Z_j (CRUZ, 1987), quando se estima D^2 (distância de Mahalanobis) entre as amostras ou observações i e k , pela fórmula:

$$D^2_{ik} = \sum_{j=1}^p (Z_{ij} - Z_{kj})^2$$

O valor 1 é dado a variável que proporciona a mais alta contribuição e o valor p é dado a característica que proporciona a mais baixa contribuição. Dessa forma, a contribuição de cada variável (C_j) para a divergência genética é expressa por

$$C_j = \frac{(\text{número de "ranks" 1 da } j\text{-ésima variável})}{g(g-1)/2} \times 100$$

em que g é o número de genótipos avaliados.

Segundo SINGH (1981), esse método é falho pois associa incorretamente as variáveis não-correlacionadas (padronizadas) com as características originais, além daquelas não possuírem qualquer significado biológico. O autor apresenta um método alternativo para a estimação da contribuição percentual de cada variável, no valor total de D^2 , obtida pelo somatório das distâncias de cada par de genótipos, pelo acréscimo de cada característica na análise.

CAMPBELL e ATCHLEY (1981) discutem a importância das variáveis na análise canônica. Segundo os autores, o método mais comumente utilizado é baseado na magnitude dos coeficientes da variável canônica para as variáveis padronizadas (Z_j) com desvio padrão residual unitário. Variáveis com maior valor absoluto são consideradas mais importantes, podendo aquelas com pequeno coeficiente serem eliminadas, porém, os autores alertam para a possível inadequação deste procedimento quando algumas das variáveis possuírem alta covariância residual entre si. Segundo SINGH (1981), deve-se considerar, na avaliação da importância dos caracteres, na análise canônica, o valor obtido pelo produto entre o desvio padrão residual do carácter e o respectivo elemento do autovalor.

SEDIYAMA et alii (1989) avaliaram a importância relativa de cada um entre 12 caracteres na diferenciação fenotípica de 16 genótipos de soja, pelo coeficiente de correlação entre cada carácter e o primeiro componente principal. Este método não apresentou consistência

entre dois experimentos realizados. Deve-se considerar que a eficiência deste método estaria condicionada ao primeiro componente reter uma alta percentagem da **variância** original total.

Segundo SNEATH e SOKAL (1973), a análise fatorial possui importantes implicações na escolha de caracteres, a qual foi utilizada por FERNANDES e LIMA (1991) para a identificação das variáveis de maior importância na caracterização de sistemas de produção de leite, procedendo, em seguida, "cluster análise", utilizando **os** escores fatoriais, a qual forneceu a estratificação das propriedades em grupos **distintos**.

HANSEN e HAMM (1983) e ENGLISH et alii (1989) utilizaram a análise discriminante para avaliar a importância de variáveis.

HUSSAINÍ et alii (1977) utilizaram uma outra metodologia para seleção prévia de variáveis. No estudo de diversidade genética de **640** genótipos de **Eleusine coracana** (L.) Gaertn, avaliados em um delineamento em blocos ao acaso com duas repetições, **os** dados dos genótipos foram sujeitos á análise de variância separadamente para cada uma das 29 características consideradas inicialmente, Valores estimados dos componentes de variância (σ^2_{gen} - variância genética entre acessos; σ^2_b - **variância** devido a bloco; e σ^2_e - variância devido ao erro experimental) foram calculados para cada caracter. Um baixo valor para a razão $\sigma^2_{gen} / (\sigma^2_e + \sigma^2_b)$ indica que as diferenças entre tratamentos são decorrentes mais dos efeitos ambientais e das interações. Foram descartados 11 caracteres cujos valores para a razão proposta foram menores que a unidade.

3. MATERIAL E MÉTODOS

3.1. Aplicação de "Cluster Análise" no Estudo do Progresso da Ferrugem (*Tranzschelia sp.*) em Cultivares de Pessegueiro (*Prunus persica*)

O trabalho foi realizado no pomar experimental do Departamento de Fitotecnia da Universidade Federal de Viçosa (MG), de outubro de **1988** a fevereiro de **1989**.

3.1.1. Material Experimental

Utilizaram-se plantas adultas, que não recebiam tratamento fitossanitário há **três** anos, dos seguintes cultivares: 'Cerrito', 'Diamante', 'Flordasun', 'Marli', 'Pérola de Itaquera', 'Premier', 'Princesa', 'Rei da Conserva', 'Sinuelo', 'Topázio' e 'Vila Nova'.

3.1.2. Delineamento Experimental

Utilizou-se o delineamento em blocos casualizados, com duas repetições, sendo a parcela representada por duas plantas.

Em cada avaliação coletaram-se **20** folhas maduras na porção mediana da copa de cada planta, a altura do terceiro e quarto internódios, considerando **os** quatro pontos cardeais na coleta.

3.1.3. Avaliação da Doença

Foram realizadas **11** avaliações de severidade de doença, a intervalos quinzenais.

Nas folhas coletadas, avaliou-se a severidade da ferrugem, em condições de laboratório. Para tanto, utilizou-se de escala de notas variando de **1** a **5**, baseada no número de pústulas por folha (Quadro 1).

3.1.4. Análise de Dados

Os valores do grau de doença foram transformados em índice de doença de McKinney (ID):

$$ID = \frac{\sum_{i=1}^n GD_i}{n \times GD_{\text{máximo}}}$$

em que **n** é o número de folhas e **GD** é o grau de doença.

Procedeu-se ao agrupamento dos cultivares quanto a intensidade de ferrugem por "cluster análise", utilizando como medidas de similaridade a "distancia de Canberra" e a distância euclidiana média e como métodos de agrupamento, o de Tocher e o do vizinho mais próximo. Cada época de avaliação foi considerada uma variável.

QUADRO 1 - Escala de Avaliação da Severidade da Ferrugem do Pessegueiro

Número de Pústulas / Folha	Grau de Doença (GD)
6 - 25	2
26 - 125	3
126 - 625	4
625 ou mais	5

Compararam-se os resultados obtidos por "cluster análise" com os resultados dos métodos tradicionalmente utilizados para comparação de epidemias (MADDEN, 1986): a área abaixo da curva de progresso de doença (AACPD); e a análise de variância e comparação de médias dos IDs entre cultivares, considerando o experimento como tendo um delineamento em parcelas subdivididas, sendo o fator tempo (época de avaliação) considerado como subparcela.

Calcularam-se os valores da AACPD, segundo o método de SHANER e FINNEY (1977) modificado:

$$AACPD = \sum_{i=1}^{n-1} \{ [(ID_i + ID_{i+1}) / 2] - C \} (t_{i+1} - t_i),$$

em que

ID_i é o índice de doença na i-ésima (i = 1, 2, 3, ..., n = 11) avaliação,

t é o tempo, em dias, decorrido entre a primeira e a i-ésima avaliação e

C é o fator de correção, correspondendo ao D₀ = 0,2 (devido a escala de avaliação de doença -

Quadro 1 - não possuir o grau zero, quando os graus de doença são transformados para ID, o menor valor possível de se obteré D₀ = 0.2. Isso implicaria na existencia de um valor adicional de AACPD, sem sentido biológico, para todos os cultivares, caso não houvesse o fator de correção).

Para uma comparação adicional, obtiveram-se os modelos estatísticos para representar as epidemias em cada cultivar. Esses modelos foram submetidos ao teste de identidade de modelos descritos por GRAYBILL (1976) e REGAZZI (1993).

A AACPD foi obtida por meio do programa AVACPD (TORRES e VENTURA, 1991). As análises de variância, a obtenção dos modelos lineares, bem como parte dos cálculos para o teste de identidade de modelos foram realizados utilizando o programa SAEG (UNIVERSIDADE FEDERAL DE VIÇOSA, s.d.).

3.2. Aplicação de Análise **Canônica** e "Cluster Análise" no Estudo da Resistência Horizontal de **Genótipos** de *Coffea arabica* a Ferrugem (*Hemileia vastatrix*)

3.2.1. Material Experimental

Foram selecionadas uma linhagem de 'Mundo Novo', uma de 'Catuaí' e seis genótipos de *C. arabica* em geração F_1 (Quadro 2), que apresentavam-se promissores para os caracteres agronômicos sob estudo no programa de melhoramento do cafeeiro desenvolvido pelo Departamento de Fitopatologia da Universidade Federal de Viçosa.

3.2.2. Técnica de Inoculação

Foram inoculadas mudas de cafeeiros provenientes de multiplicação por semente, no estágio de desenvolvimento de três a seis pares de folhas definitivas, sendo inoculado o último par de folhas **que** apresentou desenvolvimento completo, porém de aspecto tenro.

As inoculações foram realizadas no dia **16** de março de **1992**, pelo método de D'OLIVEIRA (s.d.) modificado. o qual consiste na distribuição dos uredosporos com um pincel de "pêlo de camelo" sobre a face dorsal ou inferior das folhas a serem inoculadas (FONSECA, **1979** e ALMEIDA, **1980**). Após a aplicação dos uredosporos a seco na superfície foliar, as folhas inoculadas foram levemente atomizadas com água destilada e as mudas foram levadas para a câmara de nevoeiro (22 ± 2 °C, aproximadamente 100% U.R.). Após **48** horas no escuro, as mudas permaneceram na sala de incubação, sob luz artificial a 22 ± 2 °C por **13** dias; após, foram transferidas para casa de vegetação, onde permaneceram todo o período de avaliação.

Para a padronização da quantidade de uredosporos aplicada em cada planta foi utilizado um estilete com um recipiente em sua extremidade. com capacidade para conter uma quantidade conhecida de uredosporos, sendo seu conteúdo utilizado para inocular um par de folhas. Estimou-se a quantidade de inóculo pela suspensão em um volume de 20 ml de água, de 20 amostras do volume de uredosporos do recipiente, sendo realizada, para cada amostra, cinco contagens do número de uredosporos em gotas de **5** μ l, ao microscópio óptico comum, obtendo-se a média de $22,7 \times 10^4$ uredosporos e desvio padrão de $6,7 \times 10^4$ uredosporos.

QUADRO 2 - Identificação dos Genótipos em Geração F1 e Linhagens de 'Catuaí' e 'Mundo Novo' Utilizados no Estudo de Resistência Horizontal a Ferrugem

Genótipos	Progenitores
H415-2	2143-193 EL7 ¹ X UFV 440-22 CA ²
H442-1	2150-158 EL8 X UFV 444-01 CA
H471-4	2143-195 EL7 X UFV 319-04 CA
H472-2	2150-158 EL8 X UFV 319-04 CA
H481-2	2190-305 EL6 X UFV 437-09 CA
H543-1	2190-048 EL8 X UFV 378-33 CA
UFV 2143	'Catuaí Amarelo' - LCH 2077-2-5-30
UFV 2150	'Mundo Novo' - LCMP 376-4-32

¹ Registro do número da linhagem, cova e ensaio de linhagens de 'Catuaí' (EL7) ou 'Mundo Novo' (EL8). UFV 2190 corresponde ao 'Mundo Novo' LCMP 464-18.

² Número da **introdução** e cova no campo de adaptação e seleção. UFV 319 corresponde a uma introdução do CIFC, cujo progenitor é portador do fator S_H4. Os demais correspondem a seleções de Híbrido de Timor.

3.2.3. Raça de *H. vastatrix* Utilizada

Foi utilizada a raça 11, que é a raça mais comum no Brasil e no mundo (RODRIGUES JR. et alii, 1975; RODRIGUES JR., 1983 e CARDOSO, 1986), possui o fator de virulência v5 e permite a detecção do grupo fisiológico E. A raça utilizada pertence ao banco de raças de *H. vastatrix* do Departamento de Fitopatologia da UFV.

3.2.4. Delineamento Experimental

O experimento foi inteiramente casualizado, constando de oito tratamentos e cinco repetições, sendo cada parcela constituída por dez mudas oriundas de sementes coletadas de plantas cujas origens são apresentadas no Quadro 2.

Uma semana após, foram inoculadas 40 mudas de cada genótipo, sendo avaliado posteriormente somente o grau de doença (ver item 3.2.6.).

3.2.5. Avaliação dos Componentes de Resistência

3.2.5.1. Período de Incubação

Foi expresso pelo tempo médio, em dias, decorrido entre a inoculação e o aparecimento dos primeiros sintomas visíveis da doença nas folhas inoculadas (PARLEVLIET, 1979); avaliado pela observação visual das folhas inoculadas, a partir do 12^o dia de inoculação, diariamente, até a constatação dos primeiros sintomas.

3.2.5.2. Período Latente

Foi expresso pelo tempo, em dias, desde a inoculação até a esporulação de qualquer lesão.

3.2.5.3. Esporulação

Coletaram-se uredosporos das folhas aos 36, 50 e 71 dias após a inoculação. Para a sua quantificação, foram feitas suspensões em um volume de 5 ml e realizadas cinco contagens em gotas de 5 μ l e, quando necessário, diluiu-se o volume inicial de cinco a 40 vezes. Considerou-se a soma das três coletas de uredosporos e em algumas mudas foram realizadas apenas duas coletas por causa da queda das folhas inoculadas. Foi expresso o número médio de uredosporos por cm^2 de área foliar.

A área foliar foi estimada por meio de uma equação obtida por regressão da área foliar em função do produto do comprimento e da largura da folha, a partir de dados de 50 folhas coletadas de mudas (Quadro 3).

3.2.5.4. Severidade

Foi expressa pela percentagem de área foliar lesionada. Na sua estimação, utilizou-se a escala de KUSHALAPPA (1978).

QUADRO 3 - Equações de Regressão' para Estimativa da Área de Folhas de Genótipos de *Coffea arabica* e Linhagens de 'Catuai' e 'Mundo Novo'

Genótipo	a	b**	R ² (%)	X (intervalo)
H415-2	46,096	0,6290	97.63	2070 - 6210
H442-1	116,501	0.61 12	96.96	2706 - 8580
H471-4	97,906	0,6176	97.58	2584 - 6832
H472-2	14,340	0,6320	98,58	2400 - 8906
H481-2	34,467	0,6213	97.17	2130 - 8100
UFV 2143	- 30,475	0,6468	98,95	3139 - 9261
UFV 2150	262,206	0,5391	92.96	3150 - 9338

• O modelo de regressão ajustado foi do tipo $Y = a + bX$, em que Y = área foliar (mm^2) e X = produto entre comprimento e largura foliar.

** Todos os coeficientes foram significativos pelo teste t ($P < 0,001$)

3.2.6. Avaliação do Grau de Doença

Para avaliação do grau de doença foi utilizada a escala de TAMAYO (1988), a qual foi baseada nas escalas de D'OLIVEIRA (s.d.), ESKES (1983) e ABREU (1988) (Quadro 4).

3.2.7. Análise de Dados

Para a análise dos componentes de resistência, foram utilizadas as seguintes técnicas de análise multivariada: análise canônica e "cluster análise".

Estimaram-se os intervalos de confiança (IC) do grau médio de doença de cada genótipo, avaliado aos 62 dias após a inoculação. O IC foi calculado pela equação

$$X - t_0 s / n^{1/2} < \mu < X + t_0 s / n^{1/2}$$

sendo μ = grau médio de doença no genótipo; x = estimativa de μ ; s = desvio-padrão do grau de doença no genótipo; n = número de plantas avaliadas por genótipo; e $t_0 = t_{\alpha(n-1)}$ estatística t associada a $n-1$ graus de liberdade e nível de significância α (PIMENTEL GOMES, 1987). Adotou-se $\alpha=0,05$.

QUADRO 4 - Escala Proposta por Tamayo (1988) para Avaliar a Resistência do Cafeeiro a *Hemileia vastatrix*

Grau de Doença	Descrição
1	Ausência de sintomas.
2	Manchas cloróticas pequenas (até 1 mm de diâmetro). Reação de hipersensibilidade, sem esporulação.
3	Manchas cloróticas medianas (1 - 3 mm de diâmetro). Sem esporulação.
4	Manchas cloróticas medianas (1 - 3 mm de diâmetro). Poucos uredosporos, menos de 25% de lesões esporulantes.
5	Manchas cloróticas grandes (maiores que 3 mm de diâmetro). Presença nítida de uredosporos, 25-50% de lesões esporulantes.
6	Manchas cloróticas grandes (maiores que 3 mm de diâmetro). Uredosporos abundantes, mais de 50% de lesões esporulantes.

3.3. Avaliação da Resistência de Progênes de 'Catimor' a Oito Raças Fisiológicas de *Hemileia vastatrix*

Utilizaram-se os dados obtidos por TAMAYO (1988) para agrupar progênes de 'Catimor' (a identificação das progênes é fornecida por TAMAYO, 1988) em grupos semelhantes quanto a resistência a oito raças de *H. vastatrix*. Também foram agrupadas as raças quanto as reações das progênes. Utilizou-se, para tanto, "cluster análise", baseada na distância euclidiana média e nos métodos de agrupamento do "vizinho mais próximo" e de Tocher, considerando o grau médio de doença das progênes, relativos a cada raça de *H. vastatrix*.

3.4. Utilização da Análise de Componentes Principais e "Cluster Análise" no Estudo da Resistência Horizontal de Progênes de 'Catimor' a *Hemileia vastatrix*

Utilizaram-se, os dados obtidos por ALMEIDA (1980) em seu estudo da resistência horizontal de progênes de 'Catimor' a ferrugem, o qual inoculou, com uma mistura de raças, 11 progênes segregantes para a susceptibilidade: UFV 1307, UFV 1340, UFV 3684, UFV 3686, UFV 3687, UFV 2861, UFV 2862, UFV 2863, UFV 3658, UFV 4303 e UFV 4305. Utilizaram-se 40 mudas por progênes. O cultivar 'Catuaí' (UFV 2144) foi utilizado como testemunha. A inoculação foi realizada segundo a técnica de D'OLIVEIRA (s.d.) modificada (ver seção 3.2.2.).

As progênes foram avaliadas quanto ao período de incubação (PI) e período latente médio (PLm), severidade (SEV) (ALMEIDA, 1980, denominou de período latente de infecção, período de geração e intensidade de ferrugem, respectivamente) e número de lesões esporuladas por folha (NPF). As definições de PI e SEV foram descritas na seção 3.2.5. PLm foi considerado o tempo em dias, entre a inoculação e a esporulação de 50% das lesões esporulantes. As médias das progênes em relação aos componentes de resistência podem ser consultadas em ALMEIDA (1980).

Agruparam-se as progênes mais semelhantes quanto as quatro características avaliadas pela análise de componentes principais, com os dados padronizados (com variância igual a unidade), e por "cluster análise", utilizando a distância euclidiana média, com os dados padronizados (com média igual a zero e variância igual a unidade).

3.5. Análises Multivariadas

3.5.1. Análise de Componentes Principais

Seja X_{ij} a média original e seja Z_{ij} a média padronizada da j -ésima característica ($j = 1, 2, \dots, p$) avaliada no i -ésimo indivíduo ($i = 1, 2, \dots, n$), $Z_{ij} = X_{ij} / \sigma_j$, em que X_{ij} são valores de cada indivíduo para uma dada característica. e σ_j o desvio padrão da j -ésima característica. Seja R a matriz de covariâncias entre essas características padronizadas (ou matriz de correlações entre as características com base nos dados originais). Tem-se que

$$\text{COV}(X_j; X_j) = \frac{\left[\sum_{i=1}^n (X_{ij} X_{ij}) - \left(\sum_{i=1}^n X_{ij} \sum_{i=1}^n X_{ij} \right) / n \right]}{n - 1} \quad (3.1a)$$

$$\text{COV}(X_j; X_j) = [\sigma^2(X_j + X_j) - \sigma^2(X_j) - \sigma^2(X_j)] / 2 \quad (3.1b)$$

$$r(X_j; X_j) = \text{COV}(Z_j; Z_j) = \frac{\text{COV}(X_j; X_j)}{\sqrt{\sigma^2(X_j) \sigma^2(X_j)}} \quad (3.2a)$$

ou de maneira equivalente:

$$r_{jj} = \frac{\sum_{i=1}^n (X_{ij} - \bar{X}_j) (X_{ij} - \bar{X}_j)}{\sqrt{\sum_{i=1}^n (X_{ij} - \bar{X}_j)^2 \sum_{i=1}^n (X_{ij} - \bar{X}_j)^2}} \quad (3.2b)$$

em que

$\text{COV}(X_j; X_j) = \text{COV}(j, j)$ e $\text{COV}(Z_j; Z_j)$ são estimadores da covariância entre duas variáveis originais e duas variáveis padronizadas, respectivamente; $r(X_j, X_j) = r_{jj}$ é a correlação entre duas variáveis originais (coeficiente de correlação momento-produto de Pearson); e \bar{X}_j é a média para a j -ésima variável.

Em virtude de as variáveis originais (X_i) possuírem diferentes unidades, utilizam-se as variáveis padronizadas (Z_j) na estimação dos componentes principais. A análise de componentes principais consiste na transformação do conjunto de p variáveis X_1, X_2, \dots, X_p em um novo conjunto Y_1, Y_2, \dots, Y_p , os componentes principais.

No estudo dos componentes principais, as seguintes propriedades são observadas:

a) seja Y_j um componente principal, então, Y_j é uma combinação linear de Z_j

$$Y_j = a_1 Z_1 + a_2 Z_2 + \dots + a_p Z_p;$$

b) os componentes são não-correlacionados, ou seja, são independentes entre si. Seja Y_j' outro componente, então:

$$Y_j' = b_1 Z_1 + b_2 Z_2 + \dots + b_p Z_p$$

e,

$$\sum_{j=1}^p a_j^2 = \sum_{j=1}^p b_j^2 = 1$$

$$\sum_{j=1}^p a_j b_j = 0;$$

c) entre todos os componentes, Y_1 apresenta a maior variância, Y_2 a segunda maior e assim sucessivamente.

$$\sigma^2(Y_1) > \sigma^2(Y_2) > \dots > \sigma^2(Y_p); e$$

$$d) \sum_{j=1}^p \sigma^2(Y_j) = \text{Tr}(R)$$

em que $\text{Tr}(R)$ é o traço (soma dos elementos da diagonal ou dos autovalores) da matriz de correlação R .

Com base nessas propriedades, estuda-se a dispersão dos indivíduos avaliados em sistemas cartesianos, em que o aproveitamento da variabilidade disponível seja maximizada. O problema estatístico consiste, fundamentalmente, em estimar os coeficientes de ponderação dos caracteres em cada componente, de modo que a variância a eles associada seja maximizada, observando-se as restrições descritas no item b.

MORRISON (1976) define o primeiro componente principal como uma combinação linear das variáveis estudadas cujos coeficientes a_j ($j = 1, 2, \dots, p$) são os elementos do vetor característico associado (α_1) com a maior raiz característica λ_1 da matriz de covariância (ou correlação. se for o caso) das variáveis. Se $\alpha_1' \alpha_1 = 1$, então λ_1 é interpretado como a variância amostral do primeiro componente.

MORRISON (1976) e DUNTEMAN (1984), dentre outros, demonstram que a variância associada a cada componente é estimada pelas raízes características da matriz R e os coeficientes de ponderação das características, pelos elementos dos vetores característicos correspondentes.

As estimativas dos autovalores (variância associada a cada componente principal) e autovetores (conjunto dos coeficientes dos componentes principais) são obtidas pela solução dos seguintes sistemas, respectivamente:

$$\det (R - \lambda I) = 0 \quad (3.3)$$

$$(R - \lambda_j I) a_j = 0 \quad (3.4)$$

em que

λ_j = raízes características (ou autovalores) da matriz de correlações entre as variáveis originais (ou de covariância entre as variáveis padronizadas). Existem p autovalores correspondentes às variâncias de cada um dos p componentes principais;

α_j = vetor característico (ou autovetor) que representa o conjunto de transformações ortogonais pelas quais as variáveis originais padronizadas devem ser multiplicadas para produzir as variáveis transformadas;

I: matriz identidade, de dimensão ($p \times p$);

R: matriz de correlações entre pares de variáveis originais;

e,

Φ : vetor nulo, de dimensão $p \times 1$.

A solução do sistema $(R - \lambda_j I) a_j = \Phi$ deve ser tal que $\alpha_j \neq 0$. Para tanto, impõe-se a condição $|R - \lambda_j I| = 0$, de modo que a solução possa ser escolhida entre aquelas que satisfaçam a restrição $\alpha_j' \alpha_j = 1$, que é necessária para que exista uma única solução (DUNTEMAN, 1984).

A importância relativa de um componente é avaliada pela percentagem da **variância total que ele explica**, A soma dos primeiros k autovalores dividida pela soma de todos os autovalores $(\lambda_1 + \dots + \lambda_k) / (\lambda_1 + \dots + \lambda_p)$ representa a proporção da **variância total explicada** pelos primeiros k componentes principais, ou seja, a proporção da **informação retida** na redução de p para k dimensões (**MORRISON, 1976**).

3.5.2. Análise Canônica

Seja X_{ij} a j -ésima característica ($j = 1, 2, \dots, p$) avaliada na i -ésima população ou amostra ($i = 1, 2, \dots, n$), as variáveis originais são transformadas em um novo conjunto, as **variáveis canônicas**, que possui as seguintes propriedades:

a) Se Y_j é uma variável **canônica**, então, Y_j é uma combinação linear de X .

$$V_j = a_1 X_1 + a_2 X_2 + \dots + a_p X_p$$

b) As **variáveis canônicas** são independentes entre si. Seja $Y_{j'}$ outra variável **canônica**, então

$$Y_{j'} = b_1 X_1 + b_2 X_2 + \dots + b_p X_p$$

$$\sum_{j=1}^p \sum_{j'=1}^p a_j a_{j'} \sigma_{jj'} = \sum_{j=1}^p \sum_{j'=1}^p b_j b_{j'} \sigma_{jj'} = 1;$$

$$\sum_{j=1}^p \sum_{j'=1}^p a_j b_{j'} \sigma_{jj'} = 0,$$

em que $\sigma_{jj'}$ é a **covariância residual** entre os caracteres j e j' .

c) Entre todas as **variáveis canônicas**, Y_1 apresenta a maior **variância**, Y_2 , a segunda maior e assim sucessivamente.

$$\sigma^2(Y_1) > \sigma^2(Y_2) > \dots > \sigma^2(Y_p).$$

Seja T a matriz de **variâncias e covariâncias** entre médias dos tratamentos e E a matriz de **variâncias e covariâncias** residuais, **RAO (1952)** relata que as j -ésimas **variâncias** das **variáveis canônicas** correspondem as **raízes características** de ordem correspondentes, obtidas pela **solução de:**

$$\det(T - \lambda_j E) = \det(E^{-1} T - \lambda_j I) = 0, \quad (3.5)$$

em que λ_j são as raízes características da matriz $E^{-1}T$.

As matrizes T e E são simétricas. Porém, o produto de ambas ($E^{-1}T$) não é uma matriz simétrica. O cálculo das raízes características de uma matriz assimétrica é muito complexo. Para simplificar os cálculos, os parâmetros descritos anteriormente têm sido estimados a partir de dados transformados. As funções de transformação são estabelecidas por um processo de condensação pivotal, descrito por RAO (1952), de modo que as novas variáveis apresentam covariância residual nula e variâncias residuais igual a unidade. Dessa forma, o processo de estimação toma-se idêntico ao de componentes principais (CRUZ, 1990).

As variáveis originais X_j são transformadas pelo processo de condensação pivotal (RAO, 1952), em variáveis padronizadas Z_j , com a matriz de variâncias e covariâncias entre tratamentos igual a T^* e a matriz de variâncias e covariâncias residuais igual a matriz identidade, I . Para o caso de duas variáveis, ARUNACHALAM (1981) relata outra forma de obtenção das variáveis padronizadas Z_j .

A matriz T^* é obtida a partir da matriz Z_{ij} , pelas variâncias e covariâncias entre as variáveis padronizadas Z_j .

Determinando-se os autovalores de T^* pela equação

$$\det (T^* - \lambda I) = 0, \quad (3.6)$$

obtêm-se as variâncias das j -ésimas variáveis canônicas. Os autovalores da matriz T^* correspondem aos da matriz $E''T$.

Os coeficientes de cada variável transformada por condensação pivotal são obtidos pela solução da equação:

$$(T^* - \lambda_j I) \alpha_j = 0 \quad (3.7)$$

em que

λ_j é a raiz característica que corresponde a variância da j -ésima variável canônica; e

α_j é vetor de coeficientes da j -ésima variável canônica, estabelecido com as variáveis transformadas por condensação pivotal.

As variáveis canônicas são obtidas como combinações lineares das variáveis padronizadas, Z_j , porém podem ser transformadas para combinações lineares das variáveis originais, X_i .

Considerando a seguinte variável canônica, obtida a partir de apenas três variáveis:

$$Y_1 = \alpha_1 Z_1 + \alpha_2 Z_2 + \alpha_3 Z_3.$$

Ou de maneira equivalente

$$Y_1 = a_1 X_1 + a_2 X_2 + a_3 X_3.$$

Seja as variáveis padronizadas, Z_j , obtidas pelo processo de condensação pivotal (RAO, 1952):

$$Z_1 = A_1 X_1.$$

$$Z_2 = B_1 X_1 + B_2 X_2.$$

$$Z_3 = C_1 X_1 + C_2 X_2 + C_3 X_3.$$

em que $A_1 \neq a_1$

então

$$Y_1 = (\alpha_1 A_1 + \alpha_2 B_1 + \alpha_3 C_1) X_1 + (\alpha_2 B_2 + \alpha_3 C_2) X_2 + \alpha_3 C_3 X_3,$$

sendo

$$[a_1 \ a_2 \ a_3] = [\alpha_1 \ \alpha_2 \ \alpha_3] \begin{bmatrix} A_1 & 0 & 0 \\ B_1 & B_2 & 0 \\ C_1 & C_2 & C_3 \end{bmatrix} \quad (3.8)$$

Quando existe maior número de variáveis originais essa transformação é realizada de modo análogo.

As variáveis canônicas são utilizadas em gráficos cartesianos de dispersão, possibilitando exame visual das divergências entre populações (ou amostras).

3.5.3. "Cluster Análise"

3.5.3.1. Medidas de Similaridade

3.5.3.1.1. Distância Euclidiana

Considerando um conjunto X de dados X_{ij} que representam os valores da j -ésima característica ($i = 1, 2, \dots, p$) avaliada no i -ésimo indivíduo ($i = 1, 2, 3, \dots, n$). Esse conjunto pode ser representado em forma de matriz por:

$$X = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1p} \\ X_{21} & X_{22} & \dots & X_{2p} \\ \vdots & \vdots & & \vdots \\ X_{n1} & X_{n2} & \dots & X_{np} \end{bmatrix}$$

Assim, cada vetor linha representa uma unidade amostral (indivíduo) e cada vetor coluna, uma variável.

Define-se a distância euclidiana média padronizada entre dois indivíduos i e i' pela expressão:

$$D_{ii'} = \sqrt{\frac{1}{p} \sum_{j=1}^p (Z_{ij} - Z_{i'j})^2} \quad (3.9)$$

em que

$$Z_{ij} = (X_{ij} - \bar{x}_j) / \sigma(x_j) \quad (3.10)$$

sendo $\sigma(x_j)$ o desvio-padrão da j -ésima característica,

3.5.3.1.2. Distância de Mahalanobis

Para a obtenção da distância de Mahalanobis (D^2), considera-se inicialmente um conjunto de dados X referentes as observações da j -ésima característica ($i = 1, 2, \dots, p$) da i -ésima população (ou amostra) ($i = 1, 2, \dots, n$) na k -ésima repetição ($k = 1, 2, \dots, r$). Assim, tem-se:

$$D^2_{ii'} = [d_1 \ d_2 \ \dots \ d_p] E^{-1} \quad (3.11)$$

em que

d_p : é a diferença entre médias de duas populações i e i' para uma dada característica j ; e

E: é a matriz de dispersão amostral (matriz de covariâncias residuais) comum a todas as unidades amostrais (dimensão $p \times p$), sendo constituída pelos quadrados médios e produtos médios residuais das características avaliadas.

Para a estimativa da matriz **E**, obtém-se os produtos médios (**PM**) para cada característica, por analogia aos quadrados médios (**QM**):

$$PM(X_1, X_2) = [QM(X_1 + X_2) - QM(X_1) - QM(X_2)] / 2. \quad (3.12)$$

A estimação da matriz de D^2 , pela expressão (3.11), pode tornar-se muito trabalhosa quando se estuda grande número de caracteres e, conseqüentemente, tem-se que obter a inversa de uma matriz de dispersão de ordem elevada. Entretanto, o cálculo de D^2 torna-se muito mais simples quando as características sob estudo são independentes e expressas em termos de seus respectivos erros padrões. Nesse caso, a obtenção de D^2 é expressa simplesmente pela soma dos quadrados das diferenças dos valores médios transformados dos vários caracteres de quaisquer pares de população (ARUNACHALAN, 1981 e CRUZ, 1987).

$$D^2_{ii'} = \sum_{j=1}^p (Z_{ij} - Z_{i'j})^2 \quad (3.13)$$

em que Z_j são as médias das variáveis não-correlacionadas e padronizadas, pelo processo de condensação pivotal.

Para se obter um conjunto de variáveis não-correlacionadas Z_j ($j = 1, 2, \dots, p$) a partir de um conjunto de variáveis originais X , têm sido utilizadas as informações fornecidas pela técnica de inversão por condensação pivotal da matriz de dispersão residual **E**. Esse processo permite, além da obtenção da inversa da matriz **E**, a obtenção das funções lineares para a transformação

do conjunto de variáveis correlacionadas X , em variáveis independentes Z_j e a **obtenção** da variância de cada uma das variáveis Z_j necessária a padronização das mesmas (RAO, 1952 e SINGH e CHAUDHARY, 1979).

3.5.3.1.3. "Distância de Canberra"

Seja uma matriz de dados X , a qual representa a média dos valores de intensidade de doença avaliada na i -ésima população (ou amostra) ($i = 1, 2, 3, \dots, n$) no j -ésimo período de tempo ($j = 1, 2, 3, \dots, p$),

$$X = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1p} \\ X_{21} & X_{22} & \dots & X_{2p} \\ \vdots & \vdots & & \vdots \\ X_{n1} & X_{n2} & \dots & X_{np} \end{bmatrix}$$

Define-se a distância entre duas populações i e i' pela expressão:

$$DC_{ii'} = \sum_{j=1}^p \left[\frac{|X_{ij} - X_{i'j}|}{(X_{ij} + X_{i'j})} \right] \quad (3.14)$$

3.5.3.2. Métodos de Agrupamento

Para a delimitação dos grupos, adotaram-se os métodos de agrupamento hierárquico do "vizinho mais próximo" e de otimização proposto por Tocher, citado por RAO (1952).

3.5.3.2.1. Método do Vizinho Mais Próximo

Como os demais, este método se inicia com uma matriz de distância (**dissimilaridade**) entre indivíduos (ou populações). São identificados os dois indivíduos mais similares (menor distância entre eles), os quais são reunidos em um grupo inicial. A seguir, calcula-se a distância daquele primeiro grupo em relação aos demais indivíduos.

A distância entre um grupo e um indivíduo é fornecida pela expressão:

$$d_{(ab)c} = \min \{d_{ac} ; d_{bc} \}$$

ou seja, a distância entre o grupo constituído pelos indivíduos, a e b, e o indivíduo c é fornecida pelo menor elemento do conjunto das distâncias dos pares de indivíduos (ac) e (bc). Já a distância entre dois grupos é obtida por:

$$d_{(ab)(cd)} = \min \{d_{ac} ; d_{ad} ; d_{bc} ; d_{bd} \}.$$

Constitui-se uma nova matriz de similaridade, de dimensão menor que a inicial, e identificam-se os indivíduos e, ou, grupos mais similares, incorporando-se outro indivíduo no grupo inicial ou formando um segundo grupo, caso a menor distância da nova matriz de similaridade seja entre dois outros indivíduos. Nas etapas seguintes, opera-se com matrizes de similaridade de dimensões cada vez menores, finalizando com o agrupamento de todos os indivíduos em um único grupo, formando um dendograma. **Este** método é ilustrado por MARDIA et alii (1979). MANLY (1986), CRUZ (1987) e MORAIS e SOARES (1988).

3.5.3.2.2.Método de Tocher

A partir da matriz de similaridade são identificados os dois indivíduos (ou populações) mais próximos entre si, ou seja, aqueles que apresentam a menor estimativa de distância, formando o primeiro grupo. A seguir, avalia-se a possibilidade de inclusão de novos indivíduos, respeitando-se o critério que a média das distâncias **intragrupo** seja menor que as distâncias **intergrupo**. Como a entrada de um indivíduo em um grupo sempre aumenta o valor médio da distância **intragrupos**, toma-se a decisão de incluí-lo neste grupo, por meio da comparação entre o acréscimo no valor da distância média do grupo e o nível máximo permitido. o qual pode ser estabelecido arbitrariamente ou geralmente, como tem sido feito, adotando-se a maior distância do conjunto de menores distâncias envolvendo cada indivíduo (CRUZ, 1987 e 1990).

Quando a adição de um indivíduo em particular acarreta aumento na média da distância **intragrupo** maior que a permitida, **este** indivíduo não é adicionado ao grupo. Similarmente. um segundo grupo é formado. Assim, o processo continua até que todos os indivíduos sejam incluídos em um ou em outro grupo, que são independentes entre si (SINGH e CHAUDHARY, 1979).

Ilustrações deste método são apresentadas por RAO (1952), SINGH e CHAUDHARY (1979) e CRUZ (1987).

As análises multivariadas, com exceção da estimativa da "distância de Canberra", foram realizadas pelo programa GENES, desenvolvido pelo Setor de Genética do Departamento de Biologia Geral da Universidade Federal de Viçosa.

4. RESULTADOS E DISCUSSÃO

4.1. Aplicação de 'Cluster Análise' no Estudo do Progresso da Ferrugem (*Tranzschelia sp.*) em Cultivares de Pessegueiro (*Prunus persica*)

Utilizando "cluster análise" os cultivares foram agrupados, de acordo com os IDs (índices de doença) em 11 avaliações, pelos métodos de agrupamento de Tocher (Quadro 5) e do vizinho mais próximo (Figura 1), baseados na distância euclidiana média e na "distância de Canberra".

Com o método do vizinho mais próximo, os grupos foram estabelecidos por um exame visual do dendograma, sendo considerado os pontos de alta mudança de nível como delimitadores do número de indivíduos em cada grupo (Figura 1). É uma forma subjetiva de delimitação de grupos, podendo o seu número variar de acordo com o pesquisador. Por exemplo, o grupo 3C (Figura 1) poderia ser considerado como dois grupos distintos (3C₁ e 3C₂).

"Cluster análise" apenas separa os cultivares em grupos similares quanto a resistência (progresso da ferrugem), não fornecendo indicações sobre o nível da resistência de cada grupo. Essa é uma desvantagem da análise. É necessário traçar a média da curva de progresso de doença (CPD) de cada grupo para conhecer o seu nível de resistência a doença. Neste estudo, apresentaram-se apenas as CPDs dos grupos formados com base na distância euclidiana média (Figuras 2 e 3). Considerando o método de agrupamento de Tocher e a distância euclidiana

média, as CPDs (Figura 2) apresentaram o grupo 1C ('Flordasun') como o mais susceptível, o grupo 1D ('Marli') como o mais resistente, o grupo 1B ('Princesa', 'V. Nova' e 'P. de Itaquera') como susceptível e o grupo 1A (demais cultivares) como moderadamente resistente.

Utilizaram-se outros métodos de comparação de epidemias para verificar a adequabilidade dos resultados de "cluster análise". Primeiramente, analisaram-se os IDs entre cultivares dentro de cada época de avaliação. Pela análise de variância dos dados, foram detectadas diferenças significativas entre os cultivares a partir da sexta avaliação (Quadro 6).

Comparando-se os IDs nas diferentes épocas de avaliação (Quadro 7), observam-se diferenças quanto à resistência entre os cultivares. Porém, é difícil interpretar os resultados ou estabelecer grupos, pois há alguns cultivares que diferem estatisticamente entre si em uma avaliação, mas não diferem em outra.

Pela análise da área abaixo da curva de progresso de doença (AACPD) (Quadro 8), observaram-se diferenças significativas entre os cultivares e a existência de interrelacionamento entre possíveis grupos de cultivares quanto a resistência à ferrugem. O cultivar 'Flordasun' foi mais susceptível, 'Marli' e 'Sinuelo', os mais resistentes e os demais apresentaram níveis intermediários de resistência.

Uma vantagem na utilização de "cluster análise" é não ser necessário assumir modelos matemáticos para o progresso de doença. Neste estudo, os modelos foram estimados para a verificação da eficiência de "cluster análise".

QUADRO 5 -Agrupamento de Cultivares de Pessegueiro Quanto ao Progresso da Ferrugem, pelo Método de Tocher, segundo as Medidas de Similaridade "Distância de Canberra" e Distância Euclidiana Média

Grupo	Euclidiana Média	Grupo	"Distância de Canberra"
1A	6 ' 7 8 1 0 5 9	2A	6 7 8 5 3 4
1B	3 4 2	2B	9 11
1C	1	2C	1 2
1D	11	2D	10

¹ Os cultivares foram representados pelos números: 1 (Flordasun), 2 (Princesa), 3 (V. Nova), 4 (P. Itaquera), 5 (Diamante), 6 (Cerrito), 7 (R. Conserva), 8 (Topázio), 9 (Sinuelo), 10 (Premier) e 11 (Marli).

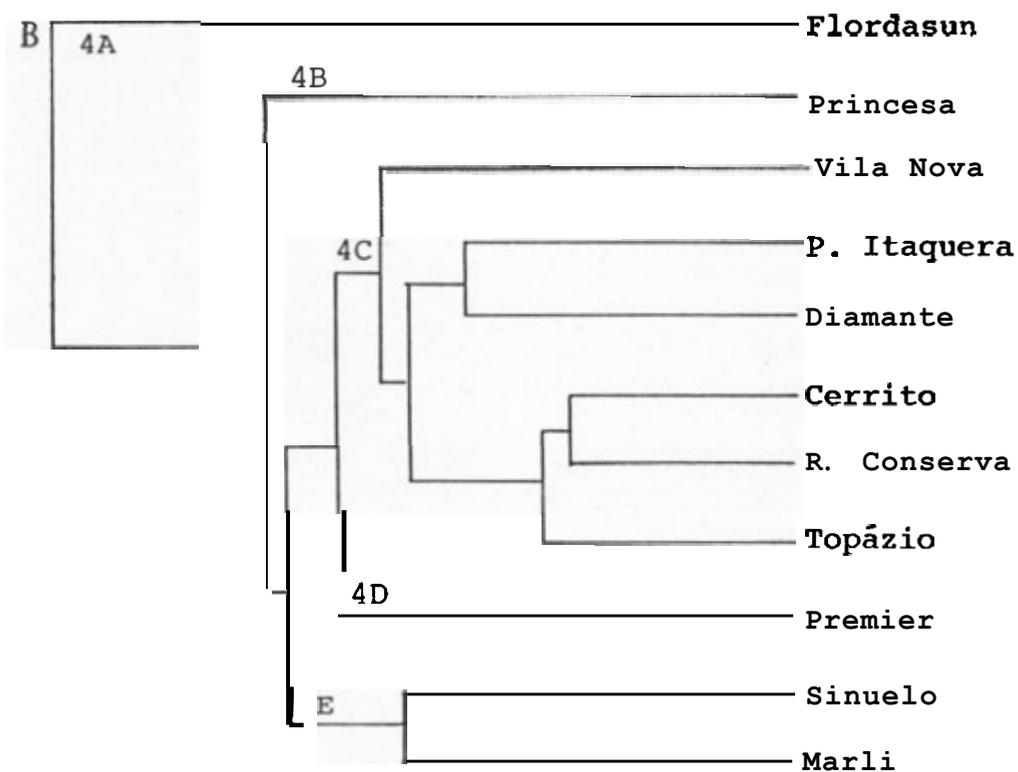
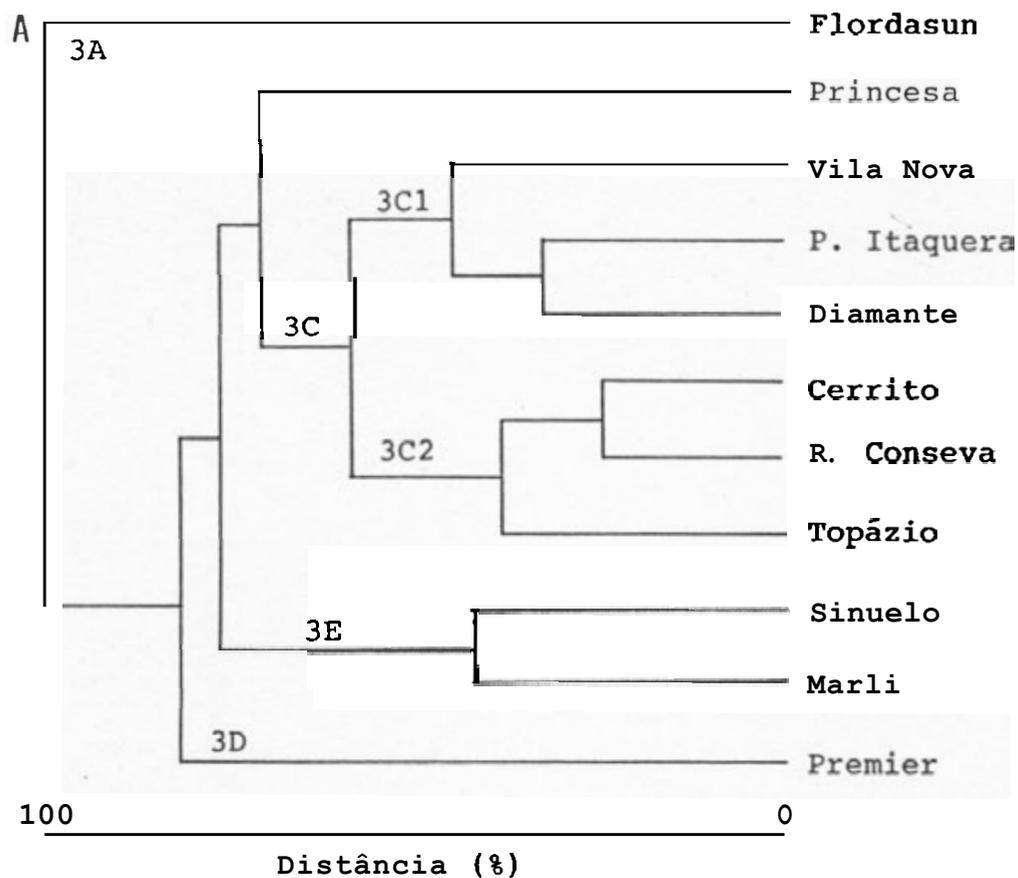


FIGURA 1 - Agrupamento pelo Método do Vizinho Mais Próximo de Cultivares de Pessegueiro Quanto ao Progresso da Ferrugem, pelas Distâncias de Canberra (A) e Euclidiana Média (B).

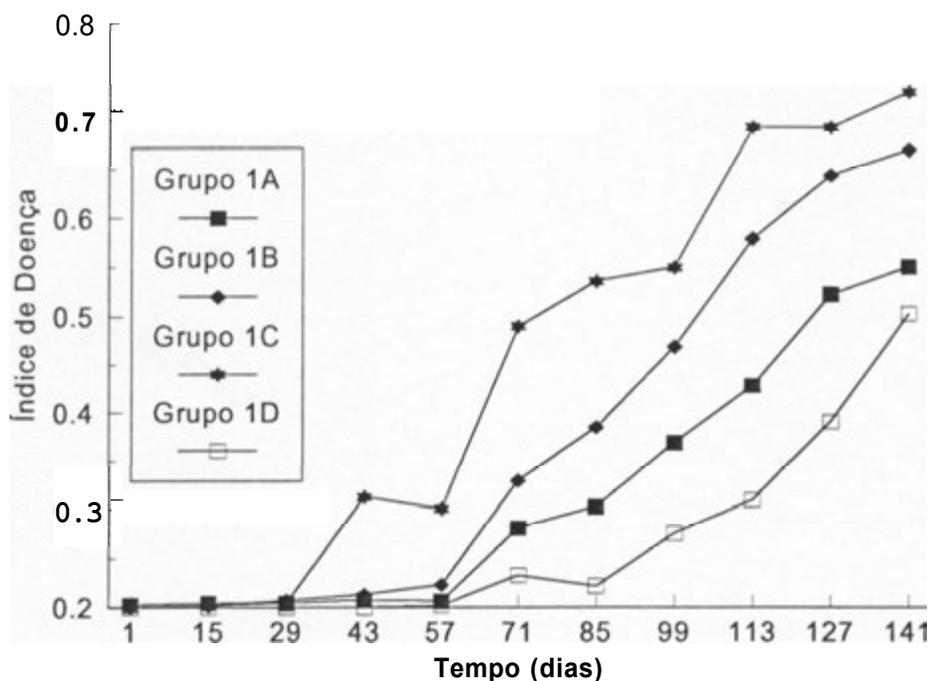


FIGURA 2 - Curvas de Progresso da Ferrugem em Grupos de Cultivares de Pessegueiro, Definidos pelo Método de Tocher, Utilizando a Distância Euclidiana Média.

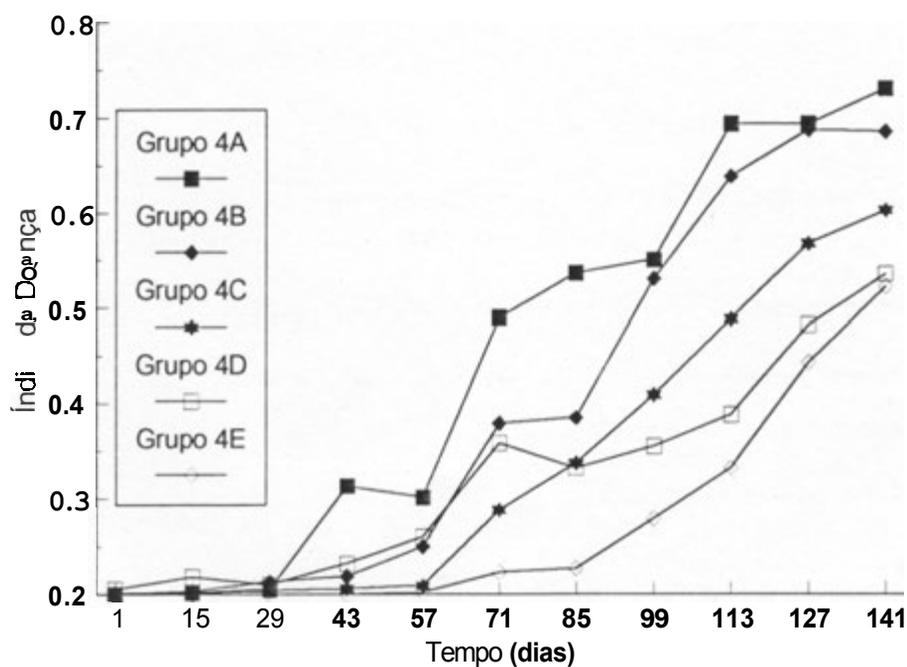


FIGURA 3 - Curvas de Progresso da Ferrugem nos Grupos de Cultivares de Pessegueiro, Definidos pelo Método do Vizinho Mais Próximo, Utilizando a Distância Euclidiana Média.

Os modelos de progresso de doença para cada cultivar são apresentados no Quadro 9. Julgou-se não adequada a utilização de modelos biológicos, como o logístico, o de Gompertz e outros, descritos por CAMPBELL e MADDEN (1990). para representar o progresso da ferrugem neste estudo. Pois, para a obtenção da forma *linearizada* desses modelos, é necessária a transformação dos IDs, que por sua vez já são uma transformação das mensurações originais GD (grau de doença).

Com o objetivo de verificar a existência de homogeneidade dentro e heterogeneidade entre os grupos formados pelos diferentes procedimentos de "cluster análise", comparou-se a igualdade dos modelos entre os cultivares pelo método descrito por GRAYBILL (1976) e REGAZZI (1993). Esse método utiliza o teste F e assume a igualdade entre os modelos como a hipótese de nulidade (H_0). Quando foram comparados os 11 modelos, simultaneamente, o valor F foi significativo, rejeitando H_0 e indicando que os modelos são diferentes (Quadro 10). Assim, não existe um modelo único que possa representar a epidemia em todos os cultivares. Como o método não discrimina se existe igualdade entre alguns modelos dentro de um conjunto de modelos cujo valor F foi significativo, considerou-se conveniente testar vários conjuntos de modelos quanto a sua igualdade, sendo esses conjuntos constituídos pelos diferentes grupos formados por "cluster análise" e as possíveis combinações entre dois diferentes grupos formados pelo mesmo método de agrupamento e medida de similaridade (Quadro 10).

QUADRO 6 - Resumo da Análise de Variância dos Cultivares (C) dentro de cada época de Avaliação (EA) em Relação ao índice de Doença

Fonte de Variação	F	Fonte de Variação	F
C / EA ₁	0,01	C / EA ₇	10,53**
C / EA ₂	0,04	C / EA ₈	10,94**
C / EA ₃	0,03	C / EA ₉	18,88**
C / EA ₄	1,47	C / EA ₁₀	11,61**
C / EA ₅	1,41	C / EA ₁₁	8,08**
C / EA ₆	8,37**		

** Significativo ($P < 0,01$).

QUADRO 7 - Índice de Doença de Cultivares de Pessegueiro da Sexta (EA₆) a Décima Primeira (EA₁₁) Época de Avaliação

Cultivar	Épocas de Avaliação ¹					
	EA ₆	EA ₇	EA ₈	EA ₉	EA ₁₀	EA ₁₁
Flordasun	0,49 a ^c	0,54 a	0,55 a	0,69 a	0,69 a	0,73 a
Princesa	0,38 ab	0,39 b	0,53 ab	0,64 ab	0,69 a	0,69 ab
Vila Nova	0,30 bcd	0,38 b	0,41 bc	0,51 bcd	0,65 ab	0,71 ab
P. Itaquera	0,31 bcd	0,39 b	0,47 abc	0,58 abc	0,59 abc	0,61 abc
Diamante	0,32 bcd	0,36 bc	0,45 abc	0,49 cde	0,57 abc	0,63 abc
Cerrito	0,26bcd	0,30 bc	0,39 cd	0,44 def	0,54 bc	0,60 bc
R. Conserva	0,25 cd	0,33 bc	0,37 cd	0,44 def	0,53 bc	0,54 c
Topázio	0,29bcd	0,27 bc	0,36 cd	0,46 def	0,52 c	0,53 c
Sinuêlo	0,21 d	0,23 c	0,28d	0,36 fg	0,50 cd	0,54 c
Premier	0,36 bc	0,33 bc	0,35 cd	0,39 efg	0,48 cd	0,54 c
Marli	0,23 cd	0,22 c	0,28d	0,3 1g	0,39 d	0,50c

¹ Correspondem a 71, 85, 99, 113, 127 e 141 dias após o início das avaliações.

² Médias seguidas pela mesma letra na vertical não diferem entre si pelo teste de Tukey (P < 0.05).

QUADRO 8 - Área Abaixo da Curva de Progresso da Doença (AACPD) para Cultivares de Pessequeiro nas Condições de Viçosa (MG)

Cultivares	A A C P D *
Flordasun	34,05 a
Princesa	27,04 a b
Pérola de Itaquera	21,83 b c
Vila Nova	21,28 b c
Diamante	19,84 b c
Premier	16,66 c d
Cerrito	15,95 c d
Rei da Conserva	15,26 c d
Topázio	14,68 c d
Sinuelo	10,29 d
Marli	7,99 d
CV** (%)	12,14

• Médias seguidas da mesma letra não diferem entre si pelo teste de Tukey ($P < 0.05$)

** Coeficiente de variação.

QUADRO 9 - Modelos de Progresso da Ferrugem em Cultivares de Pessequeiro nas condições de Viçosa (MG)

Cultivares	Modelo'	R ² (%)	ρ**
Flordasun	ID = 0,1351 + 0,0044 T	94,99	- 0,12
Princesa	ID = 0,1053 + 0,0041 T	91,31	0,44
Vila Nova	ID = 0,0931 + 0,0038 T	87,10	0,65
P. Itaquera	ID = 0,1102 + 0,0035 T	89,91	0,56
Diamante	ID = 0,1160 + 0,0033 T	90,46	0,57
Cerrito	ID = 0,1179 + 0,0029 T	85,50	0,72
R. Conserva	ID = 0,1238 + 0,0027 T	87,09	0,62
Topázio	ID = 0,1248 + 0,0026 T	85,06	0,51
Sinuêlo	ID = 0,1202 + 0,0023 T	72,59	0,77
Premier	ID = 0,1611 + 0,0023 T	91,20	0,25
Marli	ID = 0,1386 + 0,0018 T	72,24	0,78

* Os modelos (ID = a + b.T, sendo T o tempo em dias) foram estimados com médias de duas repetições. As estimativas do parâmetro b de todos os modelos foram significativas pelo teste t (P < 0,01).

** Autocorrelação.

O teste de identidade de modelos (Quadro 10) para o grupo 2A, constituído pelos cultivares 'Cerrito' (6), 'Rei da Conserva' (7), 'Sinuelo' (8), 'Diamante' (5), 'Vila Nova' (3) e 'Pérola de Itaquera' (4), obtido pelo método de Tocher utilizando a "distância de Canberra" (Quadro 5), resultou em um valor $F (1,62)$ não-significativo. Portanto, os modelos desses cultivares foram considerados iguais. Convém ressaltar que o grupo 2A (método de Tocher e distância euclidiana média) (Quadro 5) é equivalente aos grupos 3C e 4C (método do vizinho mais próximo baseado na "distância de Canberra" e na distância euclidiana média, respectivamente) (Figura 1). Procedendo-se ao teste de identidade de modelos para os demais grupos, separadamente, verificou-se que dentro de cada grupo formado (Quadro 5 e Figura 1), independente da medida de similaridade e do método de agrupamento utilizados, os modelos foram considerados iguais (Quadro 10). Conclui-se pelo teste de igualdade de modelos, que há homogeneidade quanto a resistência (progresso da doença) dentro de cada grupo formado por "cluster análise".

Para verificar a existência de heterogeneidade quanto a resistência entre os grupos formados, procedeu-se ao teste de identidade para os diferentes conjuntos de modelos relativos aos cultivares pertencentes a cada dois grupos formado pela mesma medida de similaridade e pelo mesmo método de agrupamento. O teste de identidade do conjunto de modelos relativos aos cultivares pertencentes aos grupos 2A e 2D (6-'Cerrito', 7-'Rei da Conserva', 8-'Topázio', 5-'Diamante', 3-'Vila Nova', 4-'Pérola de Itaquera' e 10-'Premier') resultou em um valor $F (1,88)$ não-significativo (Quadro 10), ou seja, os modelos dos dois respectivos grupos são iguais entre si. Em relação ao teste de identidade de modelos, há homogeneidade entre os respectivos grupos, que poderiam ser reunidos em um único grupo. Procedendo-se aos demais testes de identidade de conjuntos de modelos constituídos por dois grupos, verificou-se que, com exceção do método de Tocher utilizando a distância euclidiana média (Quadro 5), os demais procedimentos multivariados de agrupamento separaram grupos que poderiam ser considerados como um único, como é o caso dos grupos 2A e 2D, 2B e 2D (método de Tocher e "distância de Canberra"); 3A e 3B, 3C e 3D, 3D e 3E (método do vizinho mais próximo e "distância de Canberra"); e 4A e 4B, 4C e 4D, 4D e 4E (método do vizinho mais próximo e distância euclidiana média) (Quadro 10).

QUADRO 10 - Resumo do Teste de Identidade de Modelos para Diferentes Combinações de Grupos de Cultivares de Pessegueiro

Cultivares ¹	F	Grupos ²
1 2 3 4 5 6 7 8 9 1 0 1 1	6,73 **	
3 4 5 6 7 8	1,62	2A (3C)
9 11	0,53	26 (3E)
1 2	2,06	2C (3A + 3B)
1 2 3 4 5 6 7 8	4,90 **	2A + 2C
3 4 5 6 7 8 9 1 1	3,27 **	2A + 2B (3C + 3E)
3 4 5 6 7 8 1 0	1,88	2A + 2D (3C + 3D)
9 10 11	1,90	26 + 2D (3D + 3E)
1 2 9 1 1	16,95 **	26 + 2C
1 10	34,87**	3A + 3D
5 6 7 8 9 1 0	1,38	1A
2 3 4	0,82	16
2 3 4 5 6 7 8 9 1 0	3,35**	1A + 1B
5 6 7 8 9 1 0 1 1	2,24'	1A + 1D
1 2 3 4	2,75*	1B + 1C
2 3 4 1 1	7,53**	1B + 1D
1 11	45,52**	1C + 1D
1 9 11	21,65**	3A + 3E
2 10	12,84**	36 + 3D
2 9 11	11,39**	36 + 3E
1 3 4 5 6 7 8	5,02**	3A + 3C
2 3 4 5 6 7 0	2,62**	36 + 3C
5 6 10	1,58	

* ** Significativo ($P < 0,05$ e $P < 0,01$, respectivamente)

¹ Os cultivares foram representados pelos números: 1 (Flordasun), 2 (Princesa), 3 (V. Nova), 4 (P. Itaquera), 5 (Diamante), 6 (Cerrito), 7 (R. Conserva), 8 (Topázio), 9 (Sinuelo), 10 (Premier) e 11 (Marli).

² Grupos formados pelo método de Tocher: 1A a 1D (distância euclidiana média) e 2A a 2D ("distância de Canberra"); e pelo método do vizinho mais próximo: 3A a 3E ("distância de Canberra"), os quais são equivalentes aos grupos 4A a 4E (distância euclidiana média).

O teste de identidade de modelos utilizado neste estudo tem a vantagem de comparar CPDs, considerando todo o modelo de progresso de doença, enquanto o método tradicionalmente utilizado compara apenas um parâmetro do modelo: normalmente a taxa de progresso de doença (o parâmetro b no modelo utilizado neste estudo - Quadro 9) (MADDEN, 1986 e CAMPBELL e MADDEN, 1990). Este último método possui as desvantagens de comparar as CPDs duas a duas, além de poder considerar iguais duas CPDs que, embora tendo semelhantes taxas de progresso de doença, possuam diferentes níveis iniciais de doença. Ademais, há um grande potencial de erro em estudos comparando CPDs pela taxa de progresso de doença, quando o inóculo inicial não é equivalente entre tratamentos. Pois verificou-se que para o patossistema *Helminthosporium maydis* - milho (GREGORY et alii, 1981) e oídio - trigo (ROUSE et alii, 1981), a taxa de infecção aparente (VAN DER PLANK, 1963) é dependente da quantidade inicial de doença. A utilização do teste de identidade de modelos (GRAYBILL, 1976 e REGAZZI, 1993) pode contornar estes problemas.

Os padrões de agrupamento resultantes de "cluster análise" dependem do método de agrupamento e da medida de similaridade adotados, como pôde ser observado neste estudo. Houve diferenças nos padrões de agrupamento baseado na distância euclidiana média e "distância de Canberra", para ambos os métodos de agrupamento utilizados (Quadro 5 e Figura 1). Com o método do vizinho mais próximo a composição dos grupos permaneceu inalterada com relação à medida de distância utilizada, mudando apenas o relacionamento entre os mesmos. Diferentemente do método do vizinho mais próximo, o método de Tocher não revela o relacionamento entre grupos, porém pode-se obter uma idéia sobre esse relacionamento calculando-se a distância média intergrupo. SINGH e CHAUDHARY (1979) exemplificam esse cálculo utilizando a distância de Mahalanobis.

As diferenças entre os métodos de agrupamento estão nas diferentes formas de definir proximidade entre um indivíduo (neste caso cultivar) e um grupo ou entre dois grupos de indivíduos. Existe um grande número de métodos de agrupamentos propostos. A princípio, qualquer método poderia ser aplicado a uma matriz de similaridade (ou dissimilaridade), pois não existe um aceito como o melhor. Para escolher um método adequado a um dado estudo, RIBOLDI (1985) propõe aplicar vários métodos de agrupamento simultaneamente a mesma matriz de similaridade e apenas os grupos formados pela maioria dos métodos seriam aceitos. LINDE et alii

(1990) obtiveram resultados similares entre os vários métodos de agrupamento utilizados. No presente estudo, isso não ocorreu. Para os métodos hierárquicos, a eficiência com que o **dendograma** representa a matriz de similaridade pode ser avaliada pelo CCC, coeficiente de correlações cofenéticas (SOKAL e ROHLF, 1962 e ROMESBURG, 1984). Os dendogramas formados pelo método do vizinho mais próximo baseado na "**distância de Canberra**" e na distancia euclidiana média apresentaram similares valores de CCC, **0,74 e 0.75**, respectivamente.

Com relação ao método de Tocher, embora o número de grupos não tenha sido alterado, a composição dos mesmos variou de acordo com a medida de distancia adotada (Quadro 5). Porém, com exceção do fato do cultivar 'Premier' ter constituído, isoladamente, um grupo (**2D**) quando utilizou-se a "distancia de Canberra", as variações na composição dos grupos podem ser compreendidas, quando **se** comparam **estes** resultados com os das outras **técnicas** de análise utilizadas.

Os cultivares 'Premier', 'Diamante' e 'Cerrito' possuem semelhantes CPDs (Figura 4), não há diferenças significativas entre os seus IDs em qualquer época de avaliação (Quadro 7) e as suas AACPDs não diferem, estatisticamente, entre si (Quadro 8). Em virtude de até a sexta avaliação o cultivar 'Premier' possuir um maior ID que o 'Diamante' e o 'Cerrito' (Figura 4), invertendo-se esta situação nas **avaliações** posteriores, a estimativa da "**distância de Canberra**" detectou suficiente **dissimilaridade** entre **esses** cultivares (diferentemente da **distância euclidiana** média), para que o método de Tocher classificasse 'Premier' em um grupo a parte. O processo matemático de obtenção da "**distância de Canberra**" fornece maior peso **as** diferenças (entre os IDs) proporcionais que **as** absolutas, ou seja, diferença no início da epidemia possuem maior **importância** que a mesma diferença absoluta no decorrer da epidemia (ver **seção 3.5.3.1.3.**). Deve-se ressaltar que a escala de avaliação de doença utilizada neste estudo é logarítmica. Portanto, a própria escala é mais **sensível** a diferenças proporcionais que absolutas. Convém também ressaltar que o método do vizinho mais próximo classificou 'Premier' como um grupo a parte, com ambas as medidas de **distância** utilizadas (Figura 1).

Com relação a medida de similaridade a ser utilizada, a sua escolha é baseada no tipo de dados **disponíveis**. Quando mais de uma medida pode ser aplicada aos dados, a escolha é geralmente baseada na **preferência** do pesquisador. THOMPSON e REES (1979) utilizaram a "**distância de Canberra**" para estimar a similaridade entre cultivares de trigo com relação ao

progresso da ferrugem do colmo. **Os** resultados foram semelhantes quando a estimativa foi realizada utilizando a distância euclidiana, porém, essa proporcionou maior separação entre cultivares mais susceptíveis. **Os** resultados de "cluster análise" foram concordantes com o de AACPD. CAMPBELL *et alii* (1980) utilizaram a distância euclidiana para estimar a similaridade entre CPDs de podridão do hipocótilo do feijoeiro. Esses autores, porém, não consideraram as intensidades de doença em cada época de avaliação como variáveis, e sim seis elementos da CPD.

A princípio, as medidas de distância utilizadas neste estudo podem ser aplicadas a dados de progresso de doença, embora ambas não considerem a variabilidade amostral nem a correlação entre variáveis (épocas de avaliação). MADDEN e PENNYPACKER (1979), com o objetivo de caracterizar a estrutura de 18 CPDs de mancha de *Alternaria* em tomateiro, considerando a correlação entre oito variáveis (intensidade de doença avaliadas em oito épocas, a intervalos semanais), utilizaram a análise de componentes principais. Porém, as análises multivariadas que consideram a variabilidade amostral e a correlação entre variáveis, como a análise canônica ou "cluster análise" baseada na distância de Mahalanobis, ou que consideram somente a correlação entre variáveis, como a análise de componentes principais, não são adequadas para serem utilizadas em dados de progresso de doença, por causa da mensuração de cada variável ser realizada em épocas diferentes.

A correlação entre os IDs nas diferentes épocas de avaliação não afeta os resultados obtidos com "cluster análise" (utilizando como medidas de similaridade a distância euclidiana ou "distância de Canberra"). A correlação entre os IDs também não é considerada quando se utiliza a AACPD ou os níveis de doença em cada época de avaliação, para comparação de CPDs, mas é importante considerar essa correlação quando são obtidos modelos de progresso de doença pela análise de regressão (intensidade de doença em função do tempo), utilizando o método dos mínimos quadrados, pois um dos pressupostos desse método é que o resíduo de uma época de avaliação (também denominado erro, é a variabilidade não-explicada, ou seja, é a diferença entre os valores observados e os valores preditos) não seja correlacionado com o resíduo em outra época de avaliação (MADDEN, 1980 e HOFFMANN e VIEIRA, 1983).

Os níveis de doença em cada época de avaliação não são independentes entre si. Essa dependência pode levar a alta correlação dos resíduos (autocorrelação) segundo MADDEN (1980

e 1986), que afirma que quando existe uma alta autocorrelação, a estimativa dos parâmetros do modelo (os quais formam a base para a análise estatística e comparação de CPDs) pelo método dos mínimos quadrados é incorreta, pois, no mínimo, a variância e as covariâncias dos parâmetros poderão ser estimadas erroneamente. Com isso, afeta-se a comparação dos parâmetros dos modelos (MADDEN, 1986). Neste estudo, as autocorrelações (ρ), calculadas segundo MADDEN (1986), variaram de -0.12 a 0,78 (Quadro 9). Devido a variável independente possuir menos de quinze pontos de dados, não foi possível testar a significância dessas autocorrelações pelo teste de Durbin-Watson (HOFFMANN e VIEIRA, 1983). Segundo MADDEN (1986), baseado em sua experiência, valores de $\rho > 0.70$ poderão ter substancial impacto nos resultados da regressão e, nesses casos, a autocorrelação deveria ser incorporada na análise de regressão. Como esse é um critério subjetivo, desconsiderou-se, neste estudo, o efeito da autocorrelação na análise de regressão.

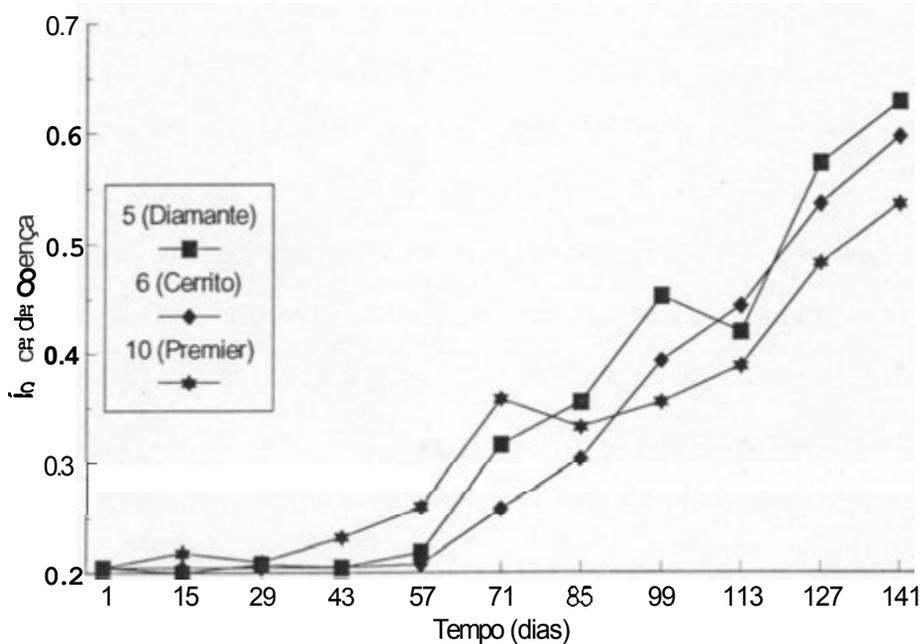


FIGURA 4 - Curvas de Progresso da Ferrugem nos Cultivares 'Diamante', 'Cerrito' e 'Premier', de Outubro de 1988 a Fevereiro de 1989, nas Condições de Viçosa (MG).

Em estudos comparativos de CPDs, o objetivo é determinar **se** duas ou mais CPDs diferem significativamente de cada outra. **Os** níveis de doença nas várias épocas de avaliação, a AACPD e os parâmetros estimados dos modelos de progresso de doença **são** características **úteis** e tradicionalmente utilizadas para comparação (CAMPBELL e MADDEN, 1990). Neste trabalho, verificou-se a **aplicabilidade** de "cluster análise" no estudo de progresso de doenças, a qual deve ser utilizada complementando os métodos tradicionalmente utilizados, pois permite obter nova perspectiva dos dados sob consideração, sem assumir um definido modelo para a CPD. Essa análise é de grande utilidade, principalmente quando o número de cultivares (ou tratamentos) for elevado.

Existem outras aplicações de "cluster análise" em epidemiologia, **além** de agrupar cultivares como realizado no presente estudo. "Cluster análise" também foi utilizada para descobrir padrões de epidemias (Kranz, 1968, citado por **HAU e KRANZ, 1990 e CAMPBELL et alii, 1980**).

4.2. Aplicação da Análise **Canônica** e "Cluster Análise " no Estudo da Resistência Horizontal de **Genótipos de Coffea arabica** a Ferrugem (*Hemileia vastatrix*)

"Cluster análise", utilizando o método de agrupamento de Tocher, revelou três grupos de genótipos. **O** genótipo **H472**, que constituiu isoladamente um primeiro grupo, apresentou maior resistência que **os** demais, com exceção de **H543** (que apresentou-se imune à raça II do patógeno, não sendo considerado nas análises **multivariadas**, pois constituiu, obviamente, um grupo a parte). **Os** genótipos **H442 e H481** constituíram um outro grupo, apresentando-se com nível intermediário de resistência. **Os** genótipos **H415, UFV2143** ('Catuaí'), **H471 e UFV2150** ('Mundo Novo') constituíram um terceiro grupo, susceptível. O método de agrupamento do vizinho mais próximo (Figura 5) dividiu esse terceiro grupo em dois grupos, sendo concordante com os demais resultados obtidos pelo método de Tocher.

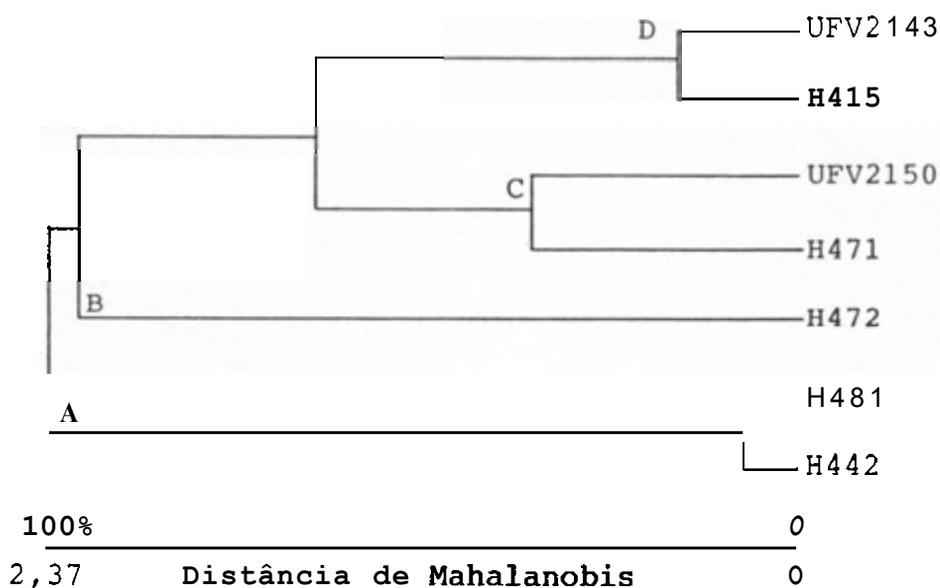


FIGURA 5 - Agrupamento pelo Método do Vizinho mais Próximo Quanto à Resistência Horizontal de Genótipos de Cafeeiros a *H. vastatrix*, Baseado em Quatro Componentes de Resistência.

QUADRO 11- Médias dos Componentes de Resistência de Genótipos de Cafeeiros a Raça II de *H. vastatrix* e Número de Plantas com Lesão Esporulada (PLE)

Genótipo	Componentes de Resistência.				
	PI	PL	SEV	ESP	PLE
	(dias)	(dias)	(%)	(esporos/cm ²)	(%)
H415-2	18,64	25,30	12,36	1027	54
H442-1	19,98	28,42	15,76	477	74
H471-4	18,64	25,66	17,22	15699	76
H472-2	19,80	30,18	10,26	6363	36
H481-2	19,90	28,50	16,54	6785	64
UFV 2143	19,10	26,26	11,28	10443	72
UFV 2150	18,64	26,85	16,18	11994	94
H543-1			0	0	0

* PI: período de incubação; PL: período latente; SEV: severidade; e ESP: esporulação.

Embora a resistência de cada grupo não seja indicada pelos métodos de agrupamento, esta pode ser conhecida pelas médias dos componentes de resistência, relativas a cada grupo. Como neste estudo o número de **genótipos** é pequeno, apresentaram-se as médias dos componentes de resistência, relativas a cada **genótipo** (Quadro 11).

A avaliação complementar do grau de doença em outras **40** mudas de cada **genótipo** (Figura 6) apresentou resultados coerentes com os de "cluster analyse" utilizando o método de Tocher. Ou seja, houve maior similaridade com relação aos graus de doença dentro que entre os grupos.

Previamente a análise **canônica**, transformaram-se as variáveis **SEV** (severidade) e **ESP** (**esporulação**) para arco seno (**SEV/100**) (= **SEV***) e log **ESP** (= **ESP***), respectivamente. Procedeu-se a análise **canônica** utilizando quatro variáveis. (Quadro 12).

A avaliação da similaridade entre os **genótipos** na análise **canônica** se faz pelo exame em dispersões gráficas. Quanto mais próximos estão dois **genótipos** maior é a similaridade entre ambos. Porém, a delimitação de grupos de similaridade é realizada de modo subjetivo, com base na inspeção visual da dispersão. Houve similaridade entre os agrupamento do vizinho mais próximo (Figura 5) e a avaliação da divergência entre os **genótipos**, baseada em dispersões gráficas em relação as duas primeiras variáveis **canônicas** (Figura 7).

Uma das vantagens das análises **multivariadas** é que **são** consideradas, simultaneamente, as informações de todas as variáveis. Quando cada variável foi analisada pela análise de **variância** univariada, não foram verificadas diferenças significativas (Quadro 13), porém a análise **multivariada** de **variância** (**MANOVA**) apresentou diferenças significativas entre os **genótipos**, considerando os quatro componentes de resistência simultaneamente. Segundo **MANLY (1986)**, é possível obter resultados não-significativos em **testes** univariados e um resultado significativo em teste **multivariado**, e também vice-versa.

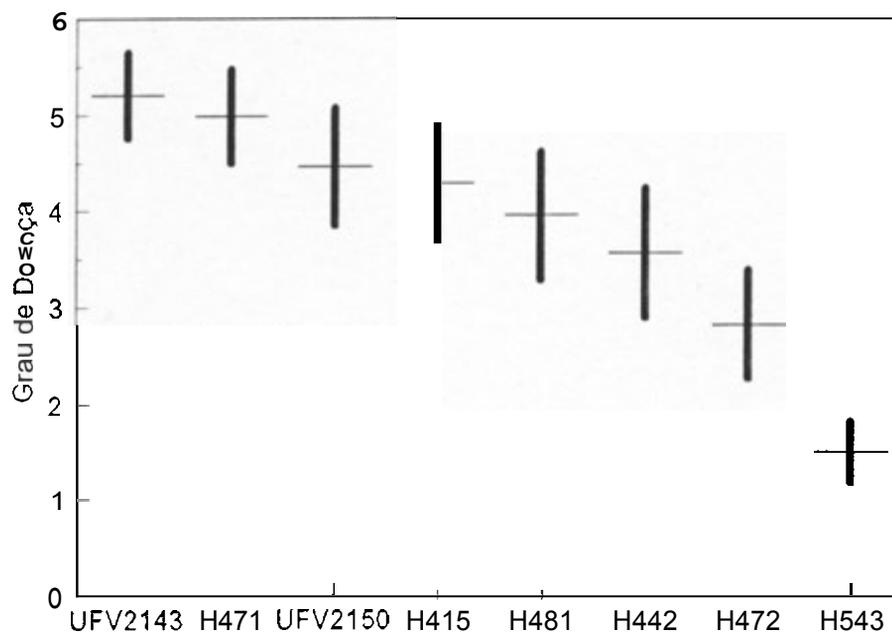


FIGURA 6 - Intervalo de Confiança do Grau Médio de Doença, 62 Dias após a Inoculação, de Genótipos de *Coffea arabica* em Relação a Raça II de *H. vastatrix*.

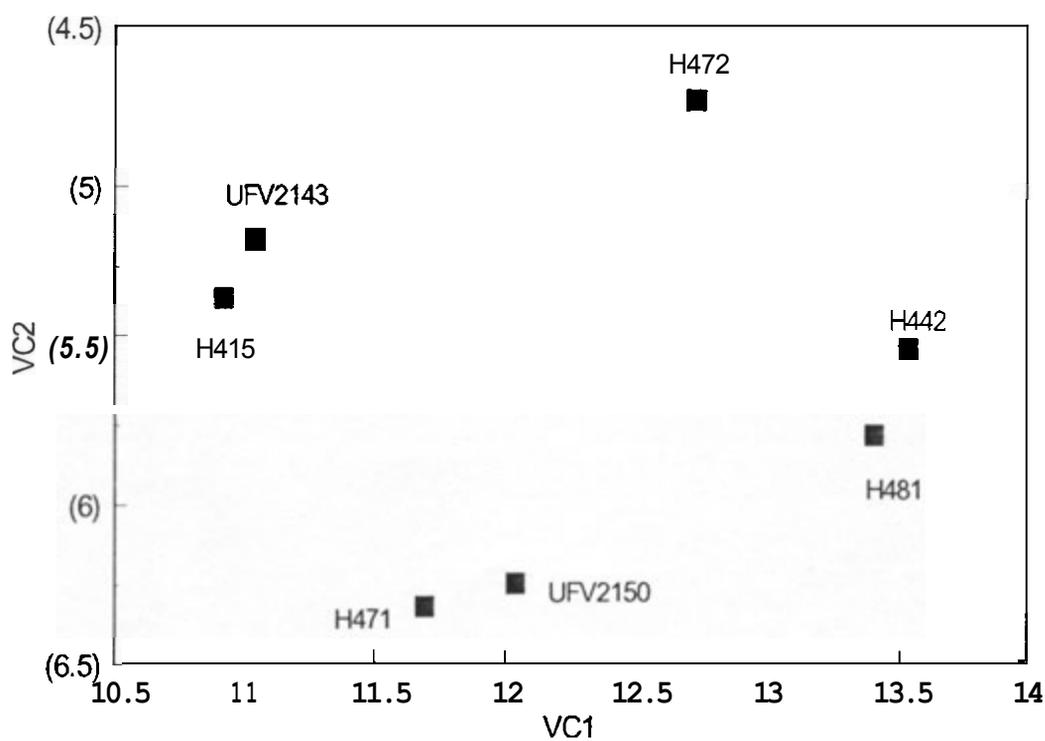


FIGURA 7 - Dispersão Gráfica entre Genótipos de *Coffea arabica* com Relação as Duas Primeiras Variáveis Canônicas (VC1 e VC2).

QUADRO 12 - Autovetores e Autovalores das Variáveis Canônicas (VC) Baseadas em Quatro Componentes de Resistência de Sete Genótipos de Cafeeiros

Autovalor	Variância Acumulada (%)	Componentes de Resistência ¹			
		PI	PL	SEV*	ESP ¹
		Variáveis Canônicas (Autovetores)			
1,135	68,29	0,220	0,309	0,248	-1,537
0,328	88,03	0,248	-0,113	-0,169	-0,946
0,160	97,65	-0,210	0,459	-0,068	2,604
0,039	100,00	1,040	-0,041	-0,066	2,359

¹ **PI**: período de incubação; **PL**: período latente; **SEV***: arco seno (severidade/100); e **ESP¹**: log (esporulação).

Quadro 13 - Resumo da MANOVA e das Análises de Variância Univariada de Quatro Componentes de Resistência de Genótipos de Coffea arabica a Hemileia vastatrix

Fonte de Variação	GL	Quadrado Médio				MANOVA ¹
		PI	PL	SEV	ESP	Teste F
Genótipo	6	1,997	15,794	40,317	7.14 x 10 ⁷	22,6 *
Resíduo	28	1,040	6,578	30,711	3.77 x 10 ⁷	
Média		19,24	27,31	14,23	9474,95	
CV ** (%)		5,30	9,39	38,95	64,78	

* Significativo (P < 0,00001).

..Coeficiente de variação.

¹ Critério de Wilks.

A ausência de diferenças significativas quando se procedeu as análises univariadas de variância com relação a SEV e ESP provavelmente é devida aos altos coeficientes de variação (CVs) dos dados (Quadro 13). Esse fato pode ser atribuído as plantas inoculadas não possuírem uniformidade genética, estando em geração F_2 , bem como ao efeito ambiental, que é grande em experimentos de inoculação em casa de vegetação. Segundo **ESKES (1983)**, o CV de uma população geneticamente homogênea de *Coffea arabica* pode ser superior a 100% para SEV. **ESKES e CARVALHO (1983)** obtiveram altos CVs para a densidade de lesões (número de lesões por unidade de área foliar) e ESP na avaliação de componentes de resistência de linhagens de *C. arabica* em experimentos conduzidos sob condições de casa de vegetação.

Em 1963, **VAN DER PLANK** defendeu a hipótese de existência de dois tipos de resistência de plantas a patógenos: resistência vertical (RV) e horizontal (RH). Segundo **VAN DER PLANK (1963, 1968 e 1975)**, a RV age no sentido de reduzir a quantidade efetiva de inóculo inicial, retardando o início da epidemia. A RH age diminuindo o tamanho e o número das lesões produzidas pelo patógeno, aumentando os períodos de incubação e latência, diminuindo o número de esporos por lesões etc. Essas manifestações podem ser expressas isoladamente ou em conjunto, e em decorrência, a epidemia progride lentamente.

Os componentes de resistência agem cumulativamente para diminuir a capacidade reprodutiva do patógeno, porém, podem ter diferentes efeitos no progresso da doença (**PARLEVLIET, 1979 e ZADOKS e SCHEIN, 1979**). A análise dos componentes de resistência é útil em determinar um eficiente critério de seleção em trabalhos de melhoramento. No estudo da associação entre os componentes de resistência e na caracterização dos genótipos, "cluster análise" e análise canônica são de grande valor. Outras técnicas também são utilizadas. **JEGER et alii (1983)** utilizaram as análises fatorial e de componentes principais para elucidar as relações entre os componentes de resistência de genótipos de trigo a *Septoria nodorum*.

Diferenças entre cultivares com relação ao PL (período latente) existem em diversos patossistemas (**PARLEVLIET, 1979**). O PL tem sido associado com "slow leaf rusting" de trigo a *Puccinia recondita* (**OHM e SHANER, 1976**), com resistência parcial de cevada (**PARLEVLIET, 1975; PARLEVLIET e VAN OMMEREN, 1975 e PARLEVLIET et alii, 1980**) e soja (**ZAMBOLIM et alii, 1983**) a ferrugem e de aveia a oídio (**JONES, 1978**). A avaliação da similaridade pela distância de Mahalanobis permitiu estimar a contribuição de cada variável para a divergência

entre os genótipos (SINGH, 1981). Os resultados indicam o PL como o componente mais importante (34,49%), embora neste estudo não tenham sido observadas diferenças significativas entre os genótipos de cafeeiros com relação ao PL. PI (período de incubação) poderia ser descartado, por causa da sua pequena importância na avaliação da divergência (8,73%). ESP e SEV contribuíram com 30,84 e 25,93%, respectivamente. Para 'Icatu' (*C. arabica* × *C. canephora*) e algumas populações de *C. arabica*, o PL avaliado em casa de vegetação apresentou **significante** correlação com o nível de doença no campo (ESKES, 1983; ESKES e CARVALHO, 1983 e ESKES e COSTA, 1983). Mais estudos devem ser realizados para **se** obter uma avaliação mais segura da **importância** de cada componente de resistência. Segundo ESKES e CARVALHO (1983), não é possível indicar um componente de resistência como critério de **seleção** mais adequado para todas as populações de *C. arabica*. A importância de cada componente varia de acordo com a população.

A ferrugem do cafeeiro constitui o maior problema **fitossanitário** do continente americano, causando perdas de produção no Brasil, de 25 a 30%. e o controle químico da ferrugem com **fungicidas** constitui a **única** medida recomendada a curto e médio prazos (ZAMBOLIM et alii, 1985a), sendo, porém, de difícil execução na prática, além de onerar os custos de produção. A utilização de variedades resistentes a ferrugem constitui-se a alternativa mais adequada e eficiente para evitar ou diminuir os prejuízos dessa enfermidade. Porém, ainda não **se** dispõe de cultivares resistentes para serem plantados em escala comercial. **Desde há** muito procura-se criar cultivares resistentes ao patógeno, **objetivando** dispensar total ou parcialmente o controle químico e substituir de maneira vantajosa os cultivares tradicionais de *C. arabica*, altamente susceptíveis (BETTENCOURT, 1981).

Por longo tempo os trabalhos realizados sobre a resistência do cafeeiro a *H. vastatrix*, desde os de Mayne (1932, 1936 e 1939) citados por CHAVES (1976) até os de BETTENCOURT e CARVALHO (1968) e BETTENCOURT (1981). trataram da RV, em virtude da mesma proporcionar imunidade da planta a raça ou raças do patógeno e envolver mecanismos cuja herança é governada por genes simples e dominantes, relativamente fáceis de serem manipulados em programas de melhoramento, muito embora fosse questionada sua validade para uma planta perene como o cafeeiro (ROBINSON, 1971). A tentativa de controle da doença pela obtenção de variedades com resistência essencialmente vertical não forneceu resultados

satisfatórios, pois esta é não-estável. Um dos problemas básicos do melhoramento utilizando RV é que, apesar dessa resistência proporcionar completo controle da doença, novas raças do patógeno capazes de superar a resistência podem aparecer. A durabilidade da RV é sempre relacionada com a extensão do uso comercial do cultivar resistente. Tratando-se de cultura perene, o processo de melhoramento é demorado e os cultivares que tiverem perdido a resistência não podem ser rapidamente substituídos.

Devido ao cafeeiro ser uma planta perene cultivada em extensas áreas que oferecem condições permanentes para a ocorrência de epidemias, pressupõe-se que a exploração da RH, por princípio mais estável, poderia ser considerada como forma de controle genético mais segura, pois não exercerá pressão seletiva intensa sobre o patógeno, reduzindo, assim, a possibilidade deste expressar sua variabilidade pelo aparecimento de novas raças fisiológicas. **H. vastatrix** possui variabilidade para virulência. Isto foi demonstrado no Brasil, onde, após 16 anos de presença do patógeno, mais de dez raças já haviam sido identificadas (CARDOSO, 1986).

Segundo BETTENCOURT (1981), há grande interesse no aproveitamento da RH isoladamente ou associada a fatores de RV. Segundo ROBINSON (1971), a RV tem maior probabilidade de ser efetiva quando reforçada por altos níveis de RH. Alguns híbridos como o 'Icatu' e o HDT (Híbrido de Timor) poderiam apresentar, além de RV complexa, altos níveis de RH (CHAVES, 1976). Em plantas susceptíveis, derivadas de HDT e **C. canephora**, tem-se procurado identificar manifestações de RH (CHAVES, 1976). Assim, a existência de fontes de altos níveis da mesma, como foi demonstrado por ABREU (1978 e 1988), FONSECA (1979), ALMEIDA (1980) e ESKES (1983), tornou esses materiais extremamente valiosos em programas de melhoramento.

O processo de seleção de híbridos interespecíficos de cafeeiros é lento e dispendioso, sendo necessário estudar uma população numerosa para se poder encontrar indivíduos que reúnam a maioria dos caracteres desejados. As avaliações de componentes individuais de resistência consomem muito tempo, limitando a quantidade de plantas avaliadas, não sendo prática a sua utilização rotineira. Além disso, segundo ESKES (1983), existem dois problemas no uso de avaliações de componentes de resistência para determinar resistência em populações: o grau de associação entre componentes e a extrapolação de avaliações monocíclicas usualmente feitas em casa de vegetação e em "seedlings".

Os componentes de resistência são bem correlacionados (coeficiente de correlação momento-produto de Pearson) entre si (ESKES, 1983). Neste estudo PI, PL e ESP apresentaram baixas correlações com SEV ($r = -0,09$; $r = -0,19$; e $r = 0,28$, respectivamente) e altas correlações entre si (PI e PL, $r = 0,86$; PI e ESP, $r = -0,90$; e PL e ESP, $r = -0,78$). No estudo de resistência horizontal de progênies de 'Catimor' (seção 4.4.), PI foi altamente correlacionado com PLm (período latente médio) ($r = 0,95$). As correlações de SEV com PI e PLm foram $r = -0,65$ e $r = -0,58$, respectivamente. Este primeiro problema pode ser contornado pela utilização das análises multivariadas que consideram a correlação entre os componentes de resistência: análise de componentes principais, análise canônica e "cluster análise" (baseada na distância de Mahalanobis). O segundo problema é mais complexo sendo discutido adiante.

Nos estudos de ESKES (1983), em geral, PL e ESP foram correlacionados com TR (tipo de reação; neste estudo avaliado como grau de doença, utilizando uma escala de notas, Quadro 4). Por definição, TR é correlacionado com ESP. Portanto, segundo ESKES (1983) e KUSHALAPPA e ESKES (1989), a maioria da variação para resistência horizontal (denominada pelos autores de resistência incompleta) poderia ser simplesmente avaliada pelo TR. Neste estudo, houve concordância entre os resultados obtidos pela análise multivariada dos componentes de resistência e o grau de doença dos genótipos, avaliado por meio de uma escala de notas. Além disso, os componentes de resistência mais importantes foram PL e ESP (Quadro 14), porém, mais estudos são necessários para que se possa recomendar a substituição da avaliação de alguns dos componentes pelo TR (ou grau de doença).

Além disso, o TR não mensura o período infeccioso, o qual é um importante componente de resistência a ser considerado no patossistema *Coffea - H. vastatrix*, avaliado como o período de retenção foliar (PRF, número de dias entre a esporulação e a queda da folha), sendo um dos principais fatores que determinam a duração da esporulação (CADENA-GÓMEZ e BURITICÁ-CÉSPEDES, 1980). Linhagens mais susceptíveis de *C. arabica* apresentaram um PRF maior em ensaios em casa de vegetação (ESKES e CARVALHO, 1983 e ESKES, 1983). sendo o PRF negativamente correlacionado com a densidade de lesões. O PRF tem sido considerado importante na explicação de diferenças no nível de doença no campo, entre variedades altamente susceptíveis (RIBEIRO et alii, 1981; ESKES, 1983 e BERGAMIN FILHO et alii, 1990). Um maior

PRF aumenta a quantidade de inóculo disponível, sendo de especial importância para manter o inóculo durante a estação seca (ESKES e CARVALHO, 1983).

Segundo RIBEIRO et alii (1981) e BERGAMIN FILHO et alii (1990), a tendência de alguns cultivares não reterem folhas infectadas por muito tempo é uma característica não-específica de raça, constituindo um tipo não-usual de RH, caracterizado por um atraso no início da epidemia. Segundo ESKES (1983), se o PRF for utilizado como critério de **seleção**, os genótipos com PRF muito longo ou muito curto deveriam ser evitados, pois um longo PRF pode estar relacionado com altos níveis de doença no campo, e curtos PRF podem estar relacionados com baixa produtividade. A duração da esporulação também pode ser afetada pela necrose da **lesão**, a qual pode ser induzida pelo ambiente, pelo **genótipo** do cafeeiro ou simplesmente pela idade da **lesão** (ESKES, 1983).

A eficiência da avaliação da resistência horizontal ou parcial em "seedlings" pode variar com o patossistema. Segundo POYNTZ e HYDE (1987), a avaliação da resistência parcial de trigo a **Puccinia recondita** em "seedlings" é inadequada, porém, segundo PARLEVLIIET et alii (1980), a **seleção** de cevada para resistência parcial a **Puccinia hordei** foi mais eficiente no estágio de "seedlings". Embora seja comum a avaliação da RV de genótipos de cafeeiros à ferrugem, em "seedlings", a predição da RH a ferrugem no campo, pelas avaliações monocíclicas em casa de vegetação, pode não ser adequada por causa da influência da produtividade (ESKES e CARVALHO, 1983 e ZAMBOLIM et alii, 1985b) e do PRF na intensidade da ferrugem. De acordo com a população de **C. arabica** considerada, nem sempre os componentes de resistência avaliados em casa de vegetação **são**, significativamente, correlacionados com o nível de doença no campo, o qual, em alguns casos, tem o PRF e a produção como melhores indicadores (ESKES e CARVALHO, 1983).

As conseqüências epidemiológicas da resistência horizontal do cafeeiro a **H. vastatrix** têm sido pouco exploradas. Mais estudos **são** necessários, com uma ampla gama de genótipos de cafeeiros e raças do patógeno, para avaliar as redundâncias entre os componentes de resistência e a real importância de cada componente na explicação de diferenças no nível de doença no campo. As técnicas de análise **multivariada** podem ser úteis nestes estudos.

4.3. Avaliação da Resistência de Progênes de 'Catimor' a Oito Raças Fisiológicas de *Hemileia vastatrix*, Utilizando "Cluster Análise"

"Cluster analise" permitiu avaliar a similaridade das progênes de 'Catimor' quanto a resistência a ferrugem, considerando simultaneamente os graus de doença relativo a oito raças do patógeno, inoculadas por TAMAYO (1988). O método de agrupamento de Tocher formou um grupo com todas as progênes e o cultivar 'Catuaí' consistiu, isoladamente, um outro grupo, mas o método do vizinho mais próximo revelou quatro grupos (Figura 8). cujas médias dos graus de doença *são* apresentadas no Quadro 14.

Em decorrência da alta susceptibilidade do cultivar 'Catuaí', a sua dissimilaridade em relação as progênes de 'Catimor' foi alta. As progênes UFV 6489 e UFV 6613 (grupos B e C, respectivamente- Quadro 14) apresentaram graus de doença ligeiramente superiores a média das demais, classificando-se em dois grupos diferentes, pelo método do vizinho mais próximo. Porém, em nível de 80% de similaridade, as progênes de 'Catimor' constituíram um só grupo (Figura 8).

QUADRO 14 - Grau Médio de Doença dos Grupos, Formados por Cluster Análise" (Método do Vizinho Mais Próximo), a Oito Raças Fisiológicas de *H. vastatrix*

Grupo	Raças							
	II	VII	XV	X	XXIV	XXIII	XXXI	XVI
A	4,8	4,8	4,2	5,0	4,2	5,2	4,0	5,8
B	2,8	3,2	2,8	3,2	2,8	3,0	2,4	2,8
C	2,8	2,4	3,0	2,0	2,4	3,0	1,8	2,8
D	2,1	2,4	2,2	2,0	2,2	2,2	2,1	2,4

* A: UFV2144 ('Catuaí'); B: UFV6489; C: UFV6613; e D: demais progênes.

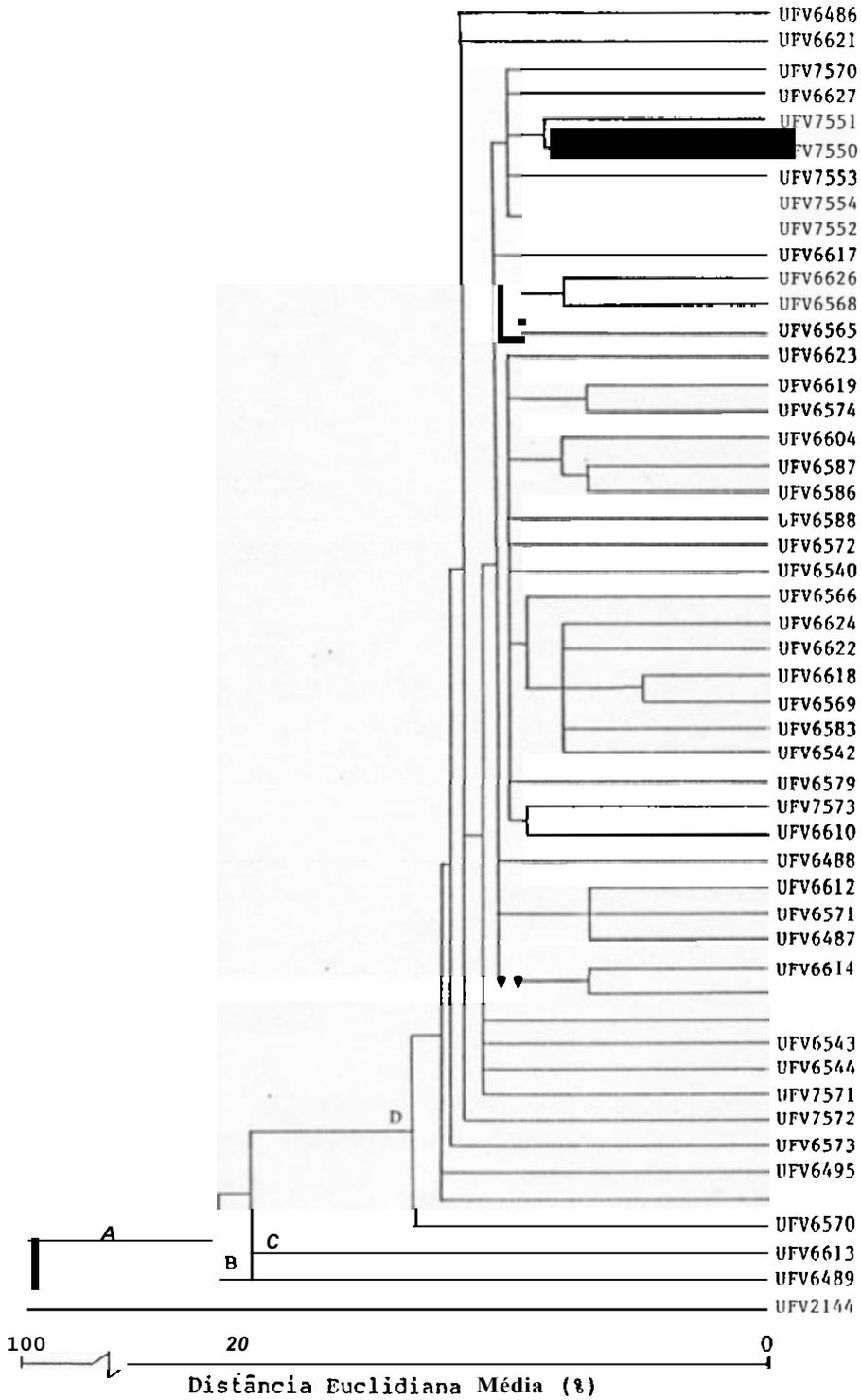


FIGURA 8 - Agrupamento pelo Método do Vizinbo Mais Próximo de 49 Progênie de 'Catimor' e o Cultivar 'Catuaí' quanto a Resistência a Oito Raças de *Hemileia vastatrix*.

A resistência do tipo vertical em cafeeiros é condicionada por fatores dominantes. Foram identificados nove genes de efeitos maiores, dominantes, *os* quais isoladamente ou associados, condicionam o espectro de resistência de *Coffea* spp. a ferrugem. Os genes S_H1 , S_H2 , S_H4 e S_H5 parecem estar associados unicamente a *C. arabica*. O gene S_H3 , encontrado apenas nos arabicas originários da Índia, e o S_H6 foram introduzidos em *C. arabica*, provavelmente, por hibridação com *C. liberica* e *C. canephora*, respectivamente (BETTENCOURT e CARVALHO, 1968). Os genes S_H6 , S_H7 , S_H8 e S_H9 são originários de *C. canephora* (BETTENCOURT e RODRIGUES JR., 1988). A população de 'Catimor' contém alguns genes de resistência de *C. canephora*, os quais diferem de S_H1 a S_H5 e provêm resistência a algumas ou todas as raças de *H. vastatrix* conhecidas (RODRIGUES JR. et alii, 1975). 'Catimor' é um híbrido derivado do cruzamento de 'Caturra' (cultivar comercial de *C. arabica*) com o híbrido interespecífico tetraplóide HDT (Híbrido de Timor), o qual é o progenitor mais utilizado para introduzir resistência em variedades comerciais, possuindo cinco genes dominantes (mas *seu* genótipo não é totalmente conhecido: S_H5 , S_H6 , S_H7 , S_H8 , S_H9 e ?), condicionando individualmente ou associados, a resistência a ferrugem. O HDT é constituído por uma população heterogênea, provavelmente derivada de cruzamento natural entre *C. arabica* e *C. canephora*, e inclui plantas do grupo fisiológico A (cerca de 95%, as quais *são* resistentes a todas as raças do patógeno e têm mantido a sua resistência por algumas décadas), R e E.

TAMAYO (1988) dividiu as progênies em três grupos, baseado no grau médio de doença e seu intervalo de confiança. Baseado apenas nas médias dos graus de doença, conclui-se que todas as progênies apresentaram resistência completa (resistência que impede a reprodução do patógeno) a todas as raças (equivalente ao grau de doença inferior a quatro na escala de TAMAYO, 1988), porém, as progênies UFV 6489 e UFV 6613 apresentaram segregação para a susceptibilidade a algumas raças, a qual é evidenciada pelo intervalo de confiança do grau médio de doença (TAMAYO, 1988).

Em um estudo como *este*, em que todas as variáveis (reações as diferentes raças do patógeno) não *são* obtidas da mesma planta, não é possível utilizar uma técnica multivariada (como a análise canônica ou "cluster análise". utilizando a distância de Mahalanobis) que considere a variabilidade amostral.

A análise das interações hospedeiro-patógeno permitiu reconhecer um sistema de relação do tipo "gene-a-gene" (FLOR, 1942 e 1955 e PERSON, 1959) mono ou oligogênico, que caracteriza a resistência específica de *Coffea* ssp. a *H. vastatrix* (NORONHA-WAGNER e BETTENCOURT, 1967). Assim, foi possível conhecer os genótipos dos cafeeiros diferenciadores e inferir sobre os genótipos das raças de *H. vastatrix* estudadas. De 32 raças já detectadas, são conhecidos os genótipos de 24 raças (RODRIGUES JR. et alii, 1975 e RODRIGUES JR., 1990). A semelhança entre as raças de *H. vastatrix*, em relação ao grau de doença em 49 progênies de 'Catimor' e o cultivar 'Catuai', foi avaliada por "cluster análise", utilizando o método do vizinho mais próximo. As raças XXIII e XXIV foram as duas mais similares entre si, sendo agrupadas a uma distância $DE_m = 0,283$ a qual corresponde a 69,2% da distância máxima do dendograma (Figura 9); portanto, aparentemente haveria grandes diferenças entre as raças. Segundo TAMAYO (1988), qualquer uma das oito raças poderia ter sido utilizada na avaliação da resistência das progênies de 'Catimor', pois não houve interação significativa entre raças e progênies.

A distância euclidiana é zero para indivíduos (neste estudo, raças) iguais, aumentando o seu valor de acordo com o aumento da dissimilaridade entre dois indivíduos. Na avaliação da dissimilaridade entre raças, o maior valor da distância euclidiana média (DE_m) foi $DE_m = 0.41$. A sua interpretação é difícil, pois, a princípio, não é possível afirmar se este valor indica alta ou baixa dissimilaridade. Essa é uma desvantagem da avaliação de similaridade pelas medidas de distância. Quando são utilizados coeficientes de associação (Seção 2.4.1.2.1.), cujo valor máximo possível é a unidade, esse problema não ocorre; portanto, é recomendável a comparação dos resultados de "cluster análise", utilizando medidas de distância, com os de outros métodos de análise.

Na avaliação da resistência das progênies de 'Catimor' a oito raças de *H. vastatrix*, TAMAYO (1988) utilizou escala de notas de um a seis (Quadro 4). Portanto, nesse estudo, a maior DE_m matematicamente possível entre duas raças é $DE_m = 5$. Pode-se afirmar, então, que o valor $DE_m = 0.41$ indica baixa dissimilaridade. Para exemplificar, incluiu-se uma raça hipotética (raça H), cujo grau de doença para todas as progênies seria o máximo da escala utilizada na avaliação (grau seis). O agrupamento pelo método do vizinho mais próximo, incluindo a raça H (Figura 10), permitiu verificar a existência de alta similaridade entre as oito raças inoculadas por TAMAYO (1988), concordando com os resultados desse autor. O método de Tocher dividiu as oito raças em

três grupos, porém, quando se adicionou a raça hipotética H a análise, as oito raças inoculadas constituíram um único grupo, diferente da raça H, concordando com os resultados do método do vizinho mais próximo.

"Cluster análise" foi utilizada por LEBEDA e JENDRULEK (1987b), considerando a presença ou ausência de cada um de 11 fatores de virulência, demonstrando haver associação destes fatores com a origem geográfica de raças ou isolados de *Bremia lactucae*. Neste estudo, não houve associação entre os fatores de virulência das raças de *H. vastatrix* e a formação de grupos: contudo, melhor avaliação da similaridade das raças seria obtida em estudos que considerassem dados de isoenzimas e. ou. de DNA das raças e reação de virulência destas a diversos genótipos de cafeeiros, como realizado por HELLMAN e CHRIST (1991) com isolados de *Ustilago hordei*; por GOTTWALD et alii (1991). com strains de *Xanthomonas campestris* pv. *citrumelo*; e por LINDE et alii (1990). com isolados de *Uromyces appendiculatus*.

Em fitopatologia é mais comum a utilização de "cluster análise" na avaliação da similaridade entre isolados patogênicos que entre raças. Em um patossistema quantitativo, em que não ocorre raças fisiológicas segundo a sua definição clássica (ROBINSON, 1969). "cluster análise" pode ser utilizada para avaliar a similaridade entre isolados, baseada em seus padrões de virulência a uma gama de genótipos hospedeiros, como realizado por SCHILDER e BERGSTROM (1990) e SAH e FEHRMANN (1992). Nesses casos, a análise pode selecionar os isolados mais apropriados para serem utilizados em programas de melhoramento.

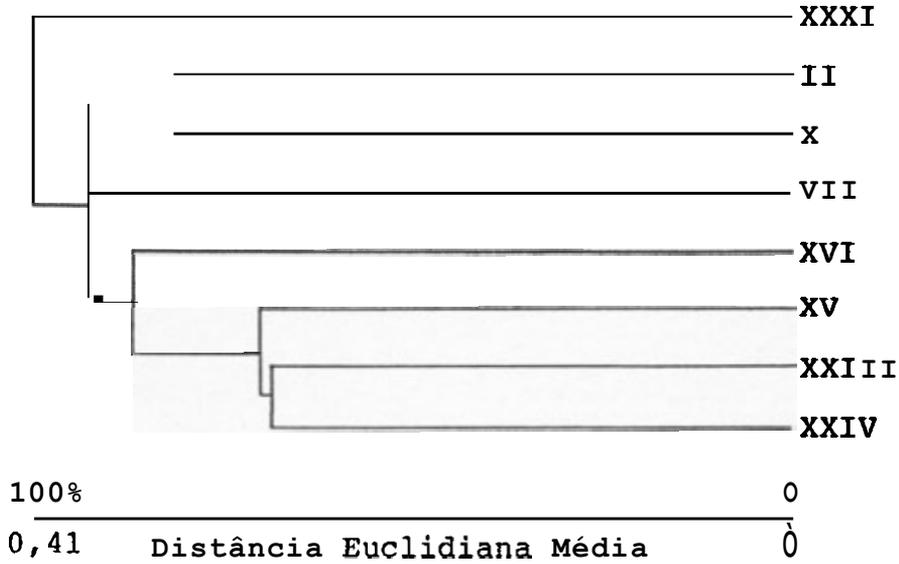


FIGURA 9 - Agrupamento pelo Método do Vizinho Mais Próximo de Oito Raças de *Hemileia vastatrix* Quanto ao Grau de Doença em 49 Progenies de 'Catimor' e no Cultivar 'Catuaí'.

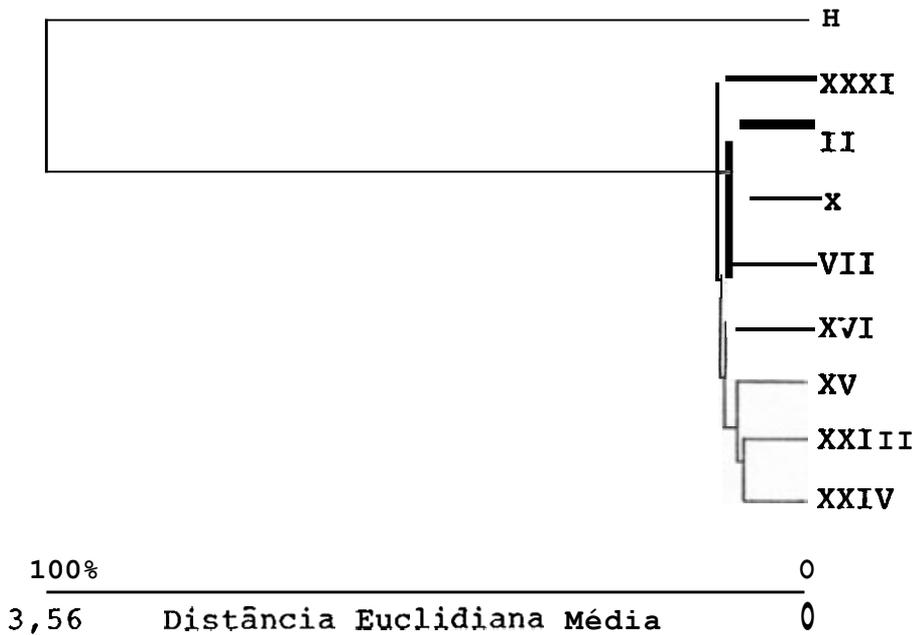


FIGURA 10 - Agrupamento pelo Método do Vizinho Mais Próximo de Oito Raças de *Hemileia vastatrix* e uma Raça Hipotética (Raça H) Quanto ao Grau de Doença em 49 Progenies de 'Catimor' e no Cultivar 'Catuaí'.

4.4. Utilização da Análise de Componentes Principais e "Cluster Análise" no Estudo da Resistência Horizontal de Progênes de 'Catimor' a *Hemileia vastatrix*

"Cluster análise" forma grupos com maior similaridade dentro que entre grupos. Baseado no dendograma resultante de "cluster análise" utilizando o método do vizinho mais próximo e a distância euclidiana média (Figura 11), o cultivar 'Catuai' constituiu um grupo e as progênes de 'Catimor' foram divididas em três grupos, cujas médias dos componentes de resistência *são* apresentadas no Quadro 15. As progênes UFV 3658, UFV 4303 e UFV 4305 constituíram o grupo mais resistente, apresentando maiores período de incubação (PI) e período latente médio (PLm) e menores número de **pústula** por folha (NPF) e severidade (SEV). O método de Tocher não revelou diferenças entre as progênes estudadas, porém, distinguiu-as do cultivar 'Catuai'.

A análise de componentes principais proporcionou simplificação na interpretação dos dados, permitindo o agrupamento das progênes mais similares, mediante exames visuais em dispersão gráfica no espaço bidimensional. Os dois primeiros componentes principais contêm quase toda a informação (98,7% da variância total) (Quadro 16). Os dois últimos, contendo 1,1 e 0,2% da variância total, praticamente não possuem significado biológico. A ordenação das progênes de 'Catimor' na análise de componentes principais (Figura 12) segue o mesmo padrão dos resultados de "cluster análise", utilizando o método de agrupamento do vizinho mais próximo (Figura 11). Os resultados das análises **multivariadas** permitiram constatar diferentes níveis de resistência horizontal entre as progênes de 'Catimor'.

A hipótese de existência de dois tipos de resistência de plantas a patógenos, resistência vertical (RV) e horizontal (RH) foi defendida por VAN DER PLANK (1963). A RV é caracterizada pela interação diferenciada entre o hospedeiro e o patógeno, ou seja, é dirigida contra algumas raças do patógeno, mas não outras, resultando na imunidade somente contra as raças nas quais atua. A RH, ao contrário, age contra todas as raças do patógeno, apesar de haver susceptibilidade **as mesmas**, causando uma **variação** nos componentes de resistência, ou seja, a RH aumenta os período de incubação e período latente, diminui a taxa de esporulação, a frequência de infecção e o período infeccioso, reduzindo, assim, o progresso da doença no campo (PARLEVLIET, 1979).

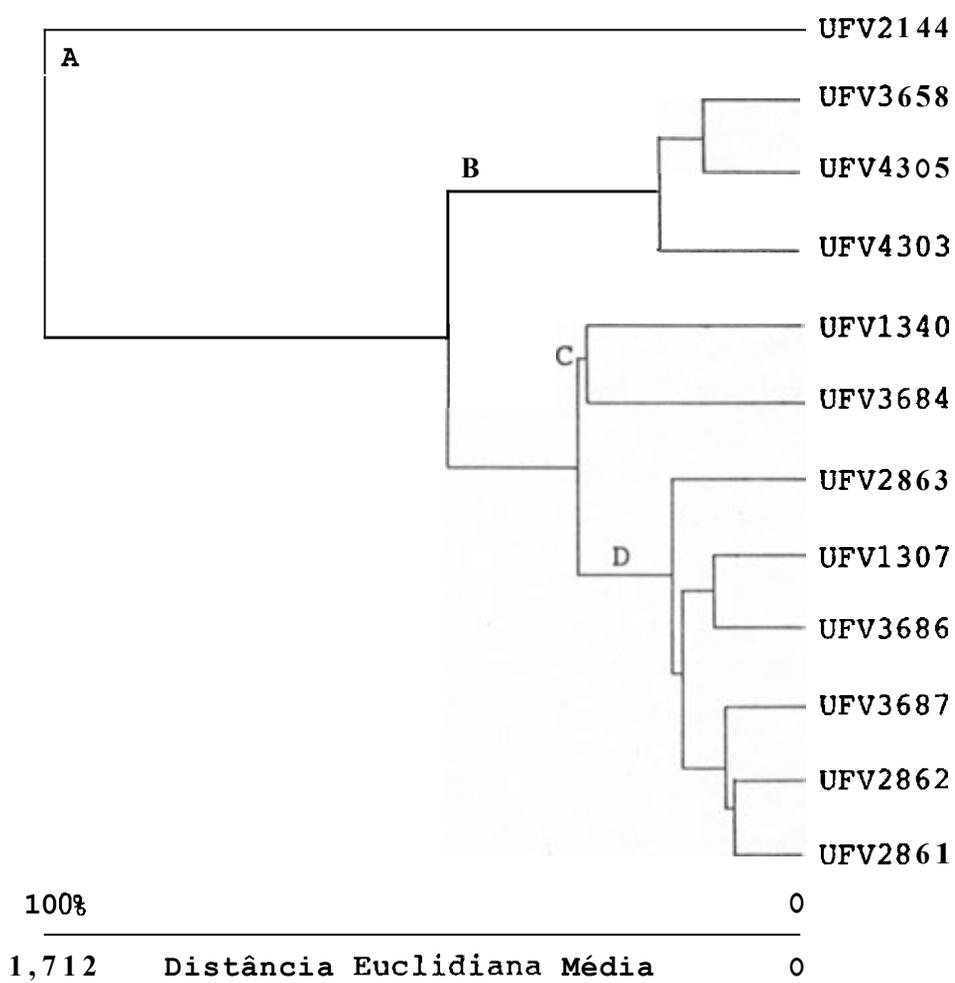


FIGURA 11 - Agrupamento pelo Método do Vizinho Mais Próximo, de Progenies de 'Catimor' quanto à Resistência Horizontal a *H. vastatrix*.

Quadro 15 - Valores Médios dos Componentes de Resistência dos Grupos de Cafeeiros Formados por "Cluster Analise" (Método do Vizinho Mais Proximo)

Grupo	Componentes de Resistência			
	PI	PLm	NPF	SEV
A UFV2144 (Catuaí)	23,2	32,0	9,1	12,0
B UFV3658; UFV4303; UFV4305	35,6	48,6	1,7	1,8
C UFV1340; UFV3684	26,1	35,3	3,9	4,1
D UFV1307; UFV2861; UFV2862; UFV2863; UFV3686; UFV3687	29,4	38,4	2,1	2,2

* PI e PLm: período de incubação e período latente médio (dias); NPF: número de pústula por folha; e SEV: severidade (%).

QUADRO 16 - Estimativas das Variâncias (Autovalores, λ_j) Associadas aos Componentes Principais e Respective Coeficientes de Ponderação (Autovetores) de Quatro Componentes de Resistência Avaliados em 11 Progenies de 'Catimor' e o Cultivar 'Catuaí'

λ_j	Variância	Z_1 *	Z_2	Z_3	Z_4
	acumulada (%)				
		Componentes Principais			
3,2289	80,72	- 0,5092	- 0,4829	0,5084	0,4991
0,7190	98,70	0,4416	0,5633	0,4726	0,5141
0,0443	99,80	- 0,7099	0,6530	0,1331	- 0,2281
0,0079	100,00	0,2045	- 0,1521	0,7074	- 0,6592

* Z_1 , Z_2 , Z_3 e Z_4 correspondem as variáveis período de incubação, período latente médio, número de pústulas por folha e severidade, todas padronizadas com variância igual a unidade, respectivamente.

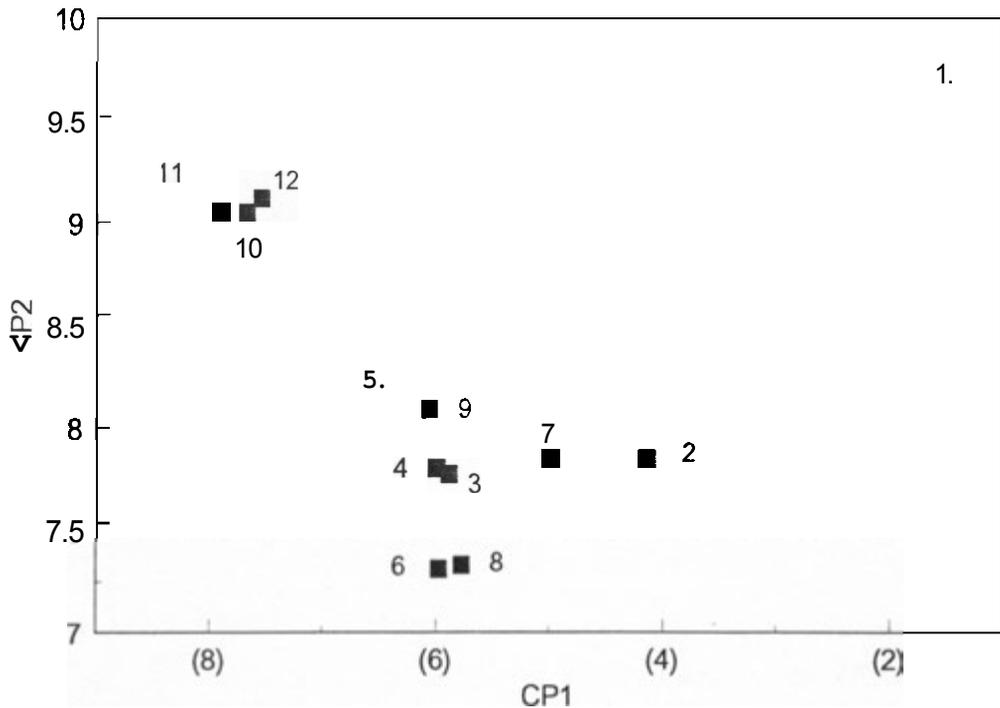


FIGURA 12 - Dispersão Gráfica de 11 Progênies de 'Catimor' e do Cultivar 'Catuaí', em Relação aos Dois Primeiros Componentes Principais (CP1 e CP2) (1 - UFV2144 ('Catuaí'); 2 - UFV1340; 3 - UFV2861; 4 - UFV2862; 5 - UFV2863; 6 - UFV1307; 7 - UFV3684; 8 - UFV3686; 9 - UFV3687; 10 - UFV3658; 11 - UFV4303; e 12 - UFV4305).

Há muito tempo procura-se criar cultivares resistentes a ferrugem (*Hemileia vastatrix*), objetivando dispensar total ou parcialmente o controle químico e substituir de maneira vantajosa os cultivares tradicionais de *Coffea arabica*, altamente susceptíveis. Por um longo tempo os trabalhos realizados sobre a resistência do cafeeiro à ferrugem, trataram da RV. A tentativa de controle da doença pela obtenção de cultivares com RV não forneceu resultados satisfatórios, pois, apesar da RV proporcionar um completo controle da doença, novas raças de *H. vastatrix* capazes de superar a resistência podem surgir. A durabilidade da resistência é relacionada com a extensão do uso comercial do cultivar resistente.

Segundo NELSON (1975), a natureza poligênica da RH é responsável pela estabilidade relativa da resistência por longo período de tempo, por causa de raças com genes necessários para vencê-la serem menos prováveis de surgirem. A perda da RH, quando ocorre, é pequena e raramente constitui perda completa. Segundo ROBINSON (1971), a RV tem maior probabilidade

de ser efetiva quando reforçada por altos níveis de RH. A demonstração que plantas susceptíveis derivadas do HDT (Híbrido de Timor) ou *C. canephora* possuem altos níveis de RH (ABREU, 1978 e 1988; FONSECA, 1979; ALMEIDA, 1980 e ESQUES, 1983) tornou **esses** materiais valiosos em programas de melhoramento.

A avaliação da RH é lenta e trabalhosa, o que limita a quantidade de plantas avaliadas. Além disso, **os** componentes **de** resistência são correlacionados entre si, o que pode gerar informações redundantes. Quanto maior for o **número** de componentes de resistência e de **genótipos** avaliados maior será o auxílio das análises **multivariadas** na interpretação dos resultados.

Além de possibilitar a avaliação da similaridade entre as progênies, a análise de componentes principais permite identificar **os** componentes de resistência que pouco contribuem para a divergência entre as progênies por serem redundantes devido a serem altamente correlacionados com outros. A identificação dos componentes de resistência redundantes é desejável a fim de facilitar a avaliação da RH. A supressão de um componente de resistência que é altamente correlacionado com outro não provocará alterações significativas na análise, Nesses casos seria preferível manter **os** componentes de resistência mais fáceis de avaliar.

As correlações (coeficiente de correlação momento-produto de Pearson) do PI com PLm, SEV e NPF foram $r = 0,95$, $r = -0,65$ e $r = -0,69$, respectivamente. As correlações de PLm com SEV e NPF foram $r = -0,58$ e $r = -0,60$, respectivamente. A correlação entre NPF e SEV foi $r = 0,99$. Segundo **os** resultados da análise de componentes principais (Quadro 16), **os** componentes de resistência NPF e PI poderiam ser descartados pois possuem **os** maiores coeficientes de ponderação absolutos associados aos componentes principais com **variância** inferior a 0,7 (JOLLIFFE, 1972 e 1973). Para a recomendação de descarte dessas variáveis em experimentos futuros, **são** necessários mais estudos envolvendo grande número de **genótipos** de cafeeiros e raças de *Hemileia vastatrix*, assim poder-se-ia avaliar **se** as correlações entre **os** componentes de resistência independem da combinação entre o **genótipo** do hospedeiro e a raça do **patógeno**. As técnicas de ordenação **e**, ou, **agrupamento** também seriam **úteis** no estudo da resistência horizontal de **genótipos** conhecidos de cafeeiros, com o objetivo de determinar **se** existe correlação entre o padrão de agrupamento **e**, ou, ordenação dos **genótipos** em relação aos genes de efeitos maiores de resistência.

5. RESUMO E CONCLUSÕES

O presente trabalho teve como objetivo **ênfatizar** a importância, os princípios, as aplicações em fitopatologia e as limitações das seguintes técnicas estatísticas de análise **multivariada**: "cluster análise", análise **canônica** e análise de componentes principais, **exemplificando o seu** uso em dois estudos: avaliação do progresso da ferrugem (***Tranzschelia* sp.**) em cultivares de pessegueiro (***Prunus persica***) e avaliação da **resistência** do cafeeiro (***Coffea arabica***) a ferrugem (***Hemileia vastatrix***).

"Cluster análise" foi utilizada para avaliar a similaridade entre 11 cultivares de pessegueiro, com relação ao progresso da ferrugem. Embora **os** grupos de cultivares formados tenham diferido com relação a medida de similaridade e ao método de agrupamento utilizado, os resultados de "cluster análise" apresentaram concordância com os resultados de AACPD e com os dos testes de identidade dos modelos de progresso de doença. "Cluster análise", baseada na distância **euclidiana** média e no método de agrupamento de Tocher, apresentou melhor concordância com os resultados dos testes de identidade de modelos, tendo classificado o cultivar '**Flordasun**' como um grupo (o mais **susceptível**), '**Marli**' como um outro grupo (o mais resistente), '**Princesa**', '**Vila Nova**' e '**Pérola de Itaquera**' como um terceiro grupo (considerado como susceptível) e '**Diamante**'. '**Cerrito**', '**Rei da Conserva**', '**Topázio**', '**Sinuelo**' e '**Premier**' em um quarto grupo (considerado moderadamente resistente).

"Cluster análise" revelou-se um método eficiente na avaliação da similaridade entre curvas de progresso da ferrugem do pessegueiro, podendo ser utilizada para comparação de epidemias, em complementação as técnicas tradicionais, sendo de grande utilidade, principalmente, quando o estudo envolver grande número de cultivares.

Avaliou-se a similaridade entre oito genótipos de cafeeiros com relação a quatro componentes de resistência, período de incubação (PI), período latente (PL), severidade (SEV) e esporulação (ESP), em mudas inoculadas com a raça II de *H. vastatrix*, em um experimento conduzido em casa de vegetação. Houve concordância entre os resultados obtidos pela análise canônica e por "cluster análise". O genótipo H472 constituiu um grupo (o mais resistente); H442 e H481 constituíram um segundo grupo; H415 e UFV 2143 ('Catuaí'), um terceiro; e H471 e UFV 2150 ('Mundo Novo'), um quarto grupo. Os últimos três grupos foram considerados susceptíveis, enquanto H543 apresentou-se imune a raça II.

Houve concordância entre os resultados das análises multivariadas utilizadas, baseadas nos componentes de resistência, e o intervalo de confiança do grau médio da ferrugem do cafeeiro, avaliado por uma escala de 1 a 6, compreendendo os tipos de reação.

PL e ESP foram os componentes de resistência que mais contribuíram para a divergência entre os genótipos (34,49 e 30,84%, respectivamente). SEV e PI contribuíram com 25,93 e 8,73%, respectivamente.

As técnicas multivariadas utilizadas foram eficientes no estudo da resistência horizontal do cafeeiro à ferrugem, em razão de considerar as correlações entre os componentes de resistência e avaliar a importância de cada um deles.

A partir de dados disponíveis na literatura (TAMAYO, 1988), utilizou-se "cluster análise" na avaliação da resistência de 49 progênies de 'Catimor' a oito raças fisiológicas de *H. vastatrix*. "Cluster análise" permitiu avaliar a similaridade entre as progênies, considerando simultaneamente as reações a oito raças do patógeno. O método de agrupamento de Tocher agrupou todas as progênies em um único grupo e a testemunha 'Catuaí' em um outro grupo. O método de agrupamento do vizinho mais próximo separou as progênies em três grupos, todos com grande dissimilaridade em relação a um quarto grupo constituído pela testemunha, o cultivar 'Catuaí'.

A partir de dados disponíveis na literatura (ALMEIDA, 1980), utilizaram-se as análises de componentes principais e "cluster análise" no estudo da resistência horizontal de progênies de 'Catimor' a ferrugem. Avaliou-se a similaridade entre 11 progênies segregantes para a susceptibilidade, quanto a quatro componentes de resistência: PI, PLm (período latente médio), SEV e NPF (número de pústulas por folha).

"Cluster análise", quando baseada no método de agrupamento de Tocher, não revelou diferenças entre as progênies de 'Catimor', porém distinguiu-as todas da testemunha. O cultivar 'Catuaí', e, quando baseada no método de agrupamento do vizinho mais próximo, apresentou quatro grupos: o primeiro grupo constituído por 'Catuaí' (o mais susceptível); o segundo grupo (o mais resistente). pelas progênies de 'Catimor' UFV3658, UFV4303 e UFV4305; o terceiro grupo, pelas progênies UFV2861, UFV2862, UFV2863 UFV1307, UFV3686 e UFV3687. O quarto grupo foi constituído pelas progênies UFV1340 e UFV 3684.

Houve concordância entre os resultados obtidos por "cluster análise" baseada no método de agrupamento do vizinho mais próximo e os resultados obtidos pela análise de componentes principais. Segundo estes últimos, PI e NPF podem ser excluídos, pois são redundantes. PI e PLm foram altamente correlacionados ($r = 0.95$). assim como SEV e NPF ($r = 0.99$).

Adicionalmente, foram descritos a importância, objetivos e aplicações em fitopatologia, de outras técnicas de análise multivariada: "path análise" (análise de trilha), análise de correlação canônica, análise fatorial, análise discriminante, MANOVA e análise de coordenada principal.

BIBLIOGRAFIA

BIBLIOGRAFIA

- ABREU, M.S. Identificação de parâmetros para avaliação da resistência horizontal de *Coffea arabica* L. a *Hemileia vastatrix* Berk. et Br. Viçosa - MG, UFV, 1978. 64p. (Tese M.S.)
- ABREU, M.S. Resistência horizontal a *Hemileia vastatrix* Berk. & Br. em cafeeiros descendentes do Híbrido de Timor. Viçosa-MG, UFV, 1988. 68p. (Tese D.S.)
- ADAMS, M.W. & WIERSMA, J.V. An adaptation of principal components analysis to an assessment of genetic distance. Research Report, **347**: 2-7, 1978.
- AKORODA, M.O. Principal component analysis and metroglyph of variation among nigerian yellow yams. *Euphytica*, **32**: 565-73, 1983.
- ALFENAS, A.C.; PETERS, I.; BRUNE, W.; PASSADOR, G.C. Eletroforese de proteínas e isoenzimas de fungos e essências florestais. Viçosa - MG, SIF, 1991. 242p.
- ALMEIDA, L.C. Resistência vertical e horizontal a *Hemileia vastatrix* Berk. et Br. em gerações F4 e F5 de progenies de cafeeiros 'Catimor'. Viçosa - MG, UFV, 1980. 40p. (Tese M.S.)
- AMORIM, L. & BERGAMIN FILHO, A. Análise comparativa de curvas de progresso do carvão em quatro variedades de cana-de-açúcar com o uso de análise de variância multivariada. *Fitopatologia Brasileira*, **17**: 42-8, 1992.
- ANDERSON, W.F.; BEUTE, M.K.; WYNNE, J.C.; WONGKAEW, S. Statistical procedures for assessment of resistance in a multiple foliar disease complex of peanut. *Phytopathology*, **80**: 1451-9, 1990.
- ARNOLD, S.F. The theory of linear models and multivariate analysis. New York, John Wiley & Sons, 1981. 475p.
- ARUNACHALAM, V. Genetic distance in plant breeding. *Indian Journal of Genetics*, **41**: 226-36, 1981.
- BAGNALL, R.H. & TAI, G.C.C. Field resistance to potato virus Y in potato assessed by cluster analysis. *Plant Disease*, **70**: 301-4, 1986.

- BEALE, E.M.L.; KENDALL, M.G.; MANN, D.W. The discarding of variables in multivariate analysis. *Biometrika*, **54**: 357-65, 1967.
- BENIGNI, R. & GIULIANI, A. Cluster analysis of short-term tests: a new methodological approach. *Mutation Research*, **147**: 139-51, 1985.
- BENIN, M.; GASQUEZ, J.; MAHFOUDI, A.; BESSIS, R. caractérisation bioquimiche des cépages de *Vitis vinifera* L. par électrophorèse d'isoenzymes foliaires: Essai de classification des variétés. *Vitis*, **27**:157-72, 1988.
- BERGAMIN FILHO, A.; SALGADO, C.L.; FEGIES, N.C.; RIBEIRO, I.J.A. Horizontal resistance in three *Coffea arabica* cultivars to *Hemileia vastatrix*. *Fitopatologia Brasileira*, **15**: 308-13, 1990.
- BETTENCOURT, A.J. Melhoramento genético do cafeeiro. **Transfêrencia de fatores de resistência a *Hemileia vastatrix* Berk. et Br. para as principais cultivares de *Coffea arabica* L.** Lisboa, Instituto Superior de Agronomia da Universidade Técnica de Lisboa, 1981. 93p. (Tese D.S.)
- BETTENCOURT, A.J. & CARVALHO, A. Melhoramento visando a resistência do cafeeiro à ferrugem. *Bragantia*, **27**: 35-68, 1968.
- BETTENCOURT, A.J. & RODRIGUES JR., C.J. Principles and practice of coffee breeding for resistance to rust and other diseases. In: CLARK, R.J. & MACRAE, R. *Coffee agronomy*, London, Elsevier, 1988. v.4, p. 199-235.
- BONDE, M.R.; PETERSON, G.L.; DOWLER, W.M.; MAY, B. Isozyme analysis to differentiate species of *Peronosclerospora* causing downy mildews of maize. *Phytopathology*, **74**: 1278-83, 1984.
- BONDE, M.R.; PETERSON, G.L.; EMMETT, R.W.; MENGE, J.A. Isozyme comparisons of *Septoria* isolates associated with citrus in Australia and the United States. *Phytopathology*. **81**: 517-21, 1991.
- BOWERS, J.H.; SNODA, R.M.; MITCHELL, D.J. Path coefficient analysis of the effect of rainfall variables on the epidemiology of *Phytophthora* blight of pepper caused by *Phytophthora capsici*. *Phytopathology*, **80**: 1439-46, 1990.
- BRUGNEROTTO, M.I.B.; LUZ, E.D.M.N.; SILVA, S.D.V.M. Uma alternativa para caracterização de populações de *Phytophthora* spp.: análise discriminante baseada no quociente das matrizes de correlação. *Fitopatologia Brasileira*, **16**: 63, 1991. (Resumo)
- CADENA-GÓMEZ, G. & BURITICÁ-CÉSPEDES, P. Expresión de resistencia horizontal a la roya (*Hemileia vastatrix* Berk. y Br.) en *Coffea canephora* variedade Conilon. *Cenicafé*, **31**: 3-28, 1980.
- CAMPBELL, N.A. & ATCHLEY, W.R. The geometry of canonical variate analysis. *Systematic Zoology*, **30**: 268-80, 1981.
- CAMPBELL, C.L.; PENNYPACKER, S.P.; MADDEN, L.V. Progression dynamics of hypocotyl rot of snapbean. *Phytopathology*, **70**: 487-94, 1980.
- CAMPBELL, C.L. & MADDEN, L.V. *Introduction to plant disease epidemiology*. New York, John Wiley & Sons, 1990. 532p.
- CAMUSSI, A.; OTTAVIANO, E.; CALINSKI, T.; KACZMAREK, Z. Genetic distances based on quantitative traits. *Genetics*, **111**: 945-62, 1985.

- CARDOSO, R.M.L. Novas raças fisiológicas de *Hemileia vastatrix* Berk. & Br. no Brasil, métodos de identificação e detecção de grupos fisiológicos em cafeeiros derivados de Híbrido de Timor. Viçosa - MG, UFV, 1986. 111p. (Tese M.S.)
- CHAVES, G.M. Melhoramento do Cafeeiro visando a obtenção de cultivares resistentes a *Hemileia vastatrix* Berk. et Br. Revista Ceres, Viçosa - MG, 23: 321-32, 1976.
- CHEN, W.; HOY, J.W.; SCHNEIDER, R.W. Comparisons of soluble proteins and isozymes for seven *Pythium* species and applications of the biochemical data to *Pythium* systematics. Mycological Research, 95: 548-55, 1991.
- CHEN, J.; CHANG, C.J.; JARRET, R.L.; GRAWEL, N. Genetic variation among *Xylella fastidiosa* strains. Phytopathology, 82: 973-7, 1992.
- COOKSEY, D.A. & GRAHAM, J.H. Genomic finger-printing of two pathovars of phytopathogenic bacteria by rare-cutting restriction enzymes and field inversion gel electrophoresis. Phytopathology, 79: 745-50, 1989.
- COSTA, J.C.B.; CRUZ, C.D.; MAFFIA, L.A. Análise de trilha entre variáveis climáticas relacionadas a temperatura, umidade relativa e precipitação pluviométrica e intensidade de vassoura-de-bruxa em *Theobroma cacao* em Altamira-PA. Fitopatologia Brasileira, 18: 347, 1993. (Suplemento)
- CRUZ, C.D. Algumas técnicas de análise **multivariada** no melhoramento de plantas. Piracicaba, ESALQ, 1987. 75p. (mimeografado)
- CRUZ, C.D. **Aplicação** de algumas técnicas **multivariadas** no melhoramento de plantas. Piracicaba, ESALQ, 1990. 188p. (Tese D.S.)
- CURI, P.R. Análise de agrupamento: métodos sequenciais, aglomerativos e hierárquicos. Ciência e Cultura, 35: 1416-29, 1983.
- DAS, P.K. & GUPTA, T.D. Multivariate analysis in black gram (*Vigna mungo* (L.) Hepper). Indian Journal of Genetics, 44: 243-7, 1984.
- DEMÉTRIO, C.G.B. Análise multidimensional para dados de **cana-de-açúcar**. Piracicaba, ESALQ, 1985. 144p. (Tese D.S.)
- D'OLIVEIRA, B. As ferrugens do cafeeiro. Oeiras, Junta de Investigações do Ultramar, Centro de Investigações das Ferrugens do Cafeeiro, s.d. 61p. Separata da Revista do Cafe Português, ■ (4): 5-13; 2 (5): 5-12; 2 (6): 5-13; 2 (7): 9-17; 2 (8): 5-22; 4 (16): 5-15; 1954-1957. (Separata nº 3)
- DUNTEMAN, G.H. Introduction to multivariate analysis. Beverly Hills, SAGE, 1984. 237p
- EDWARDS, A.W.F. & CAVALLI-SFORZA, L.L. A method for cluster analysis. Biometrics, 21: 362-75, 1965.
- ENGELBRECHT, M.C. & HATTINGH, M.J. Numerical analysis of phenotypic features of *Pseudomonas solanacearum* strains isolated from tobacco and other hosts in South Africa. Plant Disease, 73: 893-8, 1989.
- ENGLISH, J.T.; THOMAS, C.S.; MAROIS, J.J.; GUBLER, W.D. Microclimates of grapevine canopies associated with leaf removal and control of *Botrytis* bunch rot. Phytopathology, 79: 395-401, 1989.
- ESKES, A.B. Incomplete resistance to coffee leaf rust (*Hemileia vastatrix*). Wageningen, Agricultural University, 1983. 140p. (Tese D.S.)

- ESKES, A.B. & CARVALHO, A. Variation for incomplete resistance to *Hemileia vastatrix* in *Coffea arabica*. *Euphytica*, **32**: 625-37, 1983.
- ESKES, A.B. & COSTA, W.M. Characterization of incomplete resistance to *Hemileia vastatrix* in the Icatu coffee population. *Euphytica*, **32**: 649-57, 1983.
- FACHEL, J.M.G. Análise fatorial. São Paulo, USP, 1976. 81p. (Tese M.S.)
- FERNANDES, T.A.G. & LIMA, J.E. Uso de análise multivariada para identificação de sistemas de produção. Pesquisa *Agropecuária Brasileira*, **26**: 1823-36, 1991.
- FLOR, H.H. Inheritance of pathogenicity in *Melampsora lini*. *Phytopathology*, **32**: 653-69, 1942.
- FLOR, H.H. Host-parasite interaction in flax rusts - its genetics and other implications. *Phytopathology*, **45**: 680-5, 1955.
- FONSECA, S.E.A. Resistência não específica em cultivares de *Coffea arabica* L. e progênies de 'Catimor' a raças de *Hemileia vastatrix* Berk. et Br. Viçosa - MG, UFV, 1979. 42p. (Tese M.S.)
- FRAVEL, D.R.; BENSON, D.M.; BRUCK, R.I. Edaphic parameters associated with shore juniper decline. *Phytopathology*, **73**: 204-7, 1983.
- GATES, C.E. & BILBRO, J.D. Illustration of cluster analysis method for mean separation. *Agronomy journal*, **70**: 462-5, 1978.
- GOODMAN, M.M. The races of maize: I. The use of mahalanobis' generalized distances to measure morphological similarity. *Fitotecnia Latinoamericana*, **4**: 1-22, 1967.
- GOODMAN, M.M. The races of maize: II. Use of multivariate analysis of variance to measure morphological similarity, *Crop Science*, **8**: 693-8, 1968.
- GOTTWALD, T.R.; ALVAREZ, A.M.; HARTUNG, J.S.; BENEDICT, A.A. Diversity of *Xanthomonas campestris* pv. *citrumelo* strains associated with epidemics of citrus bacterial spot in Florida citrus nurseries: correlation of detached leaf, monoclonal antibody, and restriction fragment length polymorphism assays. *Phytopathology*, **81**: 749-53, 1991.
- GOWER, J.C. Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, **53**: 325-38, 1966a.
- GOWER, J.C. A Q-technique for the calculation of canonical variates. *Biometrika*, **53**: 588-90, 1966b.
- GOWER, J.C. A general coefficient of similarity and some of its properties. *Biometrics*, **27**: 857-74, 1971.
- GRAHAM, J.H.; HARTUNG, J.S.; STALL, R.E.; CHASE, A.R. Pathological, restriction-fragment length polymorphism, and fatty acid profile relationships between *Xanthomonas campestris* from citrus and noncitrus hosts. *Phytopathology*, **80**: 829-36, 1990.
- GRAYBILL, F.A. Theory and application of the linear model. Massachusetts. Duxburg Press, 1976. 704p.
- GREGORY, L.V.; AYERS, J.E.; NELSON, R.R. Reliability of apparent infection rates in epidemiological research. *Phytopathologische Zeitschrift*, **100**: 135-42, 1981.
- HAMPTON, R.O. The nature of bean yield reduction by bean yellow and bean common mosaic viruses. *Phytopathology*, **65**: 1342-6, 1975.

- HANSEN, E.M. & HAMM, P.B. Morphological differentiation of host-specialized groups of *Phytophthora megasperma*. *Phytopathology*, **73**: 129-34, 1983.
- HARRIS, R.J. A primer of multivariate statistics. New York, Academic Press, 1975. 332p
- HARTUNG, J.S. & CIVEROLO, E.L. Restriction fragment length polymorphisms distinguish *Xanthomonas campestris* strains isolated from Florida citrus nurseries from *X. C. pv. citri*. *Phytopathology*, **79**: 793-9, 1989.
- HAU, B. & KRANZ, J. Mathematics and Statistics for analysis in epidemiology, In: KRANZ, J. (ed.) Epidemics of plant diseases. Mathematical analysis and modeling. 2.ed. Berlin, Springer-Verlag, 1990. p. 12-52.
- HEADRICK, J.M.; PATAKY, J.K.; JUVIK, J.A. Relationships among carbohydrate content of kernels, condition of silks after pollination and the response of sweet corn inbred lines to infection of kernels by *Fusarium moniliforme*. *Phytopathology*, **80**: 487-94, 1990.
- HELLMANN, R. & CHRIST, B.J. Isozyme variation of physiologic races of *Ustilago hordey*. *Phytopathology*, **81**: 1536-40, 1991.
- HILL, J.H.; BENNER, H.I.; PERMAR, T.A.; BAILEY, T.B.; ANDREWS JR., R.E.; DURAND, D.P.; VAN DEUSEN, R.A. Differentiation of soybean mosaic virus isolates by one-dimensional trypsin peptide maps immunoblotted with monoclonal antibodies. *Phytopathology*, **79**: 1261-5, 1989.
- HILLIG, K.W. & IEZZONI, A.F. Multivariate analysis of a sour cherry germplasm collection. *Journal of American Society of Horticultural Science*, **113**: 928-34, 1988.
- HINDORF, H.Von. *Colletotrichum* - population auf *Coffea arabica* L. in Kenia I. Eine methode zu systematischen trennung von pilzpopulationen. *Phytopathologische Zeitschrift*, **77**: 97-116, 1973.
- HOFFMANN, R. & VIEIRA, S. Análise de regressão. Uma introdução a econometria. 2.ed. São Paulo, Hucitec, 1983. 379p.
- HUSSAINI, S.H.; GOODMAN, M.M.; TIMOTHY, D.H. Multivariate analysis and the geographical distribution of the word collection of finger millet. *Crop Science*, **17**: 257-63, 1977.
- IEZZONI, A.F. & PRITTS, M.P. Applications of principal component analysis to horticultural research. *HortScience*, **26**: 334-8, 1991.
- JAMES, F.C. & McCULLOCH, C.E. Multivariate analysis in ecology and systematics: Panacea or pandora's box? *Annual Review Ecology systematic*, **21**: 129-66, 1990.
- JEGER, M.J. Multivariate models of the components of partial resistance. *Protection Ecology*, **2**: 265-9, 1980.
- JEGER, M.J.; JONES, D.G.; GRIFFITHS, E. Components of partial resistance of wheat seedlings to *Septoria nodorum*. *Euphytica*, **32**: 575-84, 1983.
- JOHNSON, R.A. & WICHERN, D.W. Applied multivariate statistical analysis. 2.ed. Englewood Cliffs, Prentice Hall, 1988. 607p.
- JOLLIFFE, I.T. Discarding variables in a principal component analysis. I. Artificial data. *Applied Statistics*, **21**: 160-73, 1972.
- JOLLIFFE, I.T. Discarding variables in a principal component analysis. II. Real data. *Applied Statistics*, **22**: 21-31, 1973.

- JONES, I.T. Components of adult plant resistance to powdery mildew (*Erysiphe graminis* f. sp. *avenae*) in oats. *Annals of Applied Biology*, 90: 233-9, 1978.
- KUSHALAPPA, A.C. Uma escala para estimar a intensidade foliar da ferrugem do cafeeiro. *Fitopatologia Brasileira*, 3: 119, 1978. (Resumo)
- KUSHALAPPA, A.C. & ESKEES, A.B. Advances in coffee rust research. *Annual Review of Phytopathology*, 27: 503-31, 1989.
- LACHENBRUCH, P.A. & GOLDSTEIN, M. Discriminant analysis. *Biometrics*, 35: 69-85, 1979
- LALANCETTE, N.; HICKEI, K.D.; COLE JR., H. parasitic fitness and intrastain diversity of benomyl-sensitive and benomyl-resistant subpopulations of *Venturia inaequalis*. *Phytopathology*, 77: 1600-6, 1987.
- LANCE, G.N. & WILLIAMS, W.T. Mixed-data classificatory programs. I. Agglomerative systems. *The Australian Computer Journal*, 15-20, 1967.
- LEBEDA, A. & JENDRULEK, T. Application of cluster analysis for establishment of genetic similarity in gene-for-gene host-parasite relationships. *Journal of Phytopathology*, 119: 131-41, 1987a.
- LEBEDA, A. & JENDRULEK, T. Cluster analysis as a method for evaluation of genetic similarity in specific host-parasite interaction (*Lactuca sativa* - *Bremia lactucae*). *Theoretical and Applied Genetics*, 75: 194-9, 1987b.
- LEE, I.M.; DAVIS, R.E.; CHEN, T.A.; CHIYKOWSKI, L.N.; FLETCHER, J.; HIRUKI, C.; SCHAFF, D.A. A genotype-based system for identification and classification of micoplasmalike organisms (MLO) in the aster yellows MLO strains cluster. *Phytopathology*, 82: 977-86, 1992.
- LEONARD, K.J. & LEATH, S. Genetic diversity in fields populations of *Cochliobolus carbonum* on corn in North Carolina. *Phytopathology*, 80: 1154-9, 1990.
- LI, C.C. Path analysis - a primer. Boxwood Press, Pacific Grove, 1975. 346p.
- LINDE, D.C.; GROTH, J.V.; ROELFS, A.P. Comparison of isoenzyme and virulence diversity patterns in the bean rust fungus *Uromyces appendiculatus*. *Phytopathology*, 80: 141-7, 1990.
- LITTLE, T.M. & HILLS, F.J. *Agricultural experimentation*. New York, John Wiley and Sons. 1978. 350p.
- MADDEN, L.V. Quantification of disease progression. *Protection Ecology*, 2: 159-76, 1980.
- MADDEN, L.V. Statistical analysis and comparison of disease progress curves. In: LEONARD, K.J. & FRY, W.E. (eds). *Plant disease epidemiology*. New York, Macmillan Publishing, 1986. v.1, p. 55-84.
- MADDEN, L.V. & PENNYPACKER, S.P. Principal component analysis of tomato early blight epidemics. *Phytopathologische Zeitschrift*, 95: 364-9, 1979.
- MADDEN, L.V.; KNOKE, J.K.; LOUIE, R. Considerations for the use of multiple comparison procedures in phytopathological investigations. *Phytopathology*, 72: 1015-7, 1982.
- MANLY, B.F.J. *Multivariate statistical methods: A primer*. London, Chapman and Hall, 1986. 159p.
- MARDIA, K.V.; KENT, J.T.; BIBBY, J.M. *Multivariate analysis*. London, Academic Press, 1979. 521p.

- MIRANDA, J.E.C.; CRUZ, C.D.; COSTA, C.P. Predição do comportamento de híbridos de pimentão (*Capsicum annum* L.) pela divergência genética dos progenitores. *Revista Brasileira de Genética*, **11**: 929-37, 1988.
- MORAIS, O.P. & SOARES, P.C. Análise de agrupamento e suas aplicações. Viçosa - MG, UFV, 1988. 30p. (mimeografado)
- MORRISON, D.F. Multivariate statistical methods. 2.ed. New York, McGraw Hill, 1976. 415p.
- MURPHY, J.P.; COX, T.S.; RODGERS, D.M. Cluster analysis of red winter wheat cultivars based upon coefficients of parentage. *Crop Science*, **26**: 672-6, 1986.
- NEI, M. Genetic distances between populations. *The American Naturalist*, **106**: 283-92, 1972.
- NEI, M. Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences of the United States of America*, **70**: 3321-3, 1973.
- NEI, M. & LI, W.H. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceeding of National Academy of Sciences of the United States of America*, **76**: 5269-73, 1979.
- NELSON, R.R. Horizontal resistance in plants: concepts, controversies and applications. In: SEMINAR ON HORIZONTAL RESISTANCE TO THE BLAST DISEASE OF RICE. Cali, 1971. *Proceedings...* Cali, CIAT, 1975. p. 1-20. (Series CE-9)
- NOE, J.P. & BARKER, K.R. Relation of within-field spatial variation of plant-parasite nematode. *Phytopathology*, **75**: 247-52, 1985.
- NORONHA-WAGNER, M. & BETTENCOURT, A.J. Genetics study of the resistance of *Coffea* sp. to leaf rust. I. Identification and behaviour of four factors conditioning disease reaction in *Coffea arabica* to twelve physiological races of *Hemileia vastatrix*. *Canadian Journal of Botany*, **45**: 2021-31, 1967.
- NYGAARD, S.L.; ELLIOTT, C.K.; CANNON, S.J.; MAXWELL, D.P. isozyme variability among isolates of *Phytophthora megasperma*. *Phytopathology*, **79**: 773-80, 1989.
- OHM, H.W. & SHANER, G.E. Three components of slow leaf-rusting at different growth stages in wheat. *Phytopathology*, **66**: 1356-60, 1976.
- OTROSINA, W.S. & COBB JR., F.W. Analysis of allozymes of three distinct variants of *Verticicladiella wageneri* isolated from conifers in western North America. *Phytopathology*, **77**: 1360-3, 1987.
- PALMER, M.A.; STEWART, E.L.; WINGFIELD, M.J. Variation among isolates of *Sphaeropsis sapinea* in the north central United States. *Phytopathology*, **77**: 944-8, 1987.
- PARLEVIET, J.E. Partial resistance of barley to leaf rust, *Puccinia hordei*. I. Effect of cultivar and development stage on latent period. *Euphytica*, **24**: 21-7, 1975.
- PARLEVIET, J.E. Components of resistance that reduce the rate of epidemic development. *Annual Review of Phytopathology*, **17**: 202-22, 1979.
- PARLEVIET, J.E. & VAN OMMEREN, A. Partial resistance of barley to leaf rust, *Puccinia hordei*. II. Relationship between field trials, microplot tests and latent period. *Euphytica*, **24**: 293-303, 1975.
- PARLEVIET, J.E.; LINDHOUT, W.H.; VAN OMMEREN, A.; KUIPER, H.J. Level of partial resistance to leaf rust *Puccinia hordei* in west-european barley and how to Select for it. *Euphytica*, **29**: 1-8, 1980.

- PATAKY, J.K.; HEADRICK, J.M.; SUPARYONO. Classification of sweet corn hybrid reactions to common rust, northern leaf blight, Stewart's wilt, and Goss' wilt and associated yield reductions. *Phytopathology*, 78: 172-8, 1988.
- PEREIRA, A.V. Utilização de análise multivariada na caracterização de **germoplasma de mandioca**. (*Manihot esculenta* Crantz). Piracicaba, ESALQ. 1989. 180p. (Tese D.S.)
- PEREZ-GONZALES, S. Associations among morphological and phenological characteres representing apricot germplasm in central México. *Journal of American Society Horticultural Science*, 117: 486-90, 1992.
- PERSON, C. Gene-for-gene relationships in host-parasite systems. *Canadian Journal of Botany*, 37: 1101-30, 1959.
- PETRUNAK, D.M. & CHRIST, B.J. Isozyme variability in *Alternaria solani* and *A. alternata*. *Phytopathology*, 82: 1343-7, 1992.
- PIMENTEL GOMES, F. A estatística moderna na pesquisa **agropecuária**. Piracicaba, Potafos, 1984. 160p.
- PIMENTEL GOMES, F. Curso de estatística experimental. 12.ed. Piracicaba, Nobel, 1987. 467p.
- POYNTZ, B. & HYDE, P.M. The expression of partial resistance of wheat to *Puccinia recondita*. *Journal of Phytopathology*, 120: 136-42, 1987.
- PRUVOST, O.; HARTUNG, J.S.; CIVEROLO, E.L.; DUBOIS, C.; PERRIER, X. Plasmid DNA fingerprints distinguish pathotypes of *Xanthomonas campestris* pv. *citri*. The causal agent of citrus bacterial canker disease. *Phytopathology*, 82: 485-90, 1992.
- RAO, C.R. Advanced statistical methods in biometric research. New York, John Wiley & Sons, 1952. 389p.
- RAO, C.R. Linear statistical inference and its applications. 2.ed. New York, John Wiley & Sons, 1973. 625p.
- REGAZZI, A.J. Teste para verificar a identidade de modelos de regressão e a igualdade de alguns parâmetros num modelo polinomial ortogonal. *Revista Ceres, Viçosa - MG*, 40: 176-95, 1993.
- REGO, A.M. Variabilidade de *Colletotrichum orbiculare* e avaliação de **resistência de germoplasma de melancia e melão ao patógeno**. Viçosa - MG, UFV, 1993. 79p. (Tese M.S.)
- RIBEIRO, I.J.A.; BERGAMIN FILHO, A.; CARVALHO, P.C.T. Avaliação da resistência horizontal a *Hemileia vastatrix* Berk. & Br. em cultivares de *Coffea arabica* L. em condições naturais de epidemia. *Summa Phytopathologica*, 7: 80-95, 1981.
- RIBOLDI, J. Análise de agrupamento: **cluster analysis**. Piracicaba, ESALQ, 1985. 49p. (mimeografado)
- RIBOLDI, J. Análise de agrupamento: "cluster analysis" e suas **aplicações**. Piracicaba, ESALQ, 1986. 33p. (mimeografado)
- ROBINSON, R.A. Disease resistance terminology. *Review of Applied Mycology*, 48: 593-605, 1969.
- ROBINSON, R.A. Vertical resistance. *Review of Plant Pathology*, 50: 233-9, 1971

- RODRIGUES JR., C.J. Raças fisiológicas de *Hemileia vastatrix*. In: SIMPÓSIO SOBRE FERRUGENS DO CAFEIEIRO. Oeiras, 1983. Anais... Oeiras, Centro de Investigação das Ferrugens do Cafeeiro, 1983. p. 69-79.
- RODRIGUES JR., C.J. Coffee rusts: History, taxonomy, morphology, distribution and host resistance. *Fitopatologia Brasileira*, **15**: 5-9, 1990.
- RODRIGUES JR., C.J.; BETTENCOURT, A.J.; RIJO, L. Races of the pathogen and resistance to coffee rust. *Annual Review of Phytopathology*, **13**: 49-70, 1975.
- ROGERS, D.J. Taximetrics - New name, old concept. *Brittonia*, **15**: 285-90, 1963
- ROHLF, F.J. Stereograms in numerical taxonomy. *Systematic Zoology*, **17**: 246-55, 1968.
- ROMESBURG, H.C. Cluster analysis for researchers. Belmont, Lifetime Learning Publications, 1984. 334p.
- ROOS, I.M.M. & HATTINGH, M.J. Pathogenicity and numerical analysis of phenotypic features of *Pseudomonas syringae* strains isolated from deciduous fruit trees. *Phytopathology*, **77**: 900-8, 1987.
- ROUSE, D.I.; MacKENZIE, D.R.; NELSON, R.R. A relationship between initial inoculum and apparent infection rate in a set of disease progress data for powdery mildew on wheat. *Phytopathologische Zeitschrift*, **100**: 143-9, 1981.
- SAH, D.N. & FEHRMANN, H. Virulence patterns of geographically differing isolates of *Pyrenophora tritici-repentis* and sources of resistance in wheat. *Plant Disease*, **76**: 712-6, 1992.
- SCHILDER, A.M.C. & BERGSTROM, G.C. Variation in virulence within the population of *Pyrenophora tritici-repentis* in New York. *Phytopathology*, **80**: 84-90, 1990.
- SCOTT, A.J. & KNOTT, M. Cluster analysis method for grouping means in the analysis of variance. *Biometrics*, **30**: 507-12, 1974.
- SEAL, H.L. Multivariate statistical analysis for biologists. London, Methuen, 1964. 208p
- SEALY, R.L.; KENERLEY, C.M.; WILLIAMS, E.L. Evaluation of *Amaranthus* accessions for resistance to damping-off by *Pythium myriotylum*. *Plant Disease*, **72**: 985-9, 1988.
- SEDIYAMA, C.S.; SEDIYAMA, T.; SAKIYAMA, N.S. Seleção de caracteres fenotípicos para diferenciação de genótipos de soja, pela análise de componentes principais. *Revista Ceres*, Viçosa-MG, **36**: 330-5, 1989.
- SHANER, G. & FINNEY, R.E. The effect of nitrogen fertilization on the expression of slow-mildewing resistance in knox wheat. *Phytopathology*, **67**: 1051-6, 1977.
- SILVEIRA NETO, S. Análise fenética. In: ALVES, S.B. (coord.) Controle microbiano de insetos. São Paulo, Manole, 1986. p. 384-407.
- SINGH, D. The relative importance of characters affecting genetic divergence. *Indian Journal of Genetics*, **41**: 237-45. 1981.
- SINGH, R.K. & CHAUDHARY, B.D. Biometrical methods in quantitative genetic analysis. New Delhi, Kalyani Publishers, 1979. 304p.
- SMITH, H.; GNANADESIKAN, R.; HUGHES, J.B. Multivariate analyses of variance (MANOVA). *Biometrics*, **18**: 22-41, 1962.

- SMITH, J.S.C. Genetic variability within U.S. hybrid maize: Multivariate analysis of isozyme data. *Crop Science*, 24: 1041-6, 1984.
- SMITH, J.S.C.; GOODMAN, M.M.; STUBER, C.W. Genetic variability within U.S. maize germplasm. II. Widely-used inbred lines 1970 to 1979. *Crop Science*, 25: 681-5, 1985.
- SNEATH, P.H.A. & SOKAL, R.R. Numerical taxonomy. The principles and practice of numerical classification. San Francisco, W.H. Freeman, 1973. 573p.
- SNIJDERS, C.H.A. & PERKOWSKI, J. Effects of head blight caused by *Fusarium culmorum* on toxin content and weight of wheat kernels. *Phytopathology*, 80: 566-70, 1990.
- SOKAL, R.R. Classification: Purposes, principles, progress, prospects. *Science*, 185: 1115-23, 1974.
- SOKAL, R.R. & ROHLF, F.J. The comparison of dendrograms by objective methods. *Taxon*, 11: 33-40, 1962.
- SOUZA, J. Métodos estatísticos nas ciências psicossociais. V.II. Análise em componentes principais e suas aplicações. Brasília. Thesaurus, 1988a, 67p.
- SOUZA, J. Métodos estatísticos nas ciências psicossociais. V.III. Análise fatorial. Brasília, Thesaurus, 1988b, 69p.
- STAUFFER, D.F.; GARTON, E.O.; STEINHORST, R.K. A comparison of principal components from real and random data. *Ecology*, 66: 1693-8, 1985.
- STEEL, R.G.D. & TORRIE, J.H. Principles and procedures of statistics - A biometrical approach. 2.ed. New York, McGraw-Hill, 1980. 634p.
- STRAUSBAUGH, C.A.; SCHROTH, M.N.; WEINHOLD, A.R.; HANCOCK, J.G. Assessment of vegetative compatibility of *Verticillium dahliae* tester strains and isolates from California potatoes. *Phytopathology*, 82: 61-8, 1992.
- TAI, G.C.C. & YOUNG, D.A. Performance and prediction of potato genotypes tested in international trials. *Euphytica*, 42: 275-84, 1989.
- TAMAYO, P.J. Resistência de progênies de *Catimor* a oito raças de *Hemileia vastatrix* Berk. & Br. Viçosa - MG, UFV, 1988. 64p. (Tese M.S.)
- THOMPSON, J.P. & REES, R.G. Pattern analysis in epidemiological evaluation of cultivar resistance. *Phytopathology*, 69: 545-9, 1979.
- TORRES, J.C. & VENTURA, J.A. AVACPD: um programa para calcular a área e volume abaixo da curva de progresso da doença. *Fitopatologia Brasileira*, 16: 52, 1991. (Resumo)
- TUSKAN, G.A.; WALLA, J.A.; LUNDQUIST, J.E. Genetic geographic variation in *Peridermium harknessii* in the north-central United States. *Phytopathology*, 80: 857-61, 1990.
- UNIVERSIDADE FEDERAL DE VIÇOSA. Centro de Processamento de Dados. Manual do sistema de análises estatísticas e genéticas (SAEG). Viçosa - MG, UFV, s.d. 252p.
- VAN BRUGGEN, A.H.C. & ARNESON, P.A. Path coefficient analysis of effects of *Rhizoctonia solani* on growth and development of dry beans. *Phytopathology*, 76: 874-8, 1986.
- VAN DER PLANK, J.E. Plant diseases: epidemics and control. New York, Academic Press, 1963. 349p.
- VAN DER PLANK, J.E. Disease resistance in plants. New York, Academic Press, 1968. 206p

- VAN DER PLANK, J.E. Principles of plant infection. New York, Academic Press, 1975. 216p
- VARMA, N.S. & GULATI, S.C. Genetic divergence in 2-rowed and 6-rowed barley. Indian Journal of Genetics, **42**: 314-8, 1982.
- VIEIRA, S. Introdução a bioestatística. 5.ed. Rio de Janeiro, Campos, 1988. 294p
- WARD JR., J.H. Hierarchical grouping to optimize an objective function. Journal of American Statistical Association, **58**: 236-44, 1963.
- WHISSON, S.C.; MACLEAN, D.J.; MANNERS, J.M.; IRWIN, J.A.G. Genetic relationships among australian and north american isolates of *Phytophthora megasperma* f. sp. *glycinea* assessed by multicopy DNA probes. Phytopathology, **82**: 863-8, 1992.
- ZADOKS, J.C. & SCHEIN, R.D. Epidemiology and plant disease management. New York, Oxford University Press, 1979. 427p.
- ZAMBOLIM, L.; VALE, F.X.R.; CHAVES, G.M. Partial resistance of soybean cultivars to *Phakopsora pachyrhizi*. Fitopatologia Brasileira, **8**: 117-22, 1983.
- ZAMBOLIM, L.; MARTINS, M.C. del P.; CHAVES, G.M. Café. Informe **Agropecuário**, **11** (131): 64-75, 1985a.
- ZAMBOLIM, L.; ACUÑA, R.S.; VALE, F.X.R.; MACABEU, A.J.; CHAVES, G.M. Influência da carga pendente sobre o desenvolvimento da ferrugem do cafeeiro. In: CONGRESSO BRASILEIRO DE PESQUISAS CAFEEIRAS, 12. Caxambu, 1985. Anais... Caxambu, IBC, 1985b. p. 123-4.
- ZHU, H.; HIGGINBOTHAM, K.O.; DANKIK, B.P. Intraspecific genetic variability of isozymes in the ectomycorrhizal fungus *Suillus tomentosus*. Canadian Journal of Botany, **66**: 588-94, 1988.

APÊNDICES

APÊNDICE A

QUADRO 1A - Análise da Variância do Índice de Doença de Cultivares de Pessegueiro

Fonte de Variação	GL	QM	F
Blocos	1	0,0149	
Cultivares	10	0,0579	21,65**
Resíduo (a)	10	0,0027	
Época de Avaliação (EA)	10	0,4974	368,48**
EA x Cultivares	100	0,0048	3,54**
Resíduo (b)	110	0,0013	
Total	241		

CV¹ (a): 15.19 %

CV (b): 10.79%

** Significado pelo teste F (P < 0.01).

¹ Coeficiente de variação.

QUADRO 2A - Análise de Variância de AACPD para os Cultivares de Pessegueiro

Fonte de Variação	GL	SQ	QM	F
Blocos	1	29,958	29,958	
Cultivares	10	1096,283	109,628	21.50**
Resíduo	10	50,993	5,099	
Total	21	1177,234		

** Significado pelo teste F ($P < 0,01$)

QUADRO 3A - Análise de Variância com o Teste de Hipótese H_0 : as equações de todos os Cultivares de Pessegueiro são Idênticas

Fonte de Variação	GL	SQ	QM	F
Parâmetros (β)	(22)	(16,72583)		
Parâmetros (θ)	2	16,28845		
Redução (H_0)	20	0,43738	0,02187	6,73**
Resíduo	99	0,32195	0,00325	
Total	121	17.04778		

** Significado pelo teste F ($P < 0,01$).

QUADRO 4A - Análise de Variância com o Teste de Hipótese H_0 : as Equações dos Cultivares 3, 4, 5, 6, 7 e 8 de Pessegueiro são Idênticas

Fonte de Variação	GL	SQ	QM	F
Parâmetros (β)	(12)	(8,85960)		
Parâmetros (θ)	2	8,80481		
Redução (H_0)	10	0,05479	0,00548	1,62 ^{ns}
Resíduo	54	0,18261	0,00338	
Total	66	9,04221		

^{ns} Não-significativo pelo teste F ($P > 0,05$)

QUADRO 5A - Análise de Variância com o Teste de Hipótese H_0 : as Equações dos Cultivares 9 e 11 de Pessegueiro são Idênticas

Fonte de Variação	GL	SQ	QM	F
Parâmetros (β)	(4)	(1,85539)		
Parâmetros (θ)	2	1,85124		
Redução (H_0)	2	0,00415	0,00208	0,53 ^{ns}
Resíduo	18	0,07079	0,00393	
Total	22	1,92618		

^{ns} Não-significativo pelo teste F ($P > 0,05$).

QUADRO 8A - Análise de Variância com o Teste de Hipótese H_0 : as Equações dos Cultivares 3, 4, 5, 6, 7, 8, 9 e 11 de Pessegueiros são Idênticas

Fonte de Variação	GL	SQ	QM	F
Parâmetros (β)	(16)	(10,71499)		
Parâmetros (θ)	2	(10,55373)		
Redução (H_0)	14	(0,16126)	0,01152	3,27**
Resíduo	72	(0,25340)	0,00352	
Total	88	10,96839		

** Significativo pelo teste F ($P < 0,01$)

QUADRO 9A - Análise de Variância com o Teste de Hipótese H_0 : as Equações dos Cultivares 3, 4, 5, 6, 7, 8 e 10 de Pessegueiros são Idênticas

Fonte de Variação	GL	SQ	QM	F
Parâmetros (β)	(14)	(10,13650)		
Parâmetros (θ)	2	10,06720		
Redução (H_0)	12	0,06930	0,00578	1,88 ^{ns}
Resíduo	63	0,19377	0,00308	
Total	77	10,33027		

^{ns} Não-Significativo pelo teste F ($P > 0,05$).

QUADRO 10A - Análise de Variância com o Teste de Hipótese H_0 : as Equações dos Cultivares 9, 10 e 11 de Pessegueiro são Idênticas

Fonte de Variação	GL	SQ	QM	F
Parâmetros (β)	(6)	(3,13230)		
Parâmetros (θ)	2	3,10919		
Redução (H_0)	4	0,02311	0,00578	1,90 ^{ns}
Resíduo	27	0.08195	0,00304	
Total	33	3,21425		

^{ns} Não-significativo pelo teste F ($P > 0,05$)

QUADRO 11A - Análise de Variância com o Teste de Hipótese H_0 : as Equações dos Cultivares 1, 2, 9 e 11 de Pessegueiro são Idênticas

Fonte de Variação	GL	SQ	QM	F
Parametros (β)	(8)	(6,58933)		
Parâmetros (θ)	2	6,22724		
Redução (H_0)	6	0,36209	0,06035	16,95 ^{**}
Resíduo	36	0.12819	0,00356	
Total	44	6.71752		

^{**} Significativo pelo teste F ($P < 0,01$).

QUADRO 12A - Análise de Variância com o Teste de Hipótese H_0 : as Equações dos Cultivares 1 e 10 de Pessegueiro são Idênticas

Fonte de Variação	GL	SQ	QM	F
Parâmetros (β)	(4)	(3,88824)		
Parâmetros (θ)	2	3,75992		
Redução (H_0)	2	0,12832	0,06416	34,87**
Resíduo	18	0,03319	0,00184	
Total	22	3,92143		

Significativo pelo teste F ($P < 0,01$)

QUADRO 13A - Análise de Variância com o Teste de Hipótese H_0 : as Equações dos Cultivares 5, 6, 7, 8, 9 e 10 de Pessegueiro são idênticas

Fonte de Variação	GL	SQ	QM	F
Parâmetros (β)	(12)	(7,66267)		
Parâmetros (θ)	2	7,62173		
Redução (H_0)	10	0,04094	0,00409	1,38 ^{ns}
Resíduo	54	0,16022	0,00297	
Total	66	7,82289		

^{ns} Não-significativo pelo teste F ($P > 0,05$).

QUADRO 14A - Análise de Variância com o Teste de Hipótese H_0 : as Equações dos Cultivares 2, 3, e 4 de Pessegueiro são Idênticas

Fonte de Variação	GL	SQ	QM	F
Parâmetros (β)	(6)	(5,59645)		
Parâmetros (θ)	2	5,58276		
Redução (H_0)	4	0,01369	0,00342	0,82 ^{ns}
Resíduo	27	0.11239	0,00416	
Total	33	5.70884		

^{ns} Não-significativo pelo teste F ($P > 0.05$)

QUADRO 15A - Análise de Variância com o Teste de Hipótese H_0 : as Equações dos Cultivares 2, 3, 4, 5, 6.7, 8,9 e 10 de Pessegueiro são Idênticas

Fonte de Variação	GL	SQ	QM	F
Parâmetros (β)	(18)	(13,25912)		
Parâmetros (H)	2	13,07827		
Redução (H_0)	16	0,18085	0,01130	3,35 ^{**}
Resíduo	81	0,27261	0,00337	
Total	99	13,53173		

^{**} Significativo pelo teste F ($P < 0,01$)

QUADRO 16A - Análise de Variância com o Teste de Hipótese H_0 : as Equações dos Cultivares 5, 6, 7, 8, 9, 10 e 11 de Pessegueiro são Idênticas

Fonte de Variação	GL	SQ	QM	F
Parâmetros (β)	(14)	(8,51805)		
Parâmetros (H)	2	8,43809		
Redução (H_0)	12	0,07996	0,00666	2,24*
Resíduo	63	0,18753	0,00298	
Total	77	8,70558		

* Significativo pelo teste F ($P < 0,05$).

QUADRO 17A - Análise de Variância com o Teste de Hipótese H_0 : as Equações dos Cultivares 1, 2, 3 e 4, de Pessegueiro são Idênticas

Fonte de Variação	GL	SQ	QM	F
Parâmetros (β)	(8)	(8,20778)		
Parâmetros (H)	2	8,14619		
Redução (H_0)	6	0,06159	0,01027	2,75*
Resíduo	36	0,13442	0,00337	
Total	44	8,34220		

* Significativo pelo teste F ($P < 0,05$).

QUADRO 18A - Análise de Variância com o Teste de Hipótese H_0 : as Equações dos Cultivares 2, 3, 4, e 11 de Pessegueiro são Idênticas

Fonte de Variação	GL	SQ	QM	F
Parâmetros (β)	(8)	(6,45183)		
Parâmetros (H)	2	6,27649		
Redução (H_0)	6	0,17533	0,02922	7,53**
Resíduo	36	0,13971	0,00388	
Total	44	6,59153		

** Significativo pelo teste F ($P < 0,01$).

QUADRO 19A - Análise de Variância com o Teste de Hipótese H_0 : as Equações dos Cultivares 1 e 11 de Pessegueiro são Idênticas

Fonte de Variação	GL	SQ	QM	F
Parâmetros (β)	(4)	(3,46672)		
Parâmetros (R)	2	3,21730		
Redução (H_0)	2	0,24941	0,12471	45,52**
Resíduo	18	0,04934	0,00274	
Total	22	3,51606		

** Significativo pelo teste F ($P < 0,01$).

QUADRO 20A - Análise de Variância com o Teste de Hipótese H_0 : as Equações dos Cultivares 1, 9 e 11 de Pessegueiro são Idênticas

Fonte de Variação	GL	SQ	QM	F
Parâmetros (β)	(6)	(4,46673)		
Parâmetros (R)	2	4,16891		
Redução (H_0)	4	0,29782	0,07446	21,65**
Resíduo	21	0,09282	0,00344	
Total	33	4,55955		

** Significativo pelo teste F ($P < 0,01$).

QUADRO 21A - Análise de Variância com o Teste de Hipótese H_0 : as Equações dos Cultivares 2 e 10 de Pessegueiro são Idênticas

Fonte de Variação	GL	SQ	QM	F
Parâmetros (β)	(4)	(3,39950)		
Parâmetros (R)	2	3,33300		
Redução (H_0)	2	0,06651	0,03325	12,84**
Resíduo	18	0,04653	0,00259	
Total	22	3,44603		

** Significativo pelo teste F ($P < 0,01$).

QUADRO 22A -Análise de Variância com o Teste de Hipótese H_0 : as Equações dos Cultivares 2, 9 e 11 de Pessegueiro são Idênticas

Fonte de Variação	GL	SQ	QM	F
Parâmetros (β)	(6)	(3.97800)		
Parâmetros (θ)	2	3,79900		
Redução (H_0)	4	0,17899	0,04475	11,39**
Resíduo	27	0,10616	0,00393	
Total	33	4,08415		

** Significativo pelo teste F ($P < 0,01$).

QUADRO 23A - Análise de Variância com o Teste de Hipótese H_0 : as Equações dos Cultivares 1, 3,4,5,6,7 e 8 de Pessegueiro são Idênticas

Fonte de Variação	GL	SQ	QM	F
Parâmetros (β)	(14)	(11,47094)		
Parâmetros (H)	2	11,27524		
Redução (H_0)	12	0,19570	0,01631	5,02**
Resíduo	63	0,20464	0,00325	
Total	77	11.67557		

** Significativo pelo teste F ($P < 0.01$)

QUADRO 24A - Análise de Variância com o Teste de Hipótese H_0 : as Equações dos Cultivares 2, 3, 4, 5, 6, 7, e 8 de Pessegueiro são Idênticas

Fonte de Variação	GL	SQ	QM	F
Parâmetros (β)	(14)	(10,98220)		
Parâmetros (θ)	2	10,87340		
Redução (H_0)	12	0,10879	0,00907	2,62**
Resíduo	63	0,21797	0,00346	
Total	77	11,20017		

** Significativo pelo teste F ($P < 0,01$).

QUADRO 25A - Análise de Variância com o Teste de Hipótese H_0 : as Equações dos Cultivares 5, 6 e 10 de Pessegueiro são Idênticas

Fonte de Variação	GL	SQ	QM	F
Parâmetros (β)	(6)	(4,18960)		
Parâmetros (θ)	2	4,17391		
Redução (H_0)	4	0,01569	0,00392	1,58 ^{ns}
Resíduo	27	0,06683	0,00248	
Total	33	4,25643		

^{ns} Não-significativo pelo teste F ($P > 0,05$).

APÊNDICE B

QUADRO 1B - Média¹ e Variância (σ^2) de Severidade e Esporulação de Genótipos de *Coffea arabica* em relação a Raça II de *H. vastatrix*

Genótipos	Severidade		Esporulação (esporos/cm ²)	
	Média	a ²	Média	a ²
H 415	12,36	30.17	10270.9	4.35 x 10 ⁷
H 442	15,76	12.33	4770,8	1,48 x 10 ⁷
H 471	17.22	19,96	15698.9	5,08 x 10 ⁷
H 472	10,26	73.61	6362,8	5,57 x 10 ⁷
H 481	16,54	32,66	6784,6	2,87 x 10 ⁷
UFV 2143	11,28	42,51	10442,9	6,09 x 10 ⁷
UFV 2150	16,18	3,73	11993,7	0.93 x 10 ⁷

¹ Médias de cinco repetições, parcela com dez plantas.

QUADRO 2B - Valores do Teste de Bartlett Calculados para as Variáveis Período de Incubação (PI), Período latente (PL), Severidade (SEV) e Esporulação (ESP)

Variáveis	Valor Calculado
PI	11,01 ^{ns}
PL	15,76*
PL ^{1/2}	14,04*
(PL + 0,5) ^{1/2}	14,06*
SEV	8,07 ^{ns}
ESP	4,72^{ns}

* Significativo (P < 0,05).

^{ns} Não-significativo (P > 0,05).

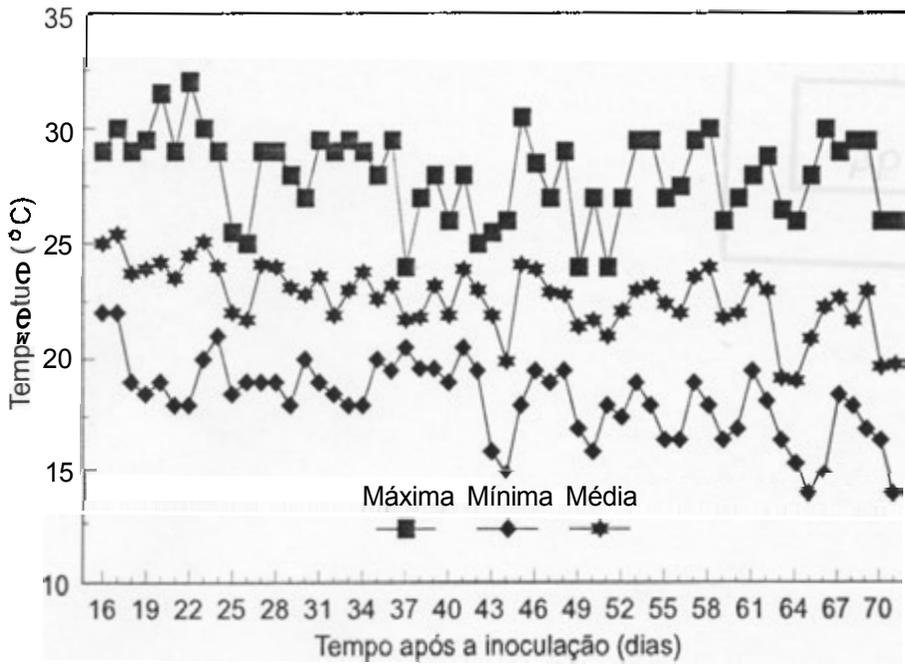


FIGURA 1B - Valores de Temperatura Média, Máxima e Mínima, Observados durante a Avaliação da Resistência Horizontal de Genótipos de *Coffea arabica* a *Hemileia vastatrix*, sob condições de Casa de Vegetação, no Período de Abril e Maio de 1992, em Viçosa-MG.

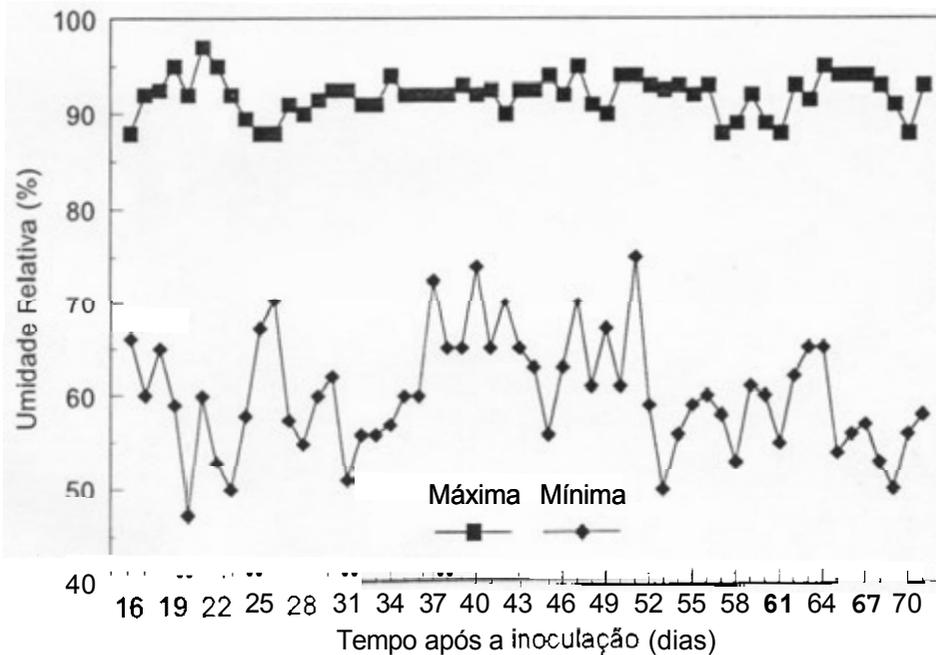


FIGURA 2B - Valores de Umidade Relativa Máxima e Mínima, Observados durante a Avaliação da Resistência Horizontal de Genótipos de *Coffea arabica* a *Hemileia vastatrix*, sob condições de Casa de Vegetação, no Período de Abril e Maio de 1992, em Viçosa-MG.