

ANDERSON RODRIGO DA SILVA

**MÉTODOS DE AGRUPAMENTO: AVALIAÇÃO E APLICAÇÃO AO
ESTUDO DE DIVERGÊNCIA GENÉTICA EM ACESSOS DE ALHO**

**Dissertação apresentada à
Universidade Federal de Viçosa,
como parte das exigências do
Programa de Pós-Graduação em
Estatística Aplicada e Biometria,
para obtenção do título de *Magister
Scientiae*.**

**VIÇOSA
MINAS GERAIS - BRASIL
2012**

**Ficha catalográfica preparada pela Seção de Catalogação e
Classificação da Biblioteca Central da UFV**

T

S586m
2012

Silva, Anderson Rodrigo da, 1987-

Métodos de agrupamento : avaliação e aplicação ao estudo de divergência genética em acessos de alho / Anderson Rodrigo da Silva. – Viçosa, MG, 2012.
xiv, 67f. : il. ; 29cm.

Inclui apêndice.

Orientador: Paulo Roberto Cecon.

Dissertação (mestrado) - Universidade Federal de Viçosa.

Referências bibliográficas: f. 56-61.

1. *Allium sativum*. 2. Análise multivariada. 3. Análise discriminatória. 4. Análise por agrupamento. I. Universidade Federal de Viçosa. II. Título.

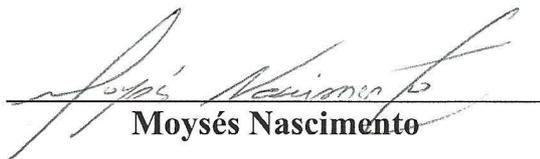
CDD 22. ed. 519.535

ANDERSON RODRIGO DA SILVA

**MÉTODOS DE AGRUPAMENTO: AVALIAÇÃO E APLICAÇÃO AO
ESTUDO DE DIVERGÊNCIA GENÉTICA EM ACESSOS DE ALHO**

Dissertação apresentada à
Universidade Federal de Viçosa,
como parte das exigências do
Programa de Pós-Graduação em
Estatística Aplicada e Biometria,
para obtenção do título de *Magister
Scientiae*.

APROVADA: 13 de fevereiro de 2012.


Moysés Nascimento


Sebastião Martins Filho


Adésio Ferreira


Paulo Roberto Cecon
(Orientador)

AGRADECIMENTOS

À Deus, Mestre dos mestres, Benfeitor do meu espírito.

À Universidade Federal de Viçosa, por me acolher como estudante.

À CAPES e Programa REUNI pela concessão da bolsa de estudo.

À FAPEMIG pelo financiamento do projeto de pesquisa.

À minha companheira de vida, Bárbara, pela compreensão, paciência e amor dedicado. Essa conquista também é sua.

À minha família, pela educação que me deram, incentivo aos meus estudos e por crer em mim, sempre.

Não se limita as seguintes poucas palavras a minha incólume gratidão ao meu orientador e amigo Paulo Roberto Cecon, por sua atenção, por seus conselhos e pelos seus valiosos ensinamentos. Agradeço também aos meus coorientadores, Antônio Policarpo Souza Carneiro e Fabyano Fonseca e Silva, pelos conhecimentos que me passaram durante o mestrado e pelas sugestões neste trabalho.

Aos professores Mário Puiatti e Fernando Luiz Finger, do Departamento de Fitotecnia da UFV, pela instalação do experimento e concessão dos equipamentos utilizados na obtenção dos dados.

À todo o corpo docente do Departamento de Estatística da UFV.

Ao meu amigo Adriano e à Renata pela grande ajuda durante a coleta de dados do experimento.

À todos os colegas do mestrado que contribuíram direta ou indiretamente para a realização desta conquista, em especial aos colegas de turma: Alyne, Danielle, Evandro, Layanne, Mariane, Rafael e Valéria.

*“Quanto mais nos elevamos, menores parecemos
aos olhos daqueles que não sabem voar.”*

F. W. Nietzsche

SUMÁRIO

LISTA DE FIGURAS.....	vii
LISTA DE QUADROS	ix
LISTA DE TABELAS.....	x
LISTA DE ABREVIATURAS E SIGLAS	xii
RESUMO.....	xiii
ABSTRACT.....	xiv
1. INTRODUÇÃO.....	1
2. REVISÃO BIBLIOGRÁFICA	4
2.1. A Cultura do Alho	4
2.2. Divergência Genética	5
2.3. Medidas de Dissimilaridade	6
2.4. Análise de Agrupamento	7
2.4.1. Métodos de Otimização.....	8
2.4.2. Métodos Hierárquicos	8
3. MATERIAL E MÉTODOS.....	11
3.1. Descrição do Experimento	11
3.2. Caracteres Avaliados.....	12
3.3. Análises Estatísticas	13
3.3.1. Análise de Variância e Estimação de Parâmetros Genéticos.....	13
3.3.2. Dissimilaridade Entre Acessos.....	15
3.3.3. Métodos de Agrupamento	20
3.3.3.1. Método de Tocher	20
3.3.3.2. Método de Tocher Modificado	22

3.3.3.3. Método UPGMA.....	24
3.3.3.4. Método da Variância Mínima de Ward	26
3.3.4. Determinação do Número de Grupos	31
3.3.5. Coeficiente de Correlação Cofenética.....	33
3.3.6. Análise Discriminante.....	34
3.3.7. Considerações Metodológicas Finais.....	36
4. RESULTADOS E DISCUSSÃO.....	37
4.1. Análise de Variância	37
4.2. Dissimilaridade Entre Acessos e Importância de Caracteres.....	38
4.3. Agrupamento com o Método de Tocher	39
4.4. Agrupamento com o Método de Tocher Modificado.....	41
4.5. Agrupamento com o Método UPGMA	44
4.6. Agrupamento com o Algoritmo de Ward.....	48
4.7. Resumo da Comparação dos Métodos	53
5. CONCLUSÕES	55
6. REFERÊNCIAS BIBLIOGRÁFICAS	56
APÊNDICE.....	62
A - Análise Discriminante.....	63
B - Script das Análises no <i>Software</i> R	65

LISTA DE FIGURAS

- Figura 1.** Dendrograma para o exemplo hipotético com o método UPGMA..... 26
- Figura 2.** Dendrograma para o exemplo hipotético com o algoritmo de Ward 30
- Figura 3.** Dendrograma obtido com o método UPGMA com base na distância generalizada de Mahalanobis (D^2) para agrupamento de 89 acessos de alho pertencentes ao BGH/UFV. UFV, Viçosa (MG), 2012 45
- Figura 4.** Correlação cofenética para as matrizes de dissimilaridade obtidas com o método UPGMA, baseado na distância generalizada de Mahalanobis (D^2); ** significativo pelo teste de Mantel a 1% de probabilidade, baseado em 1000 reamostragens. UFV, Viçosa (MG), 2012..... 47
- Figura 5.** Dendrograma obtido com o algoritmo de Ward com base na distância generalizada de Mahalanobis (D^2) para agrupamento de 89 acessos de alho pertencentes ao BGH/UFV. UFV, Viçosa (MG), 2012 49
- Figura 6.** Correlação cofenética para as matrizes de dissimilaridade obtidas com o algoritmo de Ward, baseado na distância generalizada de Mahalanobis (D^2); ** significativo pelo teste de Mantel a 1% de probabilidade, baseado em 1000 reamostragens. UFV, Viçosa (MG), 2012..... 51

Figura 7A. Escores das duas primeiras funções discriminantes (LD1 e LD2) estimadas de acordo com o agrupamento de 89 acessos de alho obtido com cada método. UFV, Viçosa (MG), 2012..... 64

LISTA DE QUADROS

- Quadro 1.** Relação dos 89 acessos de alho registrados no Banco de Germoplasmas de Hortaliças da Universidade Federal de Viçosa (BGH/UFV)..... 12
- Quadro 2.** Esquema da análise de variância e esperanças de quadrados médios de um modelo em blocos casualizados, com efeitos fixos de acessos 13

LISTA DE TABELAS

Tabela 1.	Exemplo hipotético para obtenção da matriz de distâncias de Mahalanobis: Médias dos acessos para as variáveis Y_1 , Y_2 e Y_3	16
Tabela 2.	Exemplo hipotético para obtenção da matriz de distâncias de Mahalanobis: Análise de variância	16
Tabela 3.	Exemplo hipotético para o método de Tocher: Resultado do agrupamento	22
Tabela 4.	Exemplo hipotético para o método UPGMA: Resultado.....	26
Tabela 5.	Exemplo hipotético para o algoritmo de Ward: Resultado	30
Tabela 6.	Exemplo hipotético para determinação do ponto de corte no dendrograma pelo método de Mojena (1977): Resultado.....	32
Tabela 7.	Resumo da análise de variância e estimativas dos parâmetros genéticos para caracteres produtivos de alho. UFV, Viçosa (MG), 2012.....	38
Tabela 8.	Contribuição relativa de cada caractere para a divergência genética entre 89 acessos de alho do BGH/UFV com base na estatística S_j (SINGH, 1981). UFV, Viçosa (MG), 2012	39
Tabela 9.	Formação dos grupos de 89 acessos de alho do BGH/UFV segundo o método de otimização de Tocher (original) com base na distância de Mahalanobis. UFV, Viçosa (MG), 2012.....	40
Tabela 10.	Número de acessos (N) e médias das variáveis para os grupos obtidos com o método de Tocher. UFV, Viçosa (MG), 2012.....	40

Tabela 11.	Porcentagem de classificação correta e incorreta de cada grupo segundo o agrupamento de 89 acessos de alho pelo método de Tocher (original). UFV, Viçosa (MG), 2012	41
Tabela 12.	Formação dos grupos de 89 acessos de alho do BGH/UFV segundo o método de otimização de Tocher modificado com base na distância de Mahalanobis. UFV, Viçosa (MG), 2012	42
Tabela 13.	Número de acessos (N) e médias das variáveis para os grupos obtidos com o método de Tocher modificado. UFV, Viçosa (MG), 2012.....	43
Tabela 14.	Porcentagem de classificação correta e incorreta de cada grupo segundo o agrupamento de 89 acessos de alho pelo método de Tocher modificado. UFV, Viçosa (MG), 2012	44
Tabela 15.	Formação dos grupos de 89 acessos de alho do BGH/UFV segundo o método UPGMA com base na distância de Mahalanobis. UFV, Viçosa (MG), 2012.....	46
Tabela 16.	Número de acessos (N) e médias das variáveis para os grupos obtidos com o método UPGMA. UFV, Viçosa (MG), 2012	47
Tabela 17.	Porcentagem de classificação correta e incorreta de cada grupo segundo o agrupamento de 89 acessos de alho pelo método UPGMA. UFV, Viçosa (MG), 2012	48
Tabela 18.	Formação dos grupos de 89 acessos de alho do BGH/UFV segundo o algoritmo de Ward com base na distância de Mahalanobis. UFV, Viçosa (MG), 2012.....	50
Tabela 19.	Número de acessos (N) e médias das variáveis para os grupos obtidos com o algoritmo de Ward. UFV, Viçosa (MG), 2012.....	50
Tabela 20.	Porcentagem de classificação correta e incorreta de cada grupo segundo o agrupamento de 89 acessos de alho pelo algoritmo de Ward. UFV, Viçosa (MG), 2012.....	52
Tabela 21.	Resumo dos resultados obtidos pelos métodos de agrupamento aplicados a dados de 89 acessos de alho. UFV, Viçosa (MG), 2012	53
Tabela 22A.	Coeficientes das funções discriminantes lineares estimados de acordo com o agrupamento de 89 acessos de alho obtido com cada método. UFV, Viçosa (MG), 2012	63

LISTA DE ABREVIATURAS E SIGLAS

BGH	Banco de Germoplasmas de Hortaliças
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
COMP	Comprimento do bulbo
CVe	Coefficiente de Variação Ambiental
CVg	Coefficiente de Variação Genético
DET	Departamento de Estatística
DIAM	Diâmetro do Bulbo
FAPEMIG	Fundação de Amparo à Pesquisa do Estado de Minas Gerais
LD1	Linear Discriminant function 1
NbB	Número de bulbilhos por Bulbo
PMB	Peso Médio do Bulbo
PMR	Produto Médio do Resíduo
PPESTBIO	Programa de Pós-graduação em Estatística Aplicada e Biometria
PT	Produtividade Total
QMB	Quadrado Médio de Blocos
QMG	Quadrado Médio de Genótipos (Acessos)
QMR	Quadrado Médio do Resíduo
REUNI	Programa de Reestruturação e Expansão das Universidades Federais
SQD	Soma de Quadrados de Desvios
TEA	Taxa de Erro Aparente
UFV	Universidade Federal de Viçosa
UPGMA	Unweighted Pair-Group Method using Arithmetic Averages

RESUMO

SILVA, Anderson Rodrigo da, M.Sc., Universidade Federal de Viçosa, fevereiro de 2012. **Métodos de agrupamento: avaliação e aplicação ao estudo de divergência genética em acessos de alho.** Orientador: Paulo Roberto Cecon. Coorientadores: Antônio Policarpo Souza Carneiro e Fabyano Fonseca e Silva.

Este estudo teve por objetivo avaliar, quanto a consistência do agrupamento, os métodos de agrupamentos hierárquicos UPGMA e Ward e os de otimização de Tocher e Tocher modificado, pela aplicação da análise discriminante de Fisher aos grupos obtidos com cada método, em estudo da divergência genética entre acessos de alho, identificando também os acessos mais dissimilares. Os agrupamentos foram realizados com base na distância generalizada de Mahalanobis, que também permitiu quantificar a importância relativa dos caracteres. Os acessos que apresentaram maior dissimilaridade foram os acessos 13 (BGH 4505) e 61 (BGH 5958), principalmente em relação ao peso médio do bulbo e produtividade. Os métodos de Tocher modificado, UPGMA e algoritmo de Ward apresentaram resultados concordantes entre si quanto a formação dos grupos. No entanto, pela análise discriminante de Fisher aplicada aos grupos dos métodos hierárquicos (UPGMA e Ward) observou-se as menores taxas de erro aparente, sendo, portanto, os métodos mais consistentes para o estudo da diversidade genética de acessos de alho.

ABSTRACT

SILVA, Anderson Rodrigo da, M.Sc., Universidade Federal de Viçosa, February, 2012. **Clustering methods: evaluation and application for study of genetic divergence in garlic accessions.** Adviser: Paulo Roberto Cecon. Co-advisers: Antonio Policarpo Souza Carneiro and Fabyano Fonseca e Silva.

This study aimed to assess, as the consistency of the grouping, hierarchical clustering methods UPGMA and Ward and optimization Tocher and modified Tocher by application of Fisher discriminant analysis in groups obtained with each method in the study of genetic divergence among garlic accessions, also identifying the most dissimilar access. The groupings were based on the Mahalanobis distance, which also allowed to quantify the relative importance of characters. The accessions with the highest dissimilarity accesses were 13 (BGH 4505) and 61 (BGH 5958), especially in relation to the average weight of the bulb and productivity. Modified Tocher methods, UPGMA and Ward algorithm presented results agree with each other and form groups. However, the Fisher discriminant analysis applied to groups of hierarchical methods (UPGMA and Ward) showed the lowest apparent error, therefore, more consistent methods for studying the genetic diversity of garlic accessions.

1. INTRODUÇÃO

O alho (*Allium sativum* L.) é cultivado há séculos em todo o mundo, por suas propriedades medicinais e notável valor comercial (SOUZA *et al.*, 2007). Além da importância econômica, é uma cultura de importância no sistema de agricultura familiar, por empregar mão-de-obra desde o plantio até a colheita (TRANI *et al.*, 2005). Em relação a diversidade genética da cultura, existem perspectivas para obtenção de cultivares que possam aumentar a produtividade e permitir a identificação de genótipos resistentes às principais pragas e doenças, pela exploração e manejo dos estoques genéticos. No entanto, para a cultura do alho, nota-se que há carência de referenciais teóricos e práticos acerca das metodologias adequadas para o estudo da variabilidade genética presente nos bancos de germoplasmas.

A definição de divergência genética foi postulada como “qualquer medida quantitativa ou diferença genética, estando ao nível de sequência ou nível de frequência alélica, que é calculada entre indivíduos, populações ou espécies” (MOHAMMADI e PRASANNA, 2003).

A importância dos estudos sobre a divergência genética para o melhoramento reside no fato de que cruzamentos envolvendo genitores geneticamente diferentes são os mais convenientes para produzir alto efeito heterótico e, também, maior variabilidade genética em gerações segregantes (RAO *et al.*, 1981). Neste sentido, alguns autores (PATERNIANI e

LONNQUIST, 1963; MOLL *et al.*, 1965) relatam que deve existir um grau ótimo de divergência para a expressão máxima da heterose. Para identificação desse grau ótimo, faz-se imprescindível conhecer o máximo possível o grau de semelhança entre os indivíduos do estoque genético por meio de técnicas biométricas apropriadas.

Estabelecidos os bancos de germoplasmas, a aplicação de técnicas biométricas para o estudo da divergência dos acessos tem permitido um melhor planejamento de cruzamentos a serem avaliados no programa de melhoramento. Além disso, essas análises têm possibilitado a eliminação de duplicatas e a análise de pureza dos diferentes acessos (CRUZ *et al.*, 2011).

Muitas metodologias estão disponíveis para quantificar e avaliar o grau de divergência genética a partir de informações fenotípicas e genotípicas. No entanto, Cruz *et al.* (2011) salientam que a abrangência dos estudos, de informações, de métodos e de material biológico tem levado a certa dificuldade em escolher e aplicar corretamente as metodologias disponíveis e interpretar, convenientemente, o significado dos resultados das análises biométricas.

Em muitas situações, o pesquisador está interessado em avaliar padrões de agrupamento, formular e testar hipóteses sobre a similaridade ou diversidade obtida. Nesse sentido, a literatura apresenta várias técnicas de agrupamento, que se distinguem pelo tipo de resultado a ser fornecido e pelos diferentes modos de definir a proximidade entre indivíduos e grupos deles. Nas análises de agrupamento o objetivo principal é reunir os indivíduos em um determinado número de grupos, de modo que exista homogeneidade dentro de cada grupo e heterogeneidade entre eles (JOHNSON e WICHERN, 1992; CRUZ e REGAZZI, 1994).

Dos métodos de agrupamento, destacam-se os hierárquicos e os de otimização, pelo fato de serem bastante utilizados em estudos de divergência genética. Nos métodos de otimização, os grupos são formados pela adequação de algum critério de agrupamento. Nos métodos hierárquicos, os indivíduos são agrupados por um processo que se repete em vários níveis até que seja estabelecido o dendrograma.

Diante do exposto, este estudo teve por objetivo avaliar, quanto a consistência do agrupamento, métodos de otimização e hierárquicos pela aplicação da análise discriminante de Fisher, em estudo da divergência genética entre acessos de alho, identificando também os acessos mais dissimilares.

2. REVISÃO BIBLIOGRÁFICA

2.1. A Cultura do Alho

De acordo com o IBGE (2011), na safra de 2010 o Brasil produziu 104.586 toneladas em 10.542 ha, sendo Goiás o maior produtor (39.252 t) com área plantada de 2.671 ha. O rendimento médio nacional foi estimado em 9.921 kg.ha⁻¹.

A cultura é exigente em baixa temperatura e fotoperíodo longo para que ocorra a bulbificação. Em geral, temperaturas médias de 12,8 a 23,9 °C favorecem o desenvolvimento normal do alho. Porém, existem diferenças entre os cultivares quanto à resposta à temperatura e ao fotoperíodo com consequente variação de resposta destas às regiões de cultivo e épocas de plantio (MUELLER *et al.*, 1990). Os cultivares de alho são classificados em precoces, médios e tardios, quanto à exigência em baixa temperatura e fotoperíodo longo para bulbificar, sendo que essa exigência aumenta dos precoces para os tardios (PUIATTI e FERREIRA, 2005). Cultivares de ciclo médio são adaptados ao cultivo na região sudeste e centro-oeste do Brasil, sem necessidade de “vernalização do alho-planta”. Esse é um ponto positivo em relação ao cultivo de alhos nobres em razão do menor custo de produção (TRANI *et al.*, 1997).

Em condições ambientais adequadas quanto ao clima, solo, irrigação, sanidade e tratamentos culturais, cada cultivar possui uma potencialidade genética,

expressada através da fisiologia da planta. O rendimento dos cultivares de alho está diretamente relacionado com a área foliar desenvolvida para cada cultivar, a duração da área foliar, bem como o padrão de alocação de matéria seca de diferentes órgãos da planta (STAHLSCHMIDT *et al.*, 1993).

A Universidade Federal de Viçosa é mantenedora de um Banco de Germoplasma de Hortaliças (BGH/UFV), no qual estão registrados 89 acessos de alho adaptados às condições climáticas da região sudeste. Dentre os acessos dessa coleção, encontram-se acessos considerados comuns e acessos que poderiam ser considerados “semi-nobres”. Esses acessos considerados “semi-nobres” tem bulbilhos de coloração arroxeadas, em número menor que 20 por bulbo, ausência de bulbilhos palito, raro superbrotamento e intermediária exigência em dias longos e em frio (SOUZA *et al.*, 2007). A produtividade desses alhos varia em torno de 10,86 t/ha (SILVA *et al.*, 2000), dependendo do alho-planta.

2.2. Divergência Genética

A divergência ou diversidade genética entre e dentro de populações encontradas em suas condições naturais, em bancos de germoplasma ou desenvolvidas nos programas de melhoramento genético pode ser predita pelas diferenças entre os valores fenotípicos mensurados em suas unidades (indivíduos, famílias, etc.) (DINIZ FILHO, 2000 *apud* CRUZ *et al.*, 2011).

A redução significativa no número de espécies, principalmente vegetais, torna necessário o armazenamento de sementes em coleções de bancos de germoplasma, daí sua importância. Uma vez estabelecidos os bancos, as análises de diversidade de diferentes acessos têm permitido estudos precisos da biodiversidade, de forma a orientar a utilização de diferentes genitores e, assim, um melhor planejamento de cruzamentos a serem avaliados no programa de melhoramento (CRUZ *et al.*, 2011).

Há duas maneiras básicas de se inferir sobre a diversidade genética, sendo a primeira de natureza quantitativa e a outra de natureza preditiva (CRUZ

e CARNEIRO, 2003). Os métodos preditivos de diversidade genética têm sido bastante utilizados, sobretudo pelo fato de que, ao se basearem em diferenças morfológicas, fisiológicas e moleculares dos genótipos, dispensam a obtenção das combinações híbridas entre eles, o que é vantajoso, especialmente quando o número de genitores cuja diversidade se deseja conhecer é elevado (CARVALHO *et al.*, 2003). Por esses métodos as informações múltiplas de cada cultivar são expressas em medidas de dissimilaridade, que representam a diversidade existente no conjunto de acessos estudados.

As medidas de dissimilaridade para caracteres quantitativos mais utilizadas nos estudos genéticos são: a distância euclidiana, o quadrado da distância euclidiana, a distância euclidiana média e a distância generalizada de Mahalanobis (1936). Esta última apresenta a vantagem de levar em consideração a magnitude das variâncias e covariâncias residuais que existem entre os caracteres mensurados, possíveis de serem estimadas a partir de informações de genótipos avaliados em delineamentos experimentais. Outros tipos de medidas de dissimilaridade, inclusive para caracteres qualitativos binários e multicategóricos, podem ser encontrados na literatura (CRUZ *et al.*, 2011; KHATTREE e NAIK, 2000; BUSSAB *et al.*, 1990).

Estudos sobre diversidade genética tem sido utilizados em várias culturas de grande importância econômica, como café (FONSECA *et al.*, 2006), feijão (RODRIGUES *et al.*, 2002), milho (MIRANDA *et al.*, 2003), guaraná (NASCIMENTO FILHO *et al.*, 2001), mandioca (VIEIRA *et al.*, 2008), entre outras.

2.3. Medidas de Dissimilaridade

Um critério que quantifique a distância entre dois objetos é entendido como medida de dissimilaridade e, logo, quanto menores os valores destas medidas, mais similares são os objetos comparados. Um grande número de distâncias tem sido proposto e usado em análise de agrupamento e, segundo Mingoti (2005), cada uma delas produz um determinado tipo de agrupamento.

As medidas de dissimilaridade são de grande importância em estudos de diversidade genética em que se procura identificar genitores à serem utilizados em programas de hibridação (CRUZ *et al.*, 2011). Os autores salientam também que as distâncias mais utilizadas em estudos genéticos são: a distância euclidiana, o quadrado da distância euclidiana, a distância euclidiana média, a distância ponderada e a distância generalizada de Mahalanobis (1936).

Segundo Manly (2008), na prática é usualmente desejável que todas as variáveis quantitativas tenham aproximadamente a mesma influência no cálculo da distância, o que pode ser obtido por um escalonamento preliminar dividindo cada valor da variável pelo seu desvio padrão.

2.4. Análise de Agrupamento

A técnica multivariada de análise de agrupamento é uma maneira de se obter grupos homogêneos, por um esquema que possibilite reunir os indivíduos em um determinado número de grupos, de modo que exista grande homogeneidade dentro de cada grupo e heterogeneidade entre eles (JOHNSON e WICHERN, 1992; CRUZ e REGAZZI, 1994). Manly (2008) comenta que a idéia de uma análise de agrupamento é utilizar os valores das variáveis para planejar um esquema para agrupar os objetos em classes de modo que objetos similares estejam na mesma classe, sendo que o número de classes não é usualmente conhecido. O autor salienta que a análise de agrupamento pode também ser utilizada para redução de dados.

Muitos algoritmos estão disponíveis na literatura para análise de agrupamento, os quais se distinguem pelo tipo de resultado e pelas diferentes formas de definir a proximidade entre as entidades, no entanto de acordo com Bussab *et al.* (1990), a escolha de um particular algoritmo de agrupamento exige o conhecimento de suas propriedades, aliado aos objetivos da pesquisa.

Cruz *et al.* (2011) salientam que, dos métodos de agrupamento, os mais utilizados são os de otimização e os hierárquicos.

2.4.1. Métodos de Otimização

Nos métodos de otimização os grupos são formados pela adequação de algum critério de agrupamento. Entre esses métodos, o de Tocher apresenta uma inconveniência no agrupamento dos genótipos com maior dissimilaridade: na maioria dos casos, cada genótipo forma um grupo específico (um grupo de apenas um genótipo) em virtude de este agrupamento ser influenciado pela distância dos genótipos já agrupados (VASCONCELOS *et al.*, 2007). Este método, citado por Rao (1952), utiliza um critério de agrupamento que possui a particularidade de apresentar a distância média intragrupo sempre menor que a distância média intergrupo. O método de Tocher tem sido utilizado em estudos de divergência genética de várias culturas, tais como: melancia (SOUZA *et al.*, 2005), alface (OLIVEIRA *et al.*, 2004), cana-de-açúcar (SILVA *et al.*, 2005), aveia (MARCHIORO *et al.*, 2003) e outras.

Há também na literatura o método de Tocher modificado (VASCONCELOS *et al.*, 2007), que difere do original pelo fato de ser adotado critério de aglomeração inverso, de modo que o processo de agrupamento deixa de ser simultâneo e passa a ser sequencial. Os mesmos autores relatam ainda que no método de Tocher modificado não existe influência dos genótipos já agrupados, como ocorre com o método simultâneo (original). Assim, o agrupamento de genótipos com maior dissimilaridade é realizado com maior eficácia pelo método modificado.

2.4.2. Métodos Hierárquicos

Nos métodos hierárquicos de agrupamento, os indivíduos são agrupados por um processo que se repete em vários níveis até que seja estabelecido o dendrograma (diagrama bidimensional em forma de árvore), que é o principal objetivo. Segundo Bussab *et al.* (1990), o dendrograma pode ser considerado a representação simplificada da matriz de dissimilaridade. Uma das maneiras de verificar a capacidade do dendrograma em reproduzir a matriz de

dissimilaridade é construir a matriz cofenética, que é a matriz de distâncias entre os objetos a partir do dendrograma. A verificação da proximidade das duas matrizes de dissimilaridade é fornecida pelo coeficiente de correlação entre os valores da matriz de dissimilaridade e os correspondentes da matriz cofenética, que é chamado coeficiente de correlação cofenética.

As técnicas hierárquicas são as mais amplamente difundidas (SIEGMUND *et al.*, 2004) e envolvem basicamente duas etapas. A primeira se refere à estimação de uma medida de similaridade ou dissimilaridade entre os indivíduos, e a segunda refere-se à adoção de uma técnica de formação de grupos (SANTANA e MALINOVSKI, 2002).

Segundo Cruz *et al.* (2011), os principais métodos de agrupamento hierárquicos são: o método do vizinho mais próximo, o método do vizinho mais distante, o método UPGMA (*Unweighted Pair-Group Method using Arithmetic Averages*), o método do centróide, o método da mediana (ou WPGMC) e o método da variância mínima de Ward. Todos esses métodos operam sobre a matriz de distâncias, dispensando recorrer aos dados originais.

O método da ligação média entre grupos (UPGMA) tem sido comumente utilizado em estudos de diversidade genética, tendo vantagem sobre os demais métodos por considerar médias aritméticas das medidas de dissimilaridade, o que evita caracterizar a dissimilaridade por valores extremos entre os indivíduos considerados (CRUZ e CARNEIRO, 2003), como ocorre com os métodos do vizinho mais próximo e do vizinho mais distante. Outra razão que também favorece o método UPGMA é que ele tende a gerar valores mais altos do coeficiente de correlação cofenética (SOKAL e ROHLF, 1962). Isso significa dizer que ele produz menor distorção quanto à representação das similaridades entre indivíduos de um dendrograma.

Na presença de observações discrepantes, Milligan (1980) verificou que o agrupamento com o métodos da ligação simples, centróide e mediana não foram afetados, ao contrário dos métodos de Ward e UPGMA.

O método de Ward tem sido preferido, em alguns casos, devido o efeito gráfico gerado pelo dendrograma, possibilitando a visualização de grupos bem definidos (ROSEMBURG, 1984).

De forma geral, os dendrogramas obtidos com os métodos hierárquicos apresentam estruturas de agrupamentos de objetos homogêneos, embora não exista critério objetivo para determinar um ponto de corte no dendrograma, ou seja, para determinar quais grupos foram formados.

O número de grupos nos métodos hierárquicos pode ser definido por alguns critérios: razões práticas do pesquisador; análise visual das ramificações do dendrograma, sendo o corte estabelecido em pontos onde há mudança abrupta da ramificação; e critérios estatísticos, como por exemplo: desvio padrão médio e coeficiente de determinação, propostos por Khattree e Naik (2000); método de Mojena (1977), baseado no tamanho relativo dos níveis de fusões (distâncias) no dendrograma.

Depois de obtido o dendrograma, faz-se necessário quantificar a consistência do agrupamento. O coeficiente de correlação cofenética é um coeficiente que quantifica a semelhança entre a matriz de distâncias do dendrograma (matriz cofenética) e a matriz de distâncias originais. Quanto maior for o valor da correlação, menor será a distorção provocada pelo agrupamento. Manly (2008) afirma que o valor da correlação entre as duas matrizes pode ser testado pela aplicação do teste de aleatorização de Mantel (1967).

Bussab *et al.* (1990) recomendam a aplicação de mais de um método hierárquico ao conjunto de dados e que o resultado mais concordante seja aceito como o mais adequado. Isso evita que a classificação seja um mero artefato da técnica utilizada, já que cada técnica impõe determinada estrutura aos dados.

Além dos procedimentos já citados para verificação da consistência do agrupamento, alguns autores (CRUZ *et al.*, 2011, MINGOTI, 2005, FERREIRA, 2008) salientam que a análise discriminante de Fisher (1936) pode ser aplicada utilizando-se da partição final dos indivíduos em grupos obtidos pelo método de agrupamento e verificando-se o percentual de acerto das classificações dos indivíduos para a validação do agrupamento.

3. MATERIAL E MÉTODOS

3.1. Descrição do Experimento

O experimento foi realizado no período de março a novembro de 2010 em área experimental pertencente ao setor de olericultura do Departamento de Fitotecnia da Universidade Federal de Viçosa (UFV), município de Viçosa, Zona da Mata de Minas Gerais, nas coordenadas geográficas: 20° 45' de latitude sul e 42° 51' de longitude oeste, com altitude média de 650 m.

O delineamento experimental foi o de blocos completos casualizados com quatro repetições. Foram avaliados 89 acessos de alho registrados no Banco de Germoplasma de Hortaliças (BGH/UFV), conforme o Quadro 1.

As parcelas (unidades experimentais) foram constituídas por quatro fileiras transversais de 1,0 m de comprimento, com plantas no espaçamento de 0,25 x 0,10 m, totalizando 40 plantas; foram consideradas como úteis as plantas das duas fileiras centrais.

As plantas colhidas foram submetidas ao processo de cura de campo e de galpão, por 3 e 60 dias, respectivamente, expostas ao sol e em galpão. Após a cura de campo, as plantas foram submetidas à toailete cortando a parte aérea a 1,0 cm acima dos bulbos e retirando-se as raízes.

Quadro 1. Relação dos 89 acessos de alho registrados no Banco de Germoplasmas de Hortaliças da Universidade Federal de Viçosa (BGH/UFV).

Nº	Acesso	Nº	Acesso	Nº	Acesso	Nº	Acesso	Nº	Acesso	Nº	Acesso
1	<i>Amarante Derly</i>	16	4812	31	4835	46	5941	61	5958	76	7614
2	0492	17	4814	32	4837	47	5943	62	5959	77	7615
3	4328	18	4815	33	4838	48	5944	63	5960	78	7616
4	4400	19	4816	34	4839	49	5946	64	5961	79	7617
5	4483	20	4817	35	4843	50	5947	65	5962	80	7618
6	4484	21	4818	36	4844	51	5948	66	5963	81	7619
7	4491	22	4820	37	4899	52	5949	67	5964	82	7620
8	4493	23	4822	38	5366	53	5950	68	5965	83	7621
9	4495	24	4823	39	5934	54	5951	69	5966	84	7622
10	4499	25	4825	40	5935	55	5952	70	6389	85	7623
11	4500	26	4826	41	5936	56	5953	71	6390	86	7624
12	4503	27	4827	42	5937	57	5954	72	6394	87	7625
13	4505	28	4829	43	5938	58	5955	73	6396	88	7626
14	4567	29	4830	44	5939	59	5956	74	6397	89	7627
15	4662	30	4832	45	5940	60	5957	75	6398		

3.2. Caracteres Avaliados

Após o período de cura em galpão, foram avaliados os seguintes caracteres:

- Diâmetro do bulbo (DIAM, em milímetros).
- Comprimento do bulbo (COMP, em milímetros), medido, assim como o DIAM, por meio de um paquímetro digital “Starrett”, resolução 0,01mm.
- Peso médio de bulbo (PMB, em gramas), medido por meio de uma balança de precisão eletrônica “Bioprecisa”, modelo JA3003N.
- Número de bulbilhos por bulbo (NbB).
- Produtividade (PT, em t ha⁻¹), obtida ao multiplicar o peso médio dos bulbos dos acessos pela população de plantas presentes em área equivalente a 8.000 m² (área útil do experimento).

3.3. Análises Estatísticas

3.3.1. Análise de Variância e Estimação de Parâmetros Genéticos

Foi realizada análise de variância univariada para os caracteres avaliados com base na média das parcelas, visando avaliar a existência de variabilidade genética significativa entre os acessos, utilizando-se o modelo estatístico:

$$Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$$

em que:

Y_{ij} : valor fenotípico observado do i -ésimo acesso no j -ésimo bloco;

μ : média geral do caractere;

α_i : efeito do i -ésimo acesso ($i = 1, 2, 3, \dots, g$; $g = 89$);

β_j : efeito do j -ésimo bloco ($j = 1, 2, \dots, r$; $r = 4$);

ε_{ij} : efeito do erro experimental, sendo $\varepsilon_{ij} \sim \text{NID}(0, \sigma^2)$.

Como os resultados obtidos são válidos apenas para os materiais genéticos em questão, considerou-se modelo fixo. Assim, a hipótese testada pela estatística F será $H_0: \alpha_i = 0, \forall i$.

O esquema da análise de variância com as Esperanças dos Quadrados Médios [E(QM)], foi realizado de acordo com a metodologia proposta por Steel e Torrie (1980), conforme o Quadro 2.

Quadro 2. Esquema da análise de variância e esperanças de quadrados médios de um modelo em blocos casualizados, com efeitos fixos de acessos.

FV	GL	QM	E(QM)	F
Blocos	$r - 1$	QMB	$\sigma^2 + g\Phi_b$	
Acessos	$g - 1$	QMG	$\sigma^2 + r\Phi_g$	QMG/QMR
Resíduo	$(r - 1)(g - 1)$	QMR	σ^2	

em que:

σ^2 : componente de variância residual;

Φ_b : componente de efeito quadrático associado aos blocos;

Φ_g : componente de efeito quadrático associado aos acessos (variabilidade genotípica), sendo:

$$\Phi_b = \frac{\sum_{j=1}^r \beta_j^2}{r-1} \quad \text{e} \quad \Phi_g = \frac{\sum_{i=1}^g \alpha_i^2}{g-1}$$

As estimativas dos componentes quadráticos associados aos efeitos fixos, bem como de parâmetros genéticos e não-genéticos, serão obtidas considerando-se as seguintes expressões, apresentadas por Cruz e Regazzi (1994):

a) Variância fenotípica entre médias de acessos

$$\hat{\sigma}_f^2 = \frac{QMG}{r}$$

b) Variabilidade genotípica

$$\hat{\Phi}_g = \frac{QMG - QMR}{r}$$

c) Variância de ambiente entre médias de acessos

$$\hat{\sigma}^2 = \frac{QMR}{r}$$

d) Coeficiente de determinação genotípico

$$\hat{h}^2 = \frac{\hat{\Phi}_g}{\hat{\sigma}_f^2}$$

e) Coeficiente de variação genotípico

$$CV_g (\%) = 100 \frac{(\hat{\Phi}_g)^{1/2}}{\hat{\mu}}$$

f) Coeficiente de variação experimental

$$CV_e (\%) = 100 \frac{(\hat{\sigma}^2)^{1/2}}{\hat{\mu}}$$

g) Coeficiente de variação relativa

$$CV_r = \frac{CV_g}{CV_e} = \left(\frac{\hat{\Phi}_g}{\hat{\sigma}^2} \right)^{1/2}$$

h) Acurácia seletiva (RESENDE e DUARTE, 2007)

$$\hat{r}_{gg} = (1 - 1/F)^{1/2}$$

Em que $F = QMG / QMR$.

3.3.2. Dissimilaridade Entre Acessos

A divergência genética entre acessos foi avaliada pela técnica multivariada de análise de agrupamento aplicada às médias dos acessos.

A medida de dissimilaridade adotada foi a distância generalizada de Mahalanobis (D^2), pois se pressupôs que os caracteres estudados apresentassem certo grau de correlação, não sendo totalmente independentes. Neste sentido, a distância de Mahalanobis tem a vantagem, em relação às distâncias euclidianas, de levar em consideração a correlação entre os caracteres considerados, podendo ser obtida pela expressão:

$$D_{ii'}^2 = \delta'_{ii'} \Psi^{-1} \delta_{ii'}$$

em que:

$D_{ii'}^2$ é a distância generalizada de Mahalanobis entre os acessos i e i' , $i = 1, 2, \dots, g$.

$$\delta'_{ii'} = [d_1 \ d_2 \ \dots \ d_p]_{1 \times p}, \text{ sendo } d_j = Y_{ij} - Y_{i'j} \text{ para acessos } i \text{ e } i'.$$

Y_{ij} é a média do i -ésimo acesso em relação à j -ésima variável, $j = 1, 2, \dots, p$.

Ψ^{-1} é a inversa da matriz de variâncias e covariâncias residuais, cujos elementos da diagonal principal são os quadrados médios do resíduo (QMR) da análise de variância e os elementos fora dela são os produtos médios do resíduo (PMR) entre os pares de variáveis, podendo ser obtido por:

$$\text{PMR}(Y_1, Y_2) = \frac{\text{QMR}(Y_1 + Y_2) - \text{QMR}(Y_1) - \text{QMR}(Y_2)}{2}$$

uma vez que se verifica a relação:

$$\text{Cov}(Y_1, Y_2) = \frac{V(Y_1 + Y_2) - V(Y_1) - V(Y_2)}{2}$$

Exemplo ilustrativo: Considere um experimento instalado no delineamento de blocos casualizados com quatro repetições em que foram comparados quatro acessos de alho, em relação a três variáveis (Y_1 , Y_2 e Y_3). Os resultados hipotéticos são:

Tabela 1. Exemplo hipotético para obtenção da matriz de distâncias de Mahalanobis: Médias dos acessos para as variáveis Y_1 , Y_2 e Y_3 .

Acesso	Y_1	Y_2	Y_3
1	39,42	29,54	20,81
2	39,43	30,66	23,49
3	39,14	31,56	23,59
4	52,74	39,66	47,61

Tabela 2. Exemplo hipotético para obtenção da matriz de distâncias de Mahalanobis: Análise de variância.

FV	GL	Quadrados Médios					
		Y_1	Y_2	Y_3	Y_1+Y_2	Y_1+Y_3	Y_2+Y_3
Blocos	3	110,34	20,39	191,70	223,42	585,90	331,30
Acessos	3	180,08	85,08	630,60	507,88	1480,30	1176,80
Resíduo	9	25,43	22,86	130,00	88,56	260,40	244,30

Para obtenção da matriz Ψ é necessário calcular os PMR para cada par de variáveis:

$$\text{PMR}(Y_1, Y_2) = \frac{88,56 - 25,43 - 22,86}{2} = 20,13$$

$$\text{PMR}(Y_1, Y_3) = \frac{260,40 - 25,43 - 130,00}{2} = 52,48$$

$$\text{PMR}(Y_2, Y_3) = \frac{244,30 - 22,86 - 130,00}{2} = 45,72$$

Então:

$$\Psi = \begin{array}{c} Y_1 \\ Y_2 \\ Y_3 \end{array} \begin{array}{ccc} Y_1 & Y_2 & Y_3 \\ \left[\begin{array}{ccc} 25,43 & 20,13 & 52,48 \\ 20,13 & 22,86 & 45,72 \\ 52,48 & 45,72 & 130,00 \end{array} \right]_{3 \times 3} \end{array} \quad \therefore \Psi^{-1} = \begin{array}{c} Y_1 \\ Y_2 \\ Y_3 \end{array} \begin{array}{ccc} Y_1 & Y_2 & Y_3 \\ \left[\begin{array}{ccc} 0,2597 & -0,0645 & -0,0821 \\ -0,0645 & 0,1639 & -0,0316 \\ -0,0821 & -0,0316 & 0,0519 \end{array} \right]_{3 \times 3} \end{array}$$

Calcula-se agora as distâncias entre os pares de acessos em relação a cada variável:

$$\delta_{12} = \begin{bmatrix} 39,42 - 39,43 \\ 29,54 - 30,66 \\ 20,81 - 23,49 \end{bmatrix}_{3 \times 1}, \quad \delta_{13} = \begin{bmatrix} 39,42 - 39,14 \\ 29,54 - 31,56 \\ 20,81 - 23,59 \end{bmatrix}_{3 \times 1}, \quad \dots, \delta_{34}$$

Assim, a distância D^2 entre os acessos 1 e 2 é obtida fazendo: $D_{12}^2 = \delta'_{12} \Psi^{-1} \delta_{12}$

$$D_{12}^2 = \begin{bmatrix} -0,01 & -1,12 & -2,68 \end{bmatrix}_{1 \times 3} \begin{bmatrix} 0,2597 & -0,0645 & -0,0821 \\ -0,0645 & 0,1639 & -0,0316 \\ -0,0821 & -0,0316 & 0,0519 \end{bmatrix}_{3 \times 3} \begin{bmatrix} -0,01 \\ -1,12 \\ -2,68 \end{bmatrix}_{3 \times 1} = 0,382$$

E assim sucessivamente, até o último par de acessos (3 e 4). A matriz de distâncias generalizadas de Mahalanobis é, então:

$$\mathbf{D}^2 = \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{cccc} 1 & 2 & 3 & 4 \\ \left[\begin{array}{cccc} 0 & 0,382 & 0,943 & 7,017 \\ 0,382 & 0 & 0,191 & 7,614 \\ 0,943 & 0,191 & 0 & 8,626 \\ 7,017 & 7,614 & 8,626 & 0 \end{array} \right]_{4 \times 4} \end{array}$$

Além disso, a estatística D^2 possibilitou quantificar a contribuição relativa dos caracteres para a divergência genética utilizando o critério proposto por Singh (1981), baseado na estatística S_j . Nesse caso considera-se que:

$$D_{ii'}^2 = \delta'_{ii'} \Psi^{-1} \delta_{ii'} = \sum_{j=1}^p \sum_{j'=1}^p \omega_{jj'} d_j d_{j'}$$

em que $\omega_{jj'}$ é o elemento da j -ésima linha e j' -ésima coluna da inversa da matriz de variâncias e covariâncias residuais.

Os valores percentuais de S_j constituem a medida da importância relativa da variável j para o estudo da diversidade genética, expressado por:

$$S_j = \frac{\sum_{j'=1}^p \omega_{jj'} d_j d_{j'}}{\sum_{j=1}^p S_j} \times 100$$

Para o exemplo em questão, pode-se calcular a contribuição relativa de cada variável (Y_1 , Y_2 e Y_3) fazendo:

$$D_{12}^2 = \omega_{11} d_1 d_1 + \omega_{21} d_2 d_1 + \omega_{31} d_3 d_1 + \omega_{12} d_1 d_2 + \omega_{22} d_2 d_2 + \omega_{23} d_2 d_3 + \\ \omega_{31} d_3 d_1 + \omega_{32} d_3 d_2 + \omega_{33} d_3 d_3$$

$$D_{12}^2 = (0,2597)(-0,01)(-0,01) + (-0,0645)(-1,12)(-0,01) + \\ (-0,0821)(-2,68)(-0,01) + (-0,0645)(-0,01)(-1,12) + \dots \\ + (0,0519)(-2,68)(-2,68) \\ = 0,382$$

A parte de D_{12}^2 devida a Y_1 equivale a:

$$S_{.1}(D_{12}^2) = \omega_{11} d_1 d_1 + \omega_{21} d_2 d_1 + \omega_{31} d_3 d_1 \\ = (0,2597)(-0,01)(-0,01) + (-0,0645)(-1,12)(-0,01) + \\ (-0,0821)(-2,68)(-0,01) \\ = -0,00289$$

Assim como a parte de D_{13}^2 devida a Y_1 equivale a:

$$\begin{aligned} S_{.1}(D_{13}^2) &= \omega_{11}d_1d_1 + \omega_{21}d_2d_1 + \omega_{31}d_3d_1 \\ &= (0,2597)(0,28)(0,28) + (-0,0645)(-2,02)(0,28) + \\ &\quad (-0,0821)(-2,78)(0,28) \\ &= 0,12074 \end{aligned}$$

Sucessivamente, tem-se:

$$\begin{aligned} S_{.1}(D_{14}^2) &= 8,07437 & S_{.1}(D_{24}^2) &= 11,9238 \\ S_{.1}(D_{23}^2) &= 0,04105 & S_{.1}(D_{34}^2) &= 14,10902 \end{aligned}$$

Então, a importância de cada variável na divergência entre os acessos é:

$$\begin{aligned} S_{.1} &= S_{.1}(D_{12}^2) + S_{.1}(D_{13}^2) + S_{.1}(D_{14}^2) + S_{.1}(D_{23}^2) + S_{.1}(D_{24}^2) + S_{.1}(D_{34}^2) \\ &= 34,26 \end{aligned}$$

$$\begin{aligned} S_{.2} &= S_{.2}(D_{12}^2) + S_{.2}(D_{13}^2) + S_{.2}(D_{14}^2) + S_{.2}(D_{23}^2) + S_{.2}(D_{24}^2) + S_{.2}(D_{34}^2) \\ &= 3,53 \end{aligned}$$

$$\begin{aligned} S_{.3} &= S_{.3}(D_{12}^2) + S_{.3}(D_{13}^2) + S_{.3}(D_{14}^2) + S_{.3}(D_{23}^2) + S_{.3}(D_{24}^2) + S_{.3}(D_{34}^2) \\ &= 6,06 \end{aligned}$$

Em porcentagem de importância relativa, tem-se:

$$(\%)S_{.1} = \frac{34,26}{34,26 + 3,53 + 6,06} \times 100 = 78,13$$

$$(\%)S_{.2} = \frac{3,53}{34,26 + 3,53 + 6,06} \times 100 = 8,05$$

$$(\%)S_{.3} = \frac{6,06}{34,26 + 3,53 + 6,06} \times 100 = 13,82$$

Além de levar em consideração as variâncias e covariâncias residuais das variáveis pela distância generalizada de Mahalanobis, os procedimentos de agrupamento foram realizados considerando-se as escalas das variáveis, por meio de: Y_{ij} / s_j , em que s_j é o desvio padrão da variável Y_j .

3.3.3. Métodos de Agrupamento

3.3.3.1. Método de Tocher

O método, como apresentado por Cruz *et al.* (2011), requer a obtenção da matriz de dissimilaridade, sobre a qual é identificado o par de indivíduos mais similares. Esses indivíduos formarão o grupo inicial. A partir daí é avaliada a possibilidade de inclusão de novos indivíduos, adotando-se o critério de que a distância média intragrupo deve ser menor que a distância média intergrupo.

Neste método, a entrada de um indivíduo em um grupo sempre aumenta o valor médio da distância dentro do grupo. A inclusão, ou não, do indivíduo k no grupo é, então, feita considerando:

$$\text{Se } \frac{d_{(\text{grupo})k}}{n} \leq \theta, \text{ inclui-se o indivíduo } k \text{ no grupo;}$$

$$\text{Se } \frac{d_{(\text{grupo})k}}{n} > \theta, \text{ o indivíduo } k \text{ não é incluído no grupo.}$$

sendo n o número de indivíduos que constitui o grupo original e θ é a maior dentre as menores distâncias envolvendo cada acesso (critério de agrupamento).

Nesse caso, a distância entre o indivíduo k e o grupo formado pelos indivíduos ij é dada por: $d_{(ij)k} = d_{ik} + d_{jk}$

Exemplo ilustrativo: Considerando a matriz de distâncias \mathbf{D}^2 obtida no exemplo anterior (item 3.3.2), tem-se que o método de Tocher segue os passos:

Passo 1: Identifica-se na matriz de dissimilaridade a menor distância envolvendo cada acesso.

Acesso	1	2	3	4
Menor distância	0,382	0,191	0,191	7,017

Critério de agrupamento: $\theta = 7,017$ (maior entre os mínimos)

Passo 2: Agrupa-se os indivíduos que apresentam a menor distância na matriz de dissimilaridade

Grupo I: acessos 2 e 3 ($d_{12} = 0,191$)

Distâncias entre o grupo I e os acessos restantes:

$$d_{(23)1} = d_{12} + d_{13} = 0,382 + 0,943 = 1,325$$

$$d_{(23)4} = d_{24} + d_{34} = 7,614 + 8,626 = 16,24$$

O acesso 1 é o mais similar ao grupo I. Avalia-se a possibilidade de inclusão deste acesso no grupo 1 por meio de:

$$\frac{d_{(23)1}}{2} \leq \theta \quad \therefore \quad \frac{1,325}{2} = 0,6625 \leq \theta$$

Assim o acesso 1 é incluído no grupo I.

Passo 3: Calcula-se a distância entre o grupo recém formado e o acesso 4

$$d_{(23,1)4} = d_{24} + d_{34} + d_{14} = 7,614 + 8,626 + 7,017 = 23,257$$

O acesso 4 é incluído no grupo se:

$$\frac{d_{(23,1)4}}{3} \leq \theta \therefore \frac{23,257}{3} = 7,7523 > \theta$$

Assim, o acesso 4 não é incluído no grupo.

O processo termina com a formação do grupo II constituído apenas pelo acesso 4 e o resultado é:

Tabela 3. Exemplo hipotético para o método de Tocher: Resultado do agrupamento.

Grupo	Acessos	Distância média intragrupo	Distância média intergrupo
I	1, 2 e 3	0,5033	$d_{I,II} = 7,7523$
II	4	--	--

$$d_{I,II} = \frac{d_{(123)4}}{3} = 7,7523$$

3.3.3.2. Método de Tocher Modificado

Este método, proposto por Vasconcelos *et al.* (2007), difere do original pelo fato de ser adotado critério (θ) diferenciado para inclusão de acessos em cada grupo que é formado, de modo que o processo de agrupamento deixa de ser simultâneo e passa a ser sequencial. O grupo inicial é formado da mesma forma que no método original. A decisão de incluir um indivíduo no grupo é tomada por meio da comparação entre o acréscimo no valor médio da distância dentro do grupo e um nível máximo (θ_1) da medida de dissimilaridade encontrado no conjunto das menores distâncias envolvendo cada indivíduo.

Na formação do próximo grupo, o procedimento é similar, diferindo apenas pelo fato de que é o valor máximo (θ_2) da medida de dissimilaridade encontrado no conjunto das menores distâncias envolvendo cada indivíduo,

porém excluindo as informações daqueles anteriormente agrupados, e assim sucessivamente.

Exemplo ilustrativo: Considerando a matriz de distâncias \mathbf{D}^2 obtida no exemplo do item 3.3.2, tem-se que o método de Tocher modificado segue os passos:

Passo 1: Idêntico ao passo 1 do método de Tocher original

Critério de agrupamento: $\theta_1 = 7,017$ (maior entre os mínimos)

Passo 2: Idêntico ao passo 1 do método de Tocher original

O acesso 1 é incluído ao grupo I (acessos 2 e 3).

Passo 3: Definição do novo critério de agrupamento (θ_2), que é a maior entre as menores distâncias envolvendo cada acesso, porém excluindo-se as informações dos indivíduos do grupo anterior.

Critério de agrupamento: restando apenas o acesso 4, o critério de agrupamento não muda, isto é, permanece o critério anterior (θ_1).

$$d_{(23,1)4} = d_{24} + d_{34} + d_{14} = 7,614 + 8,626 + 7,017 = 23,257$$

O acesso 4 é incluído no grupo se:

$$\frac{d_{(23,1)4}}{3} \leq \theta_1 \quad \therefore \quad \frac{23,257}{3} = 7,7523 > \theta_1$$

Assim, o acesso 4 não é incluído no grupo.

Neste caso, o processo termina de forma idêntica ao método de Tocher original.

3.3.3.3. Método UPGMA

Pelo método da ligação média entre grupos não ponderado, mais conhecido como UPGMA (*Unweighted Pair-Group Method using Arithmetic Averages*), utiliza-se as médias aritméticas (não ponderadas) das medidas de dissimilaridade, evitando assim, caracterizar a dissimilaridade por valores extremos (mínimo e máximo) entre os acessos considerados.

Como regra geral, a construção do dendrograma é estabelecida pelos acessos de menor dissimilaridade. A distância entre um acesso k e um grupo, formado pelos acessos i e j , é dada por:

$$d_{(ij)k} = \text{média}(d_{ik}; d_{jk}) = \frac{d_{ik} + d_{jk}}{2}$$

ou seja, $d_{(ij)k}$ é dada pela média do conjunto das distâncias dos pares de indivíduos (i e k) e (j e k).

Lance e Williams (1967) propuseram um método bastante geral para atualização da matriz de distâncias. Ferreira (2008) comenta que o método engloba todos os outros métodos hierárquicos aglomerativos, bastando para isso determinar adequadamente os parâmetros (α , β e Υ) da fórmula geral proposta para atualizar a distância entre as entidades i e j , recém agrupadas, e uma entidade k . A fórmula de atualização é dada por:

$$d_{(ij)k} = \alpha_i d_{ik} + \alpha_j d_{jk} + \beta d_{ij} + \Upsilon |d_{ik} - d_{jk}|$$

em que, para o método UPGMA:

$$\alpha_i = \frac{n_i}{n_i + n_j}, \alpha_j = \frac{n_j}{n_i + n_j}, \beta = 0 \text{ e } \Upsilon = 0.$$

Nesse caso, d representa a distância entre as entidades i e j , recentemente agrupadas, e a entidade k ; n_i , n_j e n_k representam o número de indivíduos de cada entidade. De forma que essa nova expressão se torne idêntica aquela apresentada anteriormente.

Exemplo ilustrativo: Considerando a matriz de distâncias \mathbf{D}^2 obtida no exemplo do item 3.3.2. e utilizando o algoritmo de Lance e Williams (1967), o método UPGMA segue os passos:

Passo 1: Identifica-se na matriz \mathbf{D}^2 os acessos mais similares: 2 e 3 (grupo inicial)

- Nível de Fusão: 0,191
- Distâncias entre os demais acessos:

$$d_{(23)1} = \frac{n_2}{n_2+n_3}d_{12} + \frac{n_3}{n_2+n_3}d_{13} = \frac{1}{1+1}0,382 + \frac{1}{1+1}0,943 = 0,6625$$

$$d_{(23)4} = \frac{n_2}{n_2+n_3}d_{24} + \frac{n_3}{n_2+n_3}d_{34} = \frac{1}{1+1}7,614 + \frac{1}{1+1}8,626 = 8,12$$

Forma-se a nova matriz de dissimilaridade, de dimensão inferior:

$$\mathbf{D}^2 = \begin{matrix} & \begin{matrix} 1 & (2,3) & 4 \end{matrix} \\ \begin{matrix} 1 \\ (2,3) \\ 4 \end{matrix} & \begin{bmatrix} 0 & 0,6625 & 7,017 \\ 0,6625 & 0 & 8,12 \\ 7,017 & 8,12 & 0 \end{bmatrix}_{3 \times 3} \end{matrix}$$

Passo 2: Identifica-se na nova matriz \mathbf{D}^2 as entidades mais similares: (2,3) e 1

- Nível de fusão: 0,6625
- Distâncias entre as entidades restantes: acesso 4

$$d_{(23,1)4} = \frac{n_{23}}{n_{23}+n_1}d_{(23)4} + \frac{n_1}{n_{23}+n_1}d_{14} = \frac{2}{2+1}8,12 + \frac{1}{2+1}7,017 = 7,7523$$

Passo 3: O processo termina quando o acesso 4 é incluído ao grupo anterior. O nível de fusão é de 7,7523.

Tabela 4. Exemplo hipotético para o método UPGMA: Resultado.

Passo	Fusão		Nível de Fusão	Distância (%)	Número de acessos
	Entidade	Entidade			
1	2	3	0,1910	2,46	2
2	(2,3)	1	0,6625	8,54	3
3	(1,2,3)	4	7,7523	100,00	4

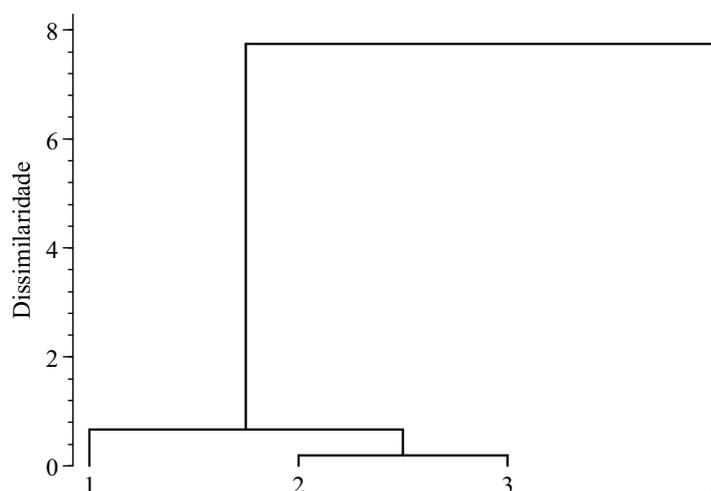


Figura 1. Dendrograma para o exemplo hipotético com o método UPGMA.

3.3.3.4. Método da Variância Mínima de Ward

Ward (1963) propôs um método de agrupamento que baseia-se na mudança de variação dentro dos grupos em formação e entre eles a cada passo do processo de agrupamento. O procedimento também é conhecido como “método da variância mínima”, pois considera como função objetivo minimizar a soma de quadrados dentro dos grupos.

Neste método considera-se, para a formação inicial do grupo, o par de indivíduos que proporciona a menor soma de quadrados dos desvios. Admite-se

que, em qualquer estágio, há perda de informações em razão do agrupamento realizado, o qual pode ser quantificado pela razão entre a soma de quadrados dos desvios dentro do grupo em formação pela soma de quadrado total dos desvios. A soma de quadrados dos desvios dentro é calculada considerando apenas os acessos dentro do grupo em formação e a soma de quadrados dos desvios total é calculada considerando todos os indivíduos disponíveis para análise de agrupamento.

Partindo-se dos dados originais, o agrupamento é feito a partir das somas de quadrados dos desvios entre acessos ou, alternativamente, a partir do quadrado da distância euclidiana, uma vez que se verifica a relação:

$$SQD_{ii'} = \frac{1}{2} d_{ii'}^2, \text{ em que: } SQD_{ii'} = \sum_{j=1}^p SQD_{j(ii')}$$

sendo $SQD_{j(ii')}$ a soma de quadrados dos desvios, para a j -ésima variável, considerando os acessos i e i' .

e

$$d_{ii'}^2 = \sum_{j=1}^p (Y_{ij} - Y_{i'j})^2$$

em que:

$d_{ii'}^2$: quadrado da distância euclidiana entre os acessos i e i' ;

p : número de caracteres avaliados;

Y_{ij} : média do caractere j para o acesso i .

A soma de quadrados dos desvios total é dada por:

$$SQD_{Total} = \frac{1}{g} \sum_{i < i'}^g \sum_{i'}^g d_{ii'}^2$$

sendo g o número de acessos.

Nesta análise de agrupamento, identifica-se na matriz de quadrados das distâncias euclidianas - $d_{ii'}^2$ (ou na matriz de somas dos quadrados dos

desvios - $SQD_{i'}$) o par de acessos que proporciona menor soma de quadrados dos desvios. Com estes acessos agrupados, uma nova matriz de dissimilaridade, de dimensão inferior, é recalculada, considerando que:

$$SQD_{(ijk)} = \frac{1}{k} d_{(ijk)}^2 \quad (k \text{ é o número de acessos no grupo, que neste caso é}$$

igual a 3).

$$d_{(ijk)}^2 = d_{(ij)}^2 + d_{(ij)k}^2 = d_{ij}^2 + d_{ik}^2 + d_{jk}^2$$

e ainda que:

$$SQD_{(ijkm)} = \frac{1}{k} d_{(ijkm)}^2 \quad (k \text{ é o número de acessos no grupo, que neste caso}$$

é igual a 4).

$$d_{(ijkm)}^2 = d_{ij}^2 + d_{ik}^2 + d_{jk}^2 + d_{im}^2 + d_{jm}^2 + d_{km}^2$$

e, assim, sucessivamente.

Como o Método de Ward (1963) originalmente baseia-se na distância euclidiana, neste estudo deve-se considerar que foi utilizado apenas o algoritmo de Ward, uma vez que a dissimilaridade entre os acessos foi quantificada por meio da distância generalizada de Mahalanobis (D^2).

Assim como o método UPGMA, o Método de Ward também pode ser definido e implementado pela família de algoritmos Lance e Williams (1967), a qual recorre a uma fórmula de atualização da distância dos grupos em cada passo do agrupamento, que consiste em otimizar a função objetivo do método. O algoritmo LW para o método de Ward segue a forma:

$$d_{(ij)k} = \alpha_i d_{ik} + \alpha_j d_{jk} + \beta d_{ij} + \Upsilon |d_{ik} - d_{jk}|$$

em que, para o algoritmo de Ward:

$$\alpha_i = \frac{n_i + n_k}{n_i + n_j + n_k}, \quad \alpha_j = \frac{n_j + n_k}{n_i + n_j + n_k}, \quad \beta = \frac{-n_k}{n_i + n_j + n_k} \quad \text{e} \quad \Upsilon = 0.$$

Nesse caso, d representa a distância entre as entidades i e j , recentemente agrupadas, e a entidade k ; n_i , n_j e n_k representam o número de indivíduos em cada entidade.

Nota-se que a distância utilizada no método de Ward leva em consideração um fator de ponderação, que é a diferença entre os tamanhos das entidades que estão sendo comparadas.

Exemplo ilustrativo: Considerando a matriz de distâncias D^2 obtida no exemplo do item 3.3.2. e utilizando o algoritmo de Lance e Williams (1967), o algoritmo de Ward segue os passos:

Passo 1: Identifica-se na matriz D^2 os acessos mais similares: 2 e 3 (grupo inicial)

- Nível de Fusão: 0,191

- Distâncias entre os demais acessos:

$$\begin{aligned} d_{(23)1} &= \frac{n_2 + n_1}{n_1 + n_2 + n_3} d_{12} + \frac{n_3 + n_1}{n_1 + n_2 + n_3} d_{13} + \frac{-n_1}{n_1 + n_2 + n_3} d_{23} \\ &= \frac{1+1}{1+1+1} 0,382 + \frac{1+1}{1+1+1} 0,943 + \frac{-1}{1+1+1} 0,191 = 0,8196 \end{aligned}$$

$$\begin{aligned} d_{(23)4} &= \frac{n_2 + n_4}{n_2 + n_3 + n_4} d_{24} + \frac{n_3 + n_4}{n_2 + n_3 + n_4} d_{34} + \frac{-n_4}{n_2 + n_3 + n_4} d_{23} \\ &= \frac{1+1}{1+1+1} 7,614 + \frac{1+1}{1+1+1} 8,626 + \frac{-1}{1+1+1} 0,191 = 10,763 \end{aligned}$$

Forma-se a nova matriz de dissimilaridade, de dimensão inferior:

$$\mathbf{D}^2 = \begin{matrix} & \begin{matrix} 1 & (2,3) & 4 \end{matrix} \\ \begin{matrix} 1 \\ (2,3) \\ 4 \end{matrix} & \begin{bmatrix} 0 & 0,8196 & 7,017 \\ 0,8196 & 0 & 10,763 \\ 7,017 & 10,763 & 0 \end{bmatrix}_{3 \times 3} \end{matrix}$$

Passo 2: Identifica-se na nova matriz D^2 as entidades mais similares: (2,3) e 1

- Nível de fusão: 0,8196
- Distâncias entre as entidades restantes: acesso 4

$$\begin{aligned}
 d_{(23,1)4} &= \frac{n_{23} + n_4}{n_1 + n_{23} + n_4} d_{(23)4} + \frac{n_1 + n_4}{n_1 + n_{23} + n_4} d_{14} + \frac{-n_4}{n_1 + n_{23} + n_4} d_{(23)1} \\
 &= \frac{2+1}{1+2+1} 10,763 + \frac{1+1}{1+2+1} 7,017 + \frac{-1}{1+2+1} 0,8196 = 11,3758
 \end{aligned}$$

Passo 3: O processo termina quando o acesso 4 é incluído ao grupo anterior. O nível de fusão é de 11,3758.

Tabela 5. Exemplo hipotético para o algoritmo de Ward: Resultado.

Passo	Fusão		Nível de Fusão	Distância (%)	Número de acessos
	Entidade	Entidade			
1	2	3	0,1910	1,68	2
2	(2,3)	1	0,8196	7,20	3
3	(1,2,3)	4	11,3758	100,00	4

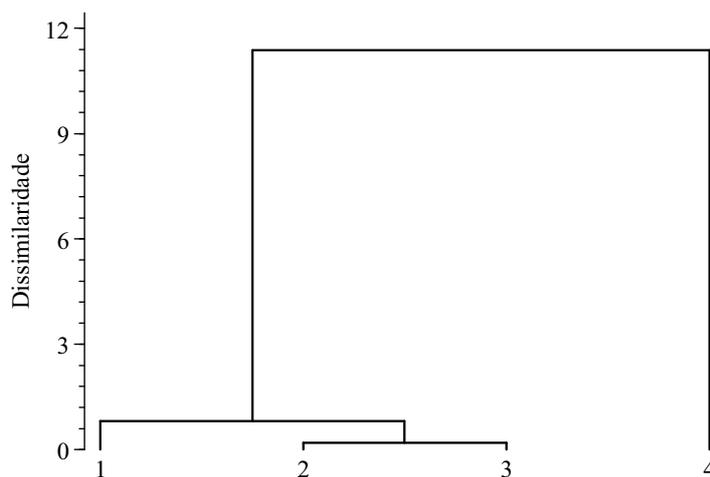


Figura 2. Dendrograma para o exemplo hipotético com o algoritmo de Ward.

3.3.4. Determinação do Número de Grupos

Para determinar o ponto de corte dos dendrogramas gerados pelos métodos hierárquicos e definir o número de grupos, Mojena (1977) sugeriu um procedimento baseado no tamanho relativo dos níveis de fusões (distâncias) no dendrograma. A proposta é selecionar o número de grupos no passo j que, primeiramente, satisfizer a seguinte inequação:

$$\alpha_j > \theta_k$$

em que α_j é o valor de distância do nível de fusão correspondente aos passo j ($j=1, 2, \dots, g-1$), θ_k é o valor referencial de corte, dado por:

$$\theta_k = \bar{\alpha} + k\hat{\sigma}_\alpha$$

onde: $\bar{\alpha}$ e $\hat{\sigma}_\alpha$ são, respectivamente, a média e o desvio padrão dos valores de α ; k é uma constante cujo valor a ser adotado, conforme sugerido por Milligan e Cooper (1985), é de $k = 1,25$ como regra de parada na definição do número de grupos.

Assim, tem-se que:

$$\bar{\alpha} = \frac{1}{g-1} \sum_{j=1}^{g-1} \alpha_j \quad \text{e} \quad \hat{\sigma}_\alpha = \sqrt{\frac{\sum_{j=1}^{g-1} \alpha_j^2 - \frac{1}{g-1} \left(\sum_{j=1}^{g-1} \alpha_j \right)^2}{g-2}}$$

Exemplo ilustrativo: Considerando o resultado obtido com o algoritmo de Ward no item 3.3.3.4, o ponto de corte pelo método de Mojena (1977) é encontrado fazendo para cada nível de fusão:

$$\theta_k = \bar{\alpha} + k\hat{\sigma}_\alpha$$

$$\bar{\alpha} = \frac{0,1910 + 0,8196}{2} = 0,5053$$

$$\hat{\sigma}_{\alpha} = \sqrt{\frac{0,1910^2 + 0,8196^2 - \frac{(0,1910 + 0,8196)^2}{2}}{2-1}} = 0,4444$$

$$\text{Para } k = 1,25, \text{ tem-se: } \theta_k = 0,5053 + 1,25(0,4444) = 1,0608$$

O nível de fusão neste passo do agrupamento ($\alpha = 0,8196$) é inferior ao critério (1,0608), então para o próximo passo tem-se:

$$\bar{\alpha} = \frac{0,1910 + 0,8196 + 11,3758}{3} = 4,1288$$

$$\hat{\sigma}_{\alpha} = \sqrt{\frac{0,1910^2 + 0,8196^2 + 11,3758^2 - \frac{(0,1910 + 0,8196 + 11,3758)^2}{3}}{3-1}} = 6,2839$$

$$\text{Para } k = 1,25, \text{ tem-se: } \theta_k = 4,1288 + 1,25(6,2839) = 11,9836$$

Como o último nível de fusão (11,3758) ainda é inferior ao critério de corte (11,9836) correspondente, pode-se adotar um valor menor para a constante k, tornando o processo mais sensível. Assim, tomando $k = 1,00$, tem-se:

Tabela 6. Exemplo hipotético para determinação do ponto de corte no dendrograma pelo método de Mojena (1977): Resultado.

Passo	α	$\bar{\alpha}$	$\hat{\sigma}_{\alpha}$	θ_k (k=1,25)	θ_k (k=1,00)
1	0,1910	--	--	--	--
2	0,8196	0,5053	0,4444	1,0608	0,9497
3	11,3758	4,1288	6,2839	11,9836	10,4127*

* ponto de corte do dendrograma

Logo o nível de fusão (α) no último passo do agrupamento superou o valor do critério θ_k correspondente e o corte no dendrograma é feito no valor 10,4127, definindo assim dois grupos, sendo o primeiro grupo constituído pelos acessos 1, 2 e 3, e o segundo grupo pelo acesso 4, como foi verificado com os métodos de Tocher e Tocher modificado.

3.3.5. Coeficiente de Correlação Cofenética

Após a aplicação de cada método de agrupamento hierárquico, foi obtido o coeficiente de correlação cofenética entre a matriz de distâncias originais e a matriz cofenética, de acordo com a expressão proposta por Bussab *et al.* (1990) e citada por Albuquerque (2005).

$$r_{\text{cof}} = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^n (c_{ij} - \bar{c})(d_{ij} - \bar{d})}{\left(\sum_{i=1}^{n-1} \sum_{j=i+1}^n (c_{ij} - \bar{c})^2 \right)^{1/2} \left(\sum_{i=1}^{n-1} \sum_{j=i+1}^n (d_{ij} - \bar{d})^2 \right)^{1/2}}$$

em que:

$$\bar{c} = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^n c_{ij}, \quad \bar{d} = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^n d_{ij},$$

c_{ij} é o valor da distância entre os indivíduos i e j na matriz cofenética; d_{ij} é o valor da distância entre os mesmos indivíduos na matriz original de distâncias; n é a dimensão da matriz.

Exemplo ilustrativo: Considerando o resultado obtido com o algoritmo de Ward no item 3.3.3.4 pode-se determinar a matriz cofenética:

$$\mathbf{Cof} = \begin{array}{c} \\ \\ \\ \\ \end{array} \begin{array}{cccc} 1 & 2 & 3 & 4 \\ \left[\begin{array}{cccc} 0 & 0,8196 & 0,8196 & 11,3758 \\ 0,8196 & 0 & 0,191 & 11,3758 \\ 0,8196 & 0,191 & 0 & 11,3758 \\ 11,3758 & 11,3758 & 11,3758 & 0 \end{array} \right]_{4 \times 4} \end{array}$$

Matriz de distâncias originais:

$$\mathbf{D}^2 = \begin{array}{c} \\ \\ \\ \\ \end{array} \begin{array}{cccc} 1 & 2 & 3 & 4 \\ \left[\begin{array}{cccc} 0 & 0,382 & 0,943 & 7,017 \\ 0,382 & 0 & 0,191 & 7,614 \\ 0,943 & 0,191 & 0 & 8,626 \\ 7,017 & 7,614 & 8,626 & 0 \end{array} \right]_{4 \times 4} \end{array}$$

Então a correlação cofenética mede simplesmente o grau de associação entre os elementos abaixo (ou acima) da diagonal principal de ambas as matrizes, necessitando, em nível de cálculos, definir:

$$\bar{c} = \frac{2}{4(4-1)} (0,8196 + \dots + 11,3758) = 5,9929$$

e

$$\bar{d} = \frac{2}{4(4-1)} (0,382 + \dots + 8,626) = 4,1288$$

Assim obtêm-se: $r_{\text{Cof}} = 0,9907$.

3.3.6. Análise Discriminante

Após a obtenção das informações dos grupos de cada agrupamento, foi aplicada a análise discriminante de Fisher (1936) para verificação do poder de separação dos métodos de agrupamento.

A partir de combinações lineares das p variáveis originais de g populações normais p -variadas com matrizes de covariâncias iguais foram estimadas as funções discriminantes canônicas de Fisher, como apresentado por Mingoti (2005):

$$\hat{Z}_j = \hat{\mathbf{a}}'_j \mathbf{X}_{p \times 1}, \quad j=1,2,\dots,s$$

Em que $s = \min(g-1, p)$, $\hat{\mathbf{a}}'_j$ é o j -ésimo autovetor correspondente ao j -ésimo maior autovalor da matriz $\mathbf{W}^{-1}\mathbf{B}$ e tal que $\hat{\mathbf{a}}'_j \mathbf{W} \hat{\mathbf{a}}_j = 1$, sendo as matrizes \mathbf{W} e \mathbf{B} respectivamente definidas por:

$$\mathbf{W}_{p \times p} = \sum_{i=1}^g \sum_{k=1}^{n_i} (\mathbf{X}_{ik} - \bar{\mathbf{X}}_i)(\mathbf{X}_{ik} - \bar{\mathbf{X}}_i)'$$

$$\mathbf{B}_{p \times p} = \sum_{i=1}^g n_i (\bar{\mathbf{X}}_i - \bar{\mathbf{X}})(\bar{\mathbf{X}}_i - \bar{\mathbf{X}})'$$

Sendo \mathbf{X}_{ik} o vetor de observações do elemento amostral k que pertence a população i , $\bar{\mathbf{X}}_i$ é o vetor de médias amostral da população i , $\bar{\mathbf{X}}$ é o vetor de médias amostral considerando todas as n observações conjuntamente e n_i o número de elementos pertencentes à amostra da população i , $i = 1, 2, \dots, g$, $\sum_{i=1}^g n_i = n$. As matrizes \mathbf{W} e \mathbf{B} são chamadas de matrizes de soma de quadrados e produtos cruzados dentro (*within*) e entre (*between*) grupos, e são análogas àquelas utilizadas em análise de variância multivariada (MANOVA).

Após a construção das funções discriminantes, para cada elemento amostral com vetor de observações \mathbf{X}_{ik} , ter-se-á um vetor com escores nestas funções, isto é, $\hat{\mathbf{Z}}'_j = [\hat{\mathbf{a}}'_1 \mathbf{X}_{ik} \quad \hat{\mathbf{a}}'_2 \mathbf{X}_{ik} \quad \dots \quad \hat{\mathbf{a}}'_s \mathbf{X}_{ik}]$. Ter-se-á também os escores das funções discriminantes aplicadas aos vetores de média amostral para cada população, isto é, $\bar{\mathbf{Z}}'_i = [\hat{\mathbf{a}}'_1 \bar{\mathbf{X}}_i \quad \hat{\mathbf{a}}'_2 \bar{\mathbf{X}}_i \quad \dots \quad \hat{\mathbf{a}}'_s \bar{\mathbf{X}}_i]$. Calcula-se então a distância euclidiana entre os vetores $\hat{\mathbf{Z}}'_j$ e $\bar{\mathbf{Z}}'_i$ para todo $i = 1, 2, \dots, g$, sendo o

correspondente elemento amostral (acesso) j classificado na população cuja distância é a menor.

Os escores das duas primeiras funções discriminantes foram plotadas em gráfico de dispersão para visualização das classificações. O poder de separação dos métodos de agrupamento foi avaliado pela taxa de erro aparente (TEA) das funções discriminantes, conforme a expressão:

$$TEA = \frac{1}{n} \sum_{i=1}^g e_i$$

em que n é o número total de classificações, e_i corresponde ao número de classificações erradas em cada grupo.

3.3.7. Considerações Metodológicas Finais

A consistência dos agrupamentos foi avaliada por meio de:

- i. Análise discriminante de Fisher com base nos resultados (grupos) de cada agrupamento, verificando-se a taxa de erro aparente das classificações;
- ii. Coeficiente de correlação cofenética (apenas para os métodos hierárquicos) e, por conseguinte, aplicação do teste de Mantel com 1000 reamostragens, considerando o nível de 5% de significância.

Os resultados obtidos com os métodos de otimização e com os métodos hierárquicos foram confrontados quanto ao poder de separação do agrupamento.

As análises foram realizadas utilizando-se os *softwares* GENES versão 2009.7.0 (CRUZ, 2006) e R (R DEVELOPMENT CORE TEAM, 2011).

4. RESULTADOS E DISCUSSÃO

4.1. Análise de Variância

Pela Tabela 7 verifica-se que houve efeito de acessos ($P < 0,01$) para todas os caracteres avaliados, evidenciando a variabilidade genotípica, o que pode ser corroborado pelos valores do coeficiente de determinação genotípico, variando de 59% (DIAM) a 81% (NbB). É possível também observar que o coeficiente de variação experimental (CVe) manteve-se abaixo de 33%, com exceção do caractere número de bulbilhos por bulbo (CVe = 45,93), provavelmente por se tratar de uma variável que, por natureza, apresenta maior variabilidade e maior suscetibilidade à erros grosseiros na coleta de dados, como ocorre frequentemente em processos de contagem. Não obstante, é interessante verificar que o coeficiente de variação genético (CVg) manteve-se proporcional ao CVe, e a razão entre eles foi maior ou igual a 0,60, o que também significa claro efeito da variabilidade genotípica. Observa-se ainda que, apenas em relação ao NbB, o CVg foi superior ao CVe (48,00 e 45,93, respectivamente).

Segundo Resende e Duarte (2007), sob quatro repetições para cada tratamento, a magnitude do coeficiente de variação experimental (CVe) é inadequada para informar sobre a precisão da avaliação genotípica. Então, tomando-se os valores de acurácia seletiva (AS) como um parâmetro adequado para resumir a avaliação da precisão do experimento, observa-se que as variáveis

relacionadas à produção (PMB, NbB e PT) apresentaram valores de AS não inferiores a 84%, sendo que todas as variáveis apresentaram valores de AS não inferiores a 76%, pertencendo as classes de precisão “alta” ou “muito alta”, de acordo com a classificação proposta por Resende e Duarte (2007) para este parâmetro.

Tabela 7. Resumo da análise de variância e estimativas dos parâmetros genéticos para caracteres produtivos de alho. UFV, Viçosa (MG), 2012.

FV	GL	Quadrados Médios				
		DIAM	COMP	PMB	NbB	PT
Blocos	3	73,0624	21,3197	341,8413	163,6311	35,0243
Acessos	88	68,2806**	44,7123**	289,0816**	753,0216**	29,6036**
Resíduo	264	27,9954	18,2736	84,9877	140,2670	8,7031
Média		43,59	32,94	28,53	25,79	9,13
CVe (%)		12,14	12,98	32,32	45,93	32,32
CVg (%)		7,28	7,80	25,04	48,00	25,04
CVg/CVe		0,60	0,60	0,77	1,05	0,77
Variabilidade		10,07	6,60	51,02	153,18	5,22
Coef. det. genotípico		0,59	0,59	0,70	0,81	0,70
Acurácia seletiva		0,76	0,76	0,84	0,90	0,84

DIAM: diâmetro do bulbo (mm); COMP: comprimento do bulbo (mm); PMB: peso médio do bulbo (g); NbB: número de bulbilhos por bulbo; PT: produtividade (t ha⁻¹); ** P < 0,01

4.2. Dissimilaridade Entre Acessos e Importância de Caracteres

De acordo com as estimativas das distâncias generalizadas de Mahalanobis (D^2) os acessos que apresentaram maior dissimilaridade foram os acessos 13 (BGH 4505) e 61 (BGH 5958), com $D^2 = 32,33$. Os que apresentaram a menor dissimilaridade foram os acessos 34 (BGH 4839) e 35 (BGH 4843), com $D^2 = 0,09$.

Na Tabela 8 é possível verificar a contribuição relativa de cada caractere para a divergência genética entre os acessos, de acordo com a estatística S_j proposta por Singh (1981). Observa-se que os caracteres produtivos (PMB, NbB e PT) foram os principais responsáveis pela divergência genética entre os 89 acessos de alho, contribuindo com 97,19%. O peso médio do bulbo (PMB) e a produtividade (PT) foram os caracteres que exerceram maior influência na

divergência genética, representando 47,46% e 39,74% do total, respectivamente; juntos estes dois últimos somam 87,2% de toda a contribuição. Os caracteres morfológicos (DIAM e COMP) juntos representaram menos de 3% de contribuição.

Tabela 8. Contribuição relativa de cada caractere para a divergência genética entre 89 acessos de alho do BGH/UFV com base na estatística S_j (SINGH, 1981). UFV, Viçosa (MG), 2012.

Caractere	Abreviatura	S_j	Percentual
Peso médio do bulbo (g)	PMB	51648,30	47,46
Produtividade ($t\ ha^{-1}$)	PT	43236,88	39,74
Número de bulbilhos por bulbo	NbB	10870,82	9,99
Comprimento do bulbo (mm)	COMP	3032,16	2,78
Diâmetro do bulbo (mm)	DIAM	14,46	0,01

4.3. Agrupamento com o Método de Tocher

Pelo método de agrupamento de Tocher (original) baseado na distância generalizada de Mahalanobis os 89 acessos de alho foram divididos em 15 grupos, onde o grupo I reuniu 22 acessos (24,7%), o grupo II reuniu 36 acessos (40,4%) e a partir do grupo IX os grupos apresentaram apenas 1 acesso cada, como pode ser visto nas Tabelas 9 e 10. Por este método de otimização verifica-se que um número bastante reduzido de grupos reúne grande parte dos acessos. Vasconcelos *et al.* (2007) argumentam que este método apresenta uma inconveniência no agrupamento de genótipos com maior dissimilaridade: na maioria dos casos, cada genótipo forma um grupo específico (com apenas um genótipo) em virtude de este agrupamento ser influenciado pela distância dos genótipos já agrupados. Isso ocorre devido ao fato de o método se utilizar de um critério global de agrupamento, ou seja, baseia-se na maior entre as menores distâncias encontradas na matriz de dissimilaridade durante todo o processo. Resultados desse tipo também foram encontrados por Fonseca *et al.* (2006) em estudo da divergência genética em café conilon; e por Nascimento Filho *et al.* (2001) em divergência genética entre clones de guaranazeiro.

Tabela 9. Formação dos grupos de 89 acessos de alho do BGH/UFV segundo o método de otimização de Tocher (original) com base na distância generalizada de Mahalanobis. UFV, Viçosa (MG), 2012.

Grupo	Acessos										
I	1	5	9	12	20	25	27	30	32	33	34
	35	37	44	60	68	69	71	76	78	82	86
II	6	8	15	16	17	18	19	22	23	24	31
	36	39	40	45	46	47	49	50	51	52	54
	55	62	63	64	66	72	73	74	77	79	83
	84	85	87								
III	53	56	57	65	75						
IV	21	29	41	42	81	89					
V	2	3	88								
VI	11	26	58	59							
VII	4	10	48								
VIII	14	28	43								
IX	38										
X	7										
XI	13										
XII	80										
XIII	67										
XIV	61										
XV	70										

Critério global de agrupamento: $\theta = 2,29$

Tabela 10. Número de acessos (N) e médias das variáveis para os grupos obtidos com o método de Tocher. UFV, Viçosa (MG), 2012.

Grupo	N	DIAM	COMP	PMB	NbB	PT
I	22	42,53	31,98	27,37	32,88	8,76
II	36	42,34	32,39	25,17	16,35	8,05
III	5	49,20	35,53	36,07	17,58	11,54
IV	6	39,62	27,88	19,54	19,08	6,25
V	3	37,96	30,30	21,15	22,13	6,77
VI	4	48,57	37,78	43,74	55,50	13,99
VII	3	51,45	40,31	47,17	53,21	15,10
VIII	3	48,51	35,94	37,92	46,13	12,14
IX	1	48,43	34,70	32,91	41,88	10,53
X	1	45,01	37,49	33,38	45,00	10,68
XI	1	55,35	41,51	59,36	59,88	18,99
XII	1	35,98	28,91	18,55	7,38	5,94
XIII	1	45,01	29,48	24,63	13,88	7,88
XIV	1	45,87	35,81	25,65	11,38	8,21
XV	1	52,38	38,94	45,05	15,13	14,42

DIAM: diâmetro do bulbo (mm); COMP: comprimento do bulbo (mm); PMB: peso médio do bulbo (g); NbB: número de bulbilhos por bulbo; PT: produtividade ($t\ ha^{-1}$)

Pelo resultado da análise discriminante de Fisher para os grupos obtidos com o agrupamento de Tocher (original) foram detectadas 24 classificações erradas, ou seja, 26,96% foi a taxa de erro aparente, indicando baixa consistência do agrupamento (Tabela 11). Nos grupos I, II, IV e V verificou-se percentuais de classificações corretas inferiores a 70%. Observa-se também que os grupos com maior número de acessos (I e II) apresentaram classificações erradas de vários outros grupos. Aproximadamente 60% das classificações erradas do método para o grupo I foram para o grupo IV (27,3%).

Tabela 11. Porcentagem de classificação correta (diagonal principal) e incorreta de cada grupo segundo o agrupamento de 89 acessos de alho pelo método de Tocher (original). UFV, Viçosa (MG), 2012.

Grupo	Classificação														
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV
I	54,6	9,1	0,0	27,3	0,0	0,0	0,0	4,6	4,6	0,0	0,0	0,0	0,0	0,0	0,0
II	2,8	69,4	8,3	8,3	2,8	0,0	0,0	0,0	0,0	0,0	0,0	2,8	0,0	5,6	0,0
III	0,0	0,0	100	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
IV	16,7	0,0	0,0	66,7	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	16,7	0,0	0,0
V	0,0	0,0	0,0	0,0	66,7	0,0	0,0	0,0	0,0	0,0	0,0	33,3	0,0	0,0	0,0
VI	0,0	0,0	0,0	0,0	0,0	100	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
VII	0,0	0,0	0,0	0,0	0,0	0,0	100	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
VIII	0,0	0,0	0,0	0,0	0,0	0,0	0,0	100	0,0	0,0	0,0	0,0	0,0	0,0	0,0
IX	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	100	0,0	0,0	0,0	0,0	0,0	0,0
X	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	100	0,0	0,0	0,0	0,0	0,0
XI	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	100	0,0	0,0	0,0	0,0
XII	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	100	0,0	0,0	0,0
XIII	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	100	0,0	0,0
XIV	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	100	0,0
XV	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	100

Taxa de erro aparente: 26,96 % (24 classificações erradas)

4.4. Agrupamento com o Método de Tocher Modificado

De acordo com o método de Tocher modificado os 89 acessos foram separados em 7 grupos, conforme a Tabela 12. É importante destacar que o

método de Tocher ao passar de simultâneo para sequencial, ou seja, quando passa-se a utilizar um critério local para definir o critério de inclusão de um acesso num grupo o número de grupos foi reduzido em mais de 50%, isto é, o que antes tinha sido dividido em 15 grupos agora encontra-se em 7, com apenas um grupo (VII) contendo um acesso (acesso 13), que, não obstante, apresenta bastante dissimilaridade em relação a vários outros acessos, sendo a maior delas com o acesso 61, como citado no item 4.2. Corrobora a afirmativa o fato de o acesso 13 também ficar isolado em um grupo no agrupamento obtido com o método original de Tocher.

Tabela 12. Formação dos grupos de 89 acessos de alho do BGH/UFV segundo o método de otimização de Tocher modificado com base na distância generalizada de Mahalanobis. UFV, Viçosa (MG), 2012.

Grupo	Acessos										θ_i
I	1	5	9	12	20	25	27	30	32	33	2,29
	34	35	37	44	60	68	69	71	76	78	
	82	86									
II	6	8	15	16	17	18	19	22	23	24	2,29
	31	36	39	40	45	46	47	49	50	51	
	52	54	55	62	63	64	66	72	73	74	
	77	79	83	84	85	87					
III	53	56	57	65	70	75					3,28
IV	2	3	21	29	41	42	80	81	88	89	5,28
V	4	7	10	11	14	26	28	43	48	58	5,28
	59										
VI	38	61	67								16,25
VII	13										--

θ_i : critério local de agrupamento

A redução no número de grupos no método sequencial ocorre devido ao fato de que o critério de agrupamento adotado (θ_i) sempre aumenta durante o processo, como pode ser observado ainda na Tabela 12. Segundo Vasconcelos *et al.* (2007) o valor do critério de agrupamento, estimado após a formação de um novo grupo, possui proporcionalidade quanto à dissimilaridade existente entre os acessos do grupo.

O grupo I reuniu 22 acessos (24,7%), de forma idêntica ao grupo I do método original de Tocher, ou seja, com os mesmos acessos. O mesmo pode ser dito em relação ao grupo II, com 36 acessos (40,4%), o que é evidente pelo fato de que o critério de agrupamento é o mesmo daquele adotado no método original ($\theta_i = 2,29$).

Na Tabela 13 encontram-se as médias dos grupos para cada variável. Destaca-se o grupo VII, constituído apenas pelo acesso 13, como sendo o grupo com as médias consideravelmente superiores às médias dos demais grupos, principalmente em relação ao peso médio do bulbo e produtividade.

Tabela 13. Número de acessos (N) e médias das variáveis para os grupos obtidos com o método de Tocher modificado. UFV, Viçosa (MG), 2012.

Grupo	N	DIAM	COMP	PMB	NbB	PT
I	22	42,53	31,98	27,37	32,88	8,76
II	36	42,34	32,39	25,17	16,35	8,05
III	6	49,73	36,10	37,56	17,17	12,02
IV	10	38,76	28,71	19,92	18,83	6,38
V	11	49,01	37,94	42,15	51,36	13,49
VI	3	46,44	33,33	27,73	22,38	8,87
VII	1	55,35	41,51	59,36	59,88	18,99

DIAM: diâmetro do bulbo (mm); COMP: comprimento do bulbo (mm); PMB: peso médio do bulbo (g); NbB: número de bulbilhos por bulbo; PT: produtividade (t ha⁻¹)

Pelo resultado da análise discriminante de Fisher para os grupos obtidos com o agrupamento de Tocher modificado foram detectadas 17 classificações erradas, ou seja, 19,10% foi a taxa de erro aparente (Tabela 14). Embora a taxa de erro tenha reduzido em sete pontos percentuais em relação ao método original de Tocher, verifica-se que ainda é baixa a consistência do agrupamento. O maior número de classificações erradas foi observado no grupo I, em que apenas 63,6% dos acessos foram corretamente classificados.

Tabela 14. Porcentagem de classificação correta (diagonal principal) e incorreta de cada grupo segundo o agrupamento de 89 acessos de alho pelo método de Tocher modificado. UFV, Viçosa (MG), 2012.

Grupo	Classificação						
	I	II	III	IV	V	VI	VII
I	63,6	9,1	0,0	22,7	4,6	0,0	0,0
II	5,6	77,8	2,8	8,3	0,0	5,6	0,0
III	0,0	0,0	100	0,0	0,0	0,0	0,0
IV	0,0	0,0	0,0	90,0	0,0	10,0	0,0
V	0,0	0,0	0,0	0,0	100	0,0	0,0
VI	0,0	0,0	0,0	0,0	0,0	100	0,0
VII	0,0	0,0	0,0	0,0	0,0	0,0	100

Taxa de erro aparente: 19,10 % (17 classificações erradas)

4.5. Agrupamento com o Método UPGMA

Utilizando o método da ligação média entre grupos (UPGMA) foi possível identificar um ponto de corte no dendrograma (Figura 3) na distância de 3,90, valor do critério θ_k do método de Mojena que corresponde a 28,2% da distância máxima observada nos níveis de fusão.

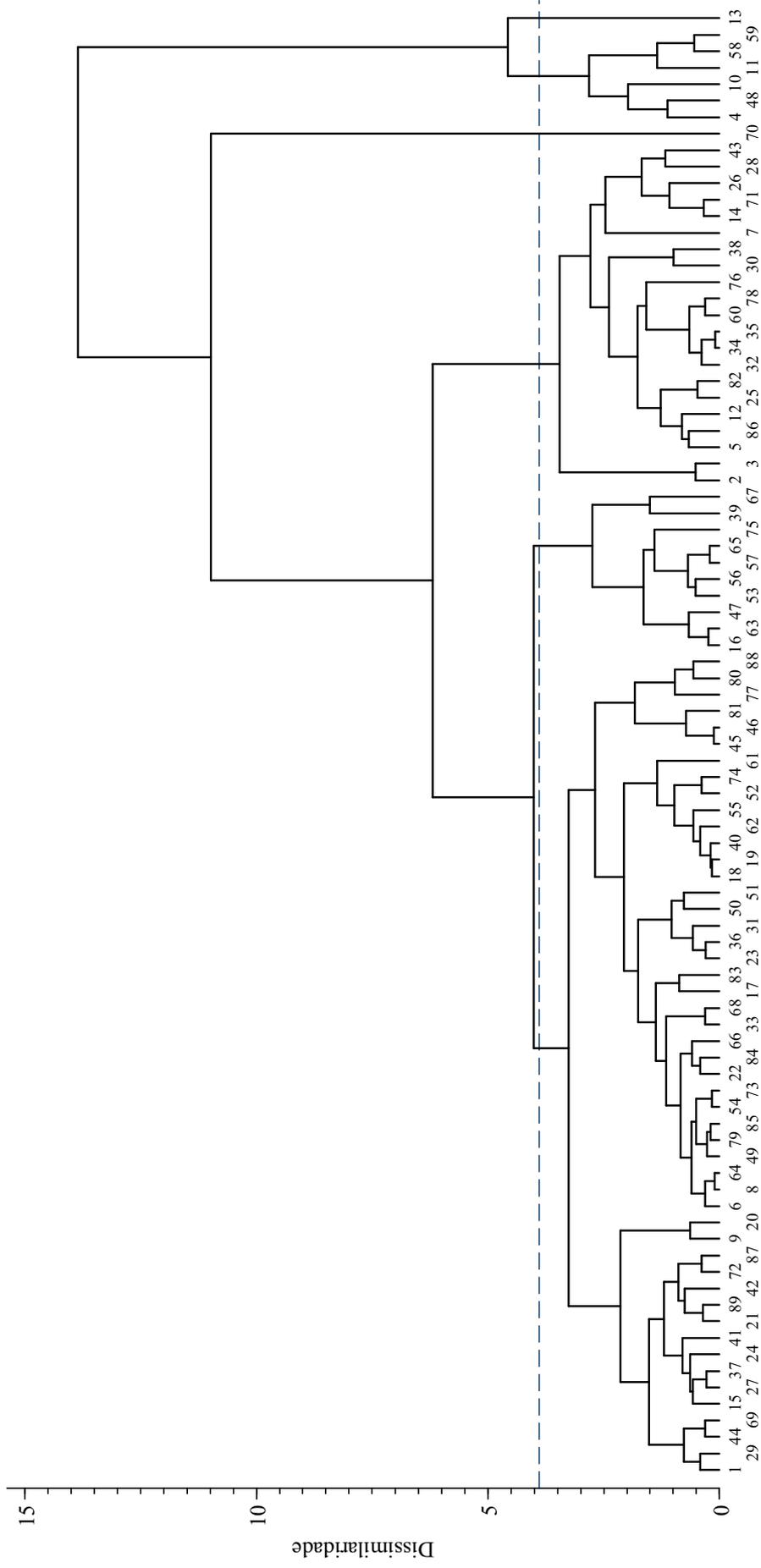


Figura 3. Dendrograma obtido com o método UPGMA com base na distância generalizada de Mahalanobis (D^2) para agrupamento de 89 acessos de alho pertencentes ao BGH/UFV. UFV, Viçosa (MG), 2012.

De acordo com o ponto de corte estabelecido pelo método de Mojena (1977) com $k = 1,25$, foram formados seis grupos de acessos, conforme a Tabela 15.

Tabela 15. Formação dos grupos de 89 acessos de alho do BGH/UFV segundo o método UPGMA com base na distância generalizada de Mahalanobis. UFV, Viçosa (MG), 2012.

Grupo	Acessos										
I	2	3	5	7	12	14	25	26	28	30	32
	34	35	38	43	60	71	76	78	82	86	
II	1	6	8	9	15	17	18	19	20	21	22
	23	24	27	29	31	33	36	37	40	41	42
	44	45	46	49	50	51	52	54	55	61	62
	64	66	68	69	72	73	74	77	79	80	81
	83	84	85	87	88	89					
III	16	39	47	53	56	57	63	65	67	75	
IV	4	10	11	48	58	59					
V	70										
VI	13										

À semelhança dos métodos de otimização de Tocher e Tocher modificado, o agrupamento com o método UPGMA reuniu no grupo I vinte e um acessos (Tabela 16). O grupo II obtido com este método foi o maior dentre os agrupamentos deste estudo, com cinquenta acessos. Os acessos 70 (BGH 6389) e 13 (BGH 4505) ficaram isolados cada um em um grupo (V e VI, respectivamente), como ocorreu quando se utilizou o método de Tocher (original). Bertan *et al.* (2006) constataram fato semelhante ao estudarem métodos de agrupamento para representar a distância morfológica entre acessos de trigo. Os autores verificaram que o método UPGMA e o método de Tocher (original) apresentaram resultados similares na representação dos genótipos mais divergentes.

Pela Tabela 16 verifica-se que o grupo VI foi o que apresentou as maiores médias para todas as variáveis, sendo constituído pelo acesso 13. O grupo V (acesso 70) também reuniu acessos com médias altas, com exceção do

NbB, o que pode ser desejável, uma vez que valores altos significam bulbilhos de tamanho fora do padrão comercial de qualidade.

Tabela 16. Número de acessos (N) e médias das variáveis para os grupos obtidos com o método UPGMA. UFV, Viçosa (MG), 2012.

Grupo	N	DIAM	COMP	PMB	NbB	PT
I	21	44,00	33,67	30,72	38,33	9,83
II	50	41,42	31,32	23,54	18,07	7,53
III	10	47,37	34,16	33,22	18,13	10,63
IV	6	50,51	39,49	46,70	55,04	14,95
V	1	52,38	38,94	45,05	15,13	14,42
VI	1	55,35	41,51	59,36	59,88	18,99

DIAM: diâmetro do bulbo (mm); COMP: comprimento do bulbo (mm); PMB: peso médio do bulbo (g); NbB: número de bulbilhos por bulbo; PT: produtividade ($t\ ha^{-1}$)

O coeficiente de correlação cofenética (r) do método UPGMA baseado na distância generalizada de Mahalanobis foi de 0,76, indicando boa representação das distâncias no dendrograma. Esse valor foi significativo pelo teste de Mantel ($P < 0,01$) baseado em 1000 reamostragens (Figura 4).

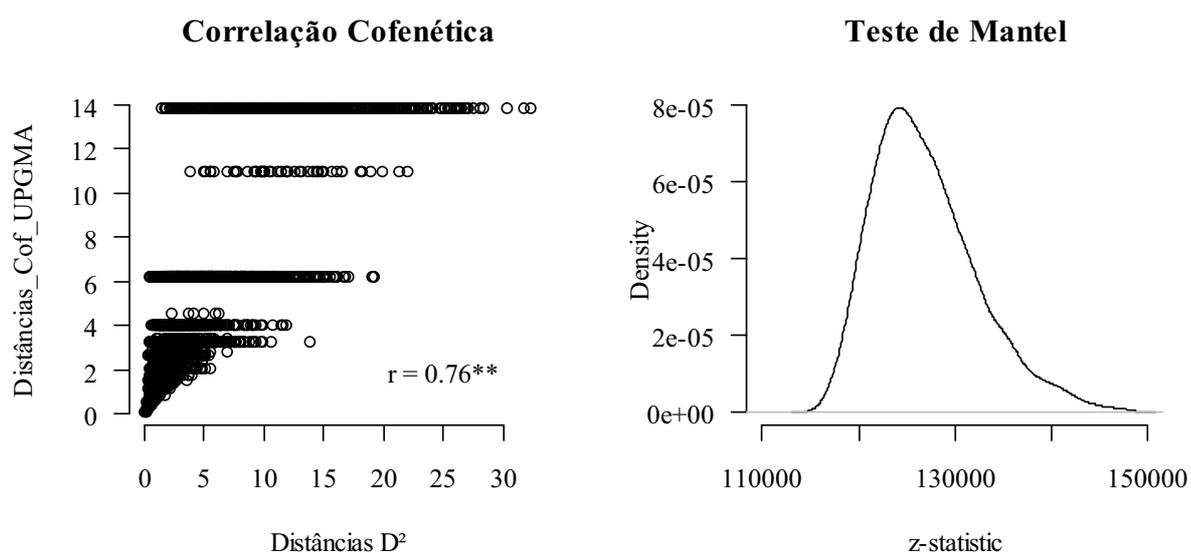


Figura 4. Correlação cofenética para as matrizes de dissimilaridade obtidas com o método UPGMA, baseado na distância generalizada de Mahalanobis (D^2); ** significativo pelo teste de Mantel a 1% de probabilidade, baseado em 1000 reamostragens. UFV, Viçosa (MG), 2012.

Na Tabela 17 é apresentado o resultado da análise discriminante de Fisher para os grupos obtidos com o agrupamento UPGMA. A análise identificou seis classificações erradas, resultando numa taxa de erro aparente igual a 6,74%. Embora o grupo II tenha apresentado um número bastante elevado de acessos (50 acessos), o percentual de classificações erradas (8%) foi ainda menor que no grupo I (9,5%), sendo estes dois grupos os únicos a apresentarem classificações erradas.

Tabela 17. Porcentagem de classificação correta (diagonal principal) e incorreta de cada grupo segundo o agrupamento de 89 acessos de alho pelo método UPGMA. UFV, Viçosa (MG), 2012.

Grupo	Classificação					
	I	II	III	IV	V	VI
I	90,5	9,5	0,0	0,0	0,0	0,0
II	4,0	92,0	4,0	0,0	0,0	0,0
III	0,0	0,0	100,0	0,0	0,0	0,0
IV	0,0	0,0	0,0	100,0	0,0	0,0
V	0,0	0,0	0,0	0,0	100,0	0,0
VI	0,0	0,0	0,0	0,0	0,0	100,0

Taxa de erro aparente: 6,74 % (6 classificações erradas)

Em comparação com os métodos de otimização, observa-se que o método UPGMA apresentou redução significativa da taxa de erro aparente, com diferença de 20,22% em relação ao método de Tocher e 12,36% em relação ao método de Tocher modificado.

4.6. Agrupamento com o Algoritmo de Ward

Pelo método de Mojena (1977) foi possível estabelecer um ponto de corte no dendrograma obtido com o algoritmo de Ward na distância 33,91, valor do critério θ_k do método de Mojena, que corresponde a 16,8% da distância máxima observada nos níveis de fusão (Figura 5).

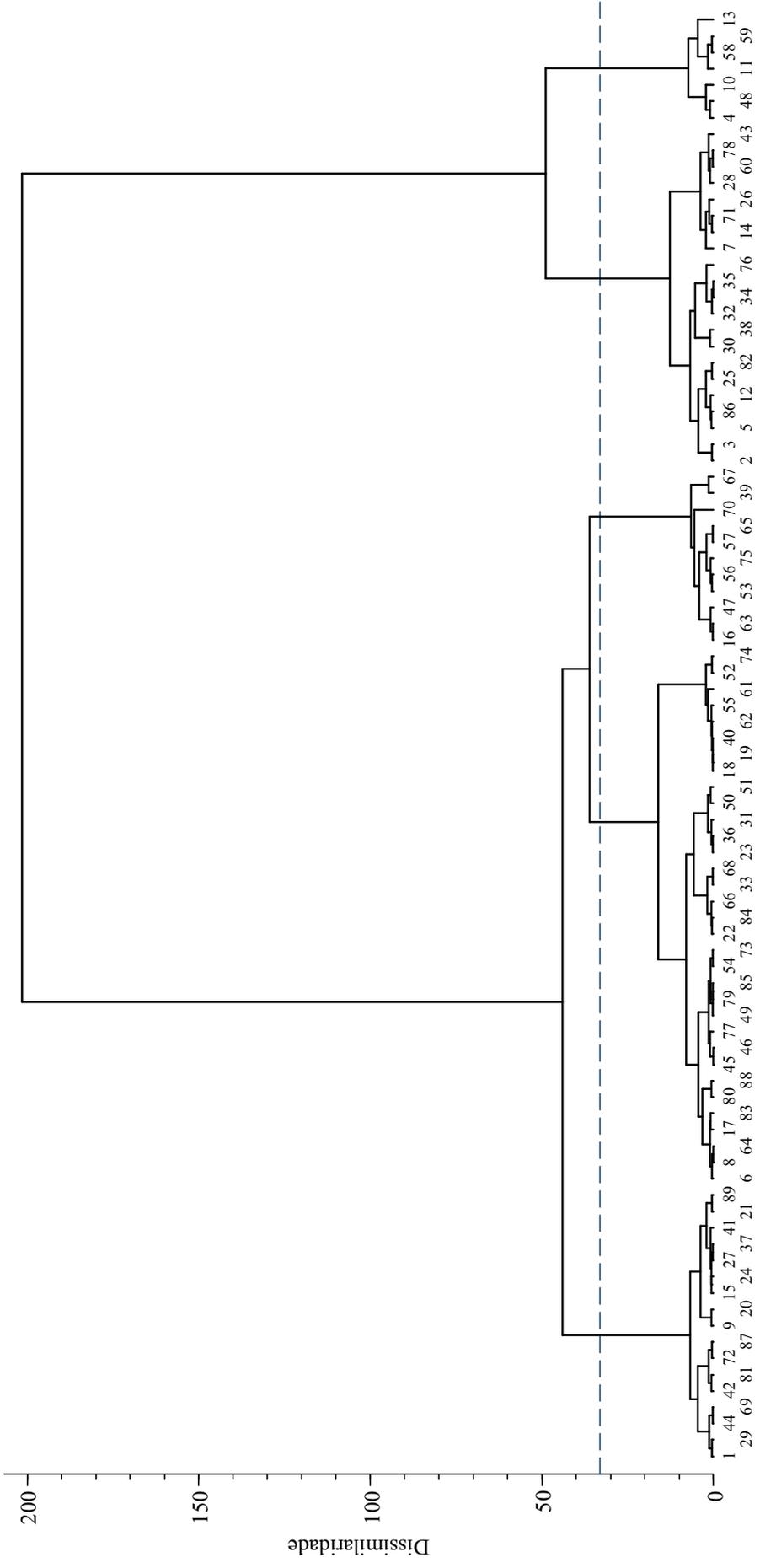


Figura 5. Dendrograma obtido com o algoritmo de Ward com base na distância generalizada de Mahalanobis (D^2) para agrupamento de 89 acessos de alho pertencentes ao BGH/UFV. UFV, Viçosa (MG), 2012.

Foram então estabelecidos cinco grupos de acessos, como pode ser visto na Tabela 18.

Tabela 18. Formação dos grupos de 89 acessos de alho do BGH/UFV segundo o algoritmo de Ward com base na distância generalizada de Mahalanobis. UFV, Viçosa (MG), 2012.

Grupo	Acessos										
I	2	3	5	7	12	14	25	26	28	30	32
	34	35	38	43	60	71	76	78	82	86	
II	6	8	17	18	19	22	23	31	33	36	40
	45	46	49	50	51	52	54	55	61	62	64
	66	68	73	74	77	79	80	83	84	85	88
III	16	39	47	53	56	57	63	65	67	70	75
IV	1	9	15	20	21	24	27	29	37	41	42
	44	69	72	81	87	89					
V	4	10	11	13	48	58	59				

Em relação às variáveis com maior importância para a dissimilaridade dos acessos (PMB e PT), pode-se caracterizar os grupos mais divergentes como: o grupo V é aquele que apresenta acessos com maiores médias (média de PMB do grupo: 48,51 g e média de PT do grupo: 15,52 t ha⁻¹); o grupo IV apresenta acessos com menores médias (média de PMB do grupo: 22,31 g e média de PT do grupo: 7,14 t ha⁻¹), conforme a Tabela 19.

Tabela 19. Número de acessos (N) e médias das variáveis para os grupos obtidos com o algoritmo de Ward. UFV, Viçosa (MG), 2012.

Grupo	N	DIAM	COMP	PMB	NbB	PT
I	21	44,00	33,67	30,72	38,33	9,83
II	33	41,72	32,49	24,17	15,55	7,73
III	11	47,82	34,59	34,30	17,85	10,98
IV	17	40,85	29,06	22,31	22,97	7,14
V	7	51,20	39,78	48,51	55,73	15,52

DIAM: diâmetro do bulbo (mm); COMP: comprimento do bulbo (mm); PMB: peso médio do bulbo (g); NbB: número de bulbilhos por bulbo; PT: produtividade (t ha⁻¹)

É interessante destacar que com o algoritmo de Ward, nenhum grupo foi constituído por apenas 1 acesso, como ocorreu com os métodos de otimização e com o método UPGMA. Isso ocorre devido ao fato de o método de Ward tender a formar grupos com o mesmo número de indivíduos, tendo como base os princípios da análise de variância (MINGOTI, 2005). Como exemplo disso, pode-se citar o resultado obtido por Rodrigues *et al.* (2002) ao utilizarem o método de Ward em estudo da divergência genética entre 37 cultivares de feijão, encontrando grupos com não menos que 5 acessos.

O coeficiente de correlação cofenético obtido com o algoritmo de Ward foi de 0,56, assim como ocorreu com o método UPGMA ($r = 0.76$), este valor foi significativo pelo teste de Mantel, baseado em 1000 reamostragens (Figura 6).

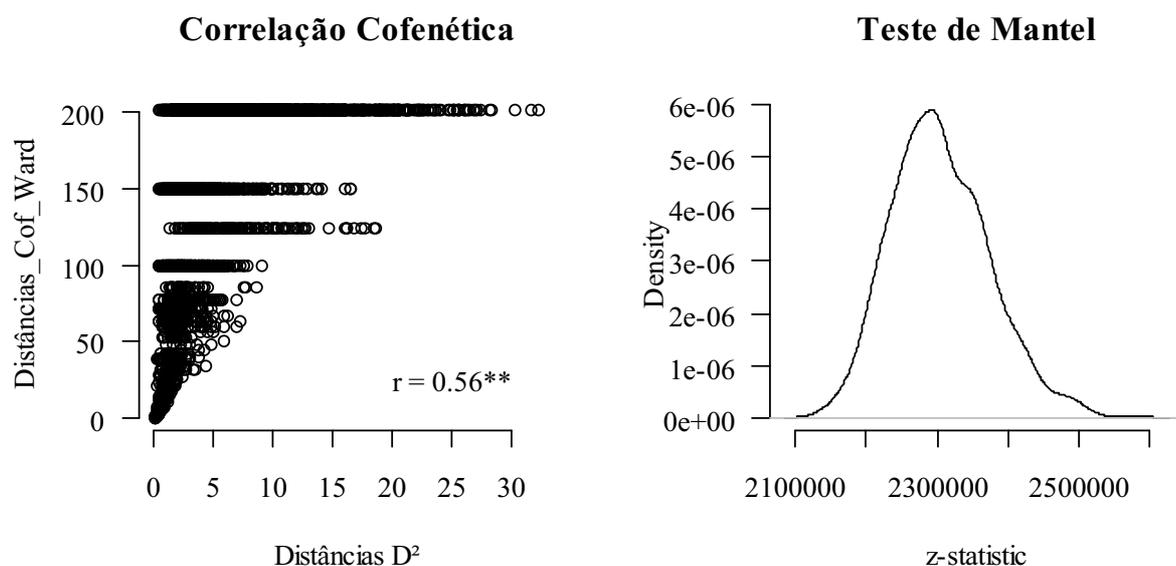


Figura 6. Correlação cofenética para as matrizes de dissimilaridade obtidas com o algoritmo de Ward, baseado na distância generalizada de Mahalanobis (D^2); ** significativo pelo teste de Mantel a 1% de probabilidade, baseado em 1000 reamostragens. UFV, Viçosa (MG), 2012.

Ainda pela Tabela 18, é possível perceber que todos os acessos do grupo II (33 acessos) constam no mesmo grupo obtido com o método UPGMA. Além disso, 28 dos 33 acessos foram coincidentes no grupo II obtido com os

métodos de otimização de Tocher e Tocher modificado, muito embora esses dois últimos métodos tenham apresentado alta taxa de erro aparente, principalmente devido a classificações erradas no grupo II, segundo os resultados das análises discriminantes.

Observa-se que o grupo III (11 acessos) foi praticamente idêntico ao grupo III obtido com o método UPGMA, com exceção do acesso 70. Nota-se também que todos os acessos do grupo IV (17 indivíduos) estão contidos no grupo II (50 indivíduos) do método UPGMA, ou seja, a união dos grupos II e IV aqui obtidos forma exatamente o grupo II do método UPGMA, evidenciando maior discriminação desses acessos pelo método de Ward.

Na Tabela 20 é apresentado o resultado da análise discriminante de Fisher para a formação obtida com o algoritmo de Ward. Observa-se que a taxa de erro aparente (4,49%) foi a menor dentre todas obtidas com os métodos de agrupamento neste estudo. Ademais, nota-se que no grupo II a análise discriminante detectou apenas 1 classificação errada (3%), sendo também a menor taxa de erro obtida para o grupo II, que em todos os métodos utilizados foi o que reuniu a maior parte dos acessos.

Tabela 20. Porcentagem de classificação correta (diagonal principal) e incorreta de cada grupo segundo o agrupamento de 89 acessos de alho pelo algoritmo de Ward. UFV, Viçosa (MG), 2012.

Grupo	Classificação				
	I	II	III	IV	V
I	90,5	4,8	0,0	4,8	0,0
II	0,0	97,0	3,0	0,0	0,0
III	0,0	0,0	90,9	9,1	0,0
IV	0,0	0,0	0,0	100,0	0,0
V	0,0	0,0	0,0	0,0	100,0

Taxa de erro aparente: 4,49 % (4 classificações erradas)

Kuiper e Fisher (1975) apud Cruz *et al.* (2011) concluíram que o método de Ward agrupa tão bem quanto a análise discriminante de Fisher quando

existe igual número de indivíduos nos grupos e quando os dados seguem distribuição normal multivariada.

Comparando o método de Ward com o método não hierárquico k-médias para agrupamento de produtores de leite, Seidel *et al.* (2008) verificaram alta coincidência nos agrupamentos. Os autores destacam ainda que o método de Ward permitiu definir com mais eficiência a quantidade de agrupamentos que devem ser utilizados, enquanto que o método k-médias classificou de forma mais adequada os produtores dentro dos agrupamentos.

4.7. Resumo da Comparação dos Métodos

Na Tabela 21 apresenta-se um resumo dos resultados obtidos (número de grupos) com os métodos de agrupamento e a avaliação desses resultados por meio da correlação cofenética e da análise discriminante de Fisher.

Tabela 21. Resumo dos resultados obtidos com os métodos de agrupamento aplicados a dados de 89 acessos de alho. UFV, Viçosa (MG), 2012.

Método de agrupamento	Número de grupos	Correlação cofenética	TEA (%)
Tocher	15	--	26,96
Tocher modificado	7	--	19,10
UPGMA	6	0,76**	6,74
Ward	5	0,56**	4,49

TEA: taxa de erro aparente da análise discriminante; ** significativo pelo teste de Mantel (P<0,01) baseado em 1000 reamostragens.

O método de Tocher apresentou um número de grupos (quinze) muito maior que os demais métodos. Sete desses grupos foram constituídos por um acesso apenas. Além disso, a aplicação da análise discriminante aos grupos formados indicou uma taxa de erro aparente de 26,96%, sendo considerada alta.

Apesar de o método sequencial de Tocher formar grupos mais concordantes com os obtidos com os métodos hierárquicos que o método original de Tocher, pela análise discriminante observou-se muitos erros de classificação

(TEA = 19,10%). Ademais, destaca-se o fato de o método ter se mostrado eficiente no que diz respeito à identificação dos acessos mais dissimilares.

Quanto aos métodos hierárquicos, observou-se resultados concordantes, tanto em relação aos grupos como em relação a taxa de erro aparente das análises discriminantes. Os coeficientes de correlação cofenética foram significativos ($P < 0,01$) pelo teste de Mantel baseado em 1000 reamostragens. Não obstante a correlação obtida com o método UPGMA foi numericamente superior a do algoritmo de Ward. Isso pode ser explicado pelo fato de a medida utilizada no método de Ward depender do quadrado da distância euclidiana, os valores da matriz cofenética são numericamente maiores que os observados no método UPGMA. Isso provavelmente explica o fato de o coeficiente de correlação cofenética obtido com o algoritmo de Ward (0,56) ser, numericamente, inferior ao obtido com o método UPGMA (0,76), embora Rohlf (1982) alerte para o fato de que uma correlação cofenética próxima de 0,9 não garante que o dendrograma tenha sido capaz de sintetizar a relação fenética dos indivíduos.

5. CONCLUSÕES

Os métodos de Tocher modificado, UPGMA e algoritmo de Ward apresentaram resultados concordantes entre si quanto a formação dos grupos. No entanto, pela análise discriminante de Fisher aplicada aos grupos dos métodos hierárquicos (UPGMA e Ward) observou-se as menores taxas de erro aparente, sendo, portanto, os métodos mais consistentes para este estudo de divergência genética de acessos de alho.

Os acessos que apresentaram maior dissimilaridade foram os acessos 13 (BGH 4505) e 61 (BGH 5958), principalmente em relação ao peso médio do bulbo e produtividade.

6. REFERÊNCIAS BIBLIOGRÁFICAS

ALBUQUERQUE, M. A. **Estabilidade em análise de agrupamento (cluster analysis)**. Dissertação (Mestrado em Biometria) – Universidade Federal Rural de Pernambuco. Orientador: Rinaldo Luiz Caraciolo Ferreira, Recife, 2005. 62f.

BERTAN, I.; CARVALHO, F. I. F.; OLIVEIRA, A. C.; VIEIRA, E. A.; HARTWING, I.; SILVA, J. A. G.; SHIMIDT, D. A. M.; VALÉRIO, I. P.; BUSATO, C. C.; RIBEIRO, G. Comparação de métodos de agrupamento na representação da distância morfológica entre genótipos de trigo. **Revista Brasileira de Agrociência**, v.12, n.3, p 279-286, 2006.

BUSSAB, W. O.; MIAZAKI, E. S.; ANDRADE, D. **Introdução à análise de agrupamentos**. São Paulo: Associação Brasileira de Estatística, 1990. 105p.

CARVALHO, L. P. ;LANZA, M. A.; FALIERI, J.; SANTOS, J. W. Análise da diversidade genética entre acessos do banco ativo de germoplasma de algodão. **Pesquisa Agropecuária Brasileira**, v.38, n.10, p.1149-1155, 2003.

CRUZ, C. D. **Programa Genes: Biometria**. Viçosa: Editora UFV. 382p. 2006

CRUZ, C. D.; CARNEIRO, P. C. S. **Modelos biométricos aplicados ao melhoramento genético**.volume 2. Viçosa: UFV, 2003. 623p.

CRUZ, C. D.; FERREIRA, F. M.; PESSONI, L. A. **Biometria aplicada ao estudo da diversidade genética**. Visconde do Rio Branco: Suprema, 2011. 620p.

CRUZ, C. D.; REGAZZI, A. J. **Modelos biométricos aplicados ao melhoramento genético**. 2 ed. Viçosa: UFV, 1994. 390p.

DINIZ FILHO, J. A. **Métodos filogenéticos comparativos**. Ribeirão Preto: Holos, 2000. 120p.

FERREIRA, D. F. **Estatística Multivariada**. 1 ed. Lavras: Editora UFLA, 2008. 662p.

FISHER, R. A. The Use of Multiple Measurement in Taxonomic Problems. **Annals of Eugenics**, v.7, p.179–188, 1936.

FONSECA, A. F. A. *et al.* Divergência genética em café conilon. **Pesquisa Agropecuária Brasileira**, v.41, n.4, p.599-605, 2006.

IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Levantamento Sistemático da Produção Agrícola. Rio de Janeiro, v.24, n.10, p.1-80, out. 2011. Disponível em: http://www.ibge.gov.br/home/estatistica/indicadores/agropecuaria/lspa/lspa_2011_10.pdf. Acesso em: 18, dez. 2011.

JOHNSON, R. A.; WICHERN, D. W. **Applied Multivariate Statistical Analysis**. New Jersey, USA: Englewood Cliffs, 642p. 1992.

KHATTREE, R.; NAIK, D. N. **Multivariate data reduction and discrimination with SAS Software**. Cary, NC: SAS Institute Inc., 2000. 338p.

KUIPER, F. K.; FISHER, L. A Monte Carlo comparison of six clustering procedures. **Biometrics**, v.31, p.777-783, 1975.

LANCE, G. N.; WILLIAMS, W. T. A General Theory of Classificatory Sorting Strategies. 1. Hierarchical Systems, **Computer Journal**, v.9, p.373-380, 1967.

MAHALANOBIS, P. C. On the generalized distance in statistics. **Proceedings of The National Institute of Sciences of India**, v.12, p.49-55, 1936.

MANLY, B. F. J. **Métodos estatísticos multivariados**: uma introdução (tradução Sara Ianda Carmona). 3 ed. Porto Alegre: Bookman, 2008. 229p.

MANTEL, N. The detection of disease clustering and generalized regression approach. **Cancer Res.**, v.27, p.209-220, 1967.

MARCHIORO, V. S. *et al.* Dissimilaridade genética entre genótipos de aveia. **Ciência e Agrotecnologia**, v.27, p.285-294, 2003.

MILLIGAN, G. W. Na examination of the effect of six types of error perturbation on fifteen clustering algorithms. **Psychometrika**, v.45, p.325-342, 1980.

MILLIGAN, G. W.; COOPER, M. C. An examination of procedures for determining the number of cluster in a data set. **Psychometrika**, v.50, p.159-179, 1985.

MINGOTI, S. A. **Análise de dados através de métodos de estatística multivariada**: uma abordagem aplicada. Belo Horizonte: Editora UFMG, 2005. 297p.

MIRANDA, G. V. *et al.* Potencial de melhoramento e divergência genética de cultivares de milho-pipoca. **Pesq. agropec. bras.**, v.38, n.6, p.681-688, 2003.

MOHAMMADI, S. A.; PRASANNA, B. M. Analysis of genetic diversity in crop plants – Salient statistical tools and considerations. **Crop Science**, v.43, p.1235-1248, 2003.

MOJENA, R. Hierárquical grouping method and stopping rules: an evaluation. **Computer Journal**, v.20, p.359-363, 1977.

MOLL, R. H.; LONNQUIST, J. H.; FORTUNO, J. V.; JOHNSON, E. C. The relationship of heterosis and genetic divergence in maize. **Genetics**, v.52, 139-144, 1965.

MUELLER, S.; BIASI, J.; MENEZES SOBRINHO, J. A.; MULLER, J. J. V. Comportamento de cultivares de alho, plantio de junho. **Pesquisa Agropecuária Brasileira**, v.25, n.11, p.1561-1567. 1990.

NASCIMENTO FILHO, F. J. *et al.* Divergência genética entre clones de guaranazeiro. **Pesq. agropec. bras.**, v.36, n.3, p.501-506, 2001.

OLIVEIRA, A. C. B.; SEDIYAMA, M. A. N.; PEDROSA, M. W.; GARCIA, N. C. P.; GARCIA, S. L. R. Divergência genética e descarte de variáveis em alface cultivada sob sistema hidropônico. **Acta Scientiarum. Agronomy**, v.26, p.211-217, 2004.

PATERNIANI, E.; LONNQUIST, J. H. Heterosis in interracial crosses of corn (*Zea mays* L.). **Crop Science**, v.3, p.504-507, 1963.

PUIATTI, M.; FERREIRA, F. A. Cultura do alho. In: FONTES, P. C. R. (eds.). **Olericultura: teoria e prática**. 1 ed. Viçosa: UFV, 2005. p.299-322.

R DEVELOPMENT CORE TEAM (2011). **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.

RAO, C. R.; **An advanced statistical method in biometric research**. New York, Ed. John Wiley e Sons, 1952. 390p.

RAO, N. K. S.; SWAMY, R. D.; CHACO, E. K. Differentiation of plantlets in hybrid embryo callus of pineapple. **Scientia Horticulturae**, Amsterdam, v.15, p.235-238, 1981.

RESENDE, M. D. V.; DUARTE, J. B. Precisão e controle de qualidade em experimentos de avaliação de cultivares. **Pesquisa Agropecuária Tropical**, v.37, n.3, p.182-194, 2007.

RODRIGUES, L. S.; ANTUNES, I. F.; TEIXEIRA, M. G.; SILVA, J. B. Divergência genética entre cultivares locais e cultivares melhoradas de feijão. **Pesquisa Agropecuária Brasileira**, v.37, n.9, p.1275-1284, 2002.

ROHLF, F. J. Consensus índices for comparing classifications. **Mathematical Bioscience**, v.59, p.131-144, 1982.

ROSEMBURG, H. C. **Cluster analysis for researchers**. California, EUA: Lifetime Learning, 1984. 334p.

SANTANA, C. M.; MALINOVSKI, J. R. Uso da análise multivariada no estudo de fatores humanos em operadores de motosserra, **Cerne**, v.8, n.2, p.101-107, 2002.

SEIDEL, E. J.; MOREIRA JÚNIOR, F. J.; ANSUI, A. P.; NOAL, M. R. C. Comparação entre o método Ward e o método K-médias no agrupamento de produtores de leite. **Ciência e Natura**, UFSM, v.30, n.1, 7- 15, 2008.

SIEGMUND, K. D.; LAIRD, P. W.; LAIRD OFFRINGA, I. A. A comparison of cluster analysis methods using DNA methylation data. **Bioinformatics**, v.20, n.12, p.1896-1904, 2004.

SILVA, C. M. *et al.* Genetic diversity among sugarcane clones (*Saccharum* spp.). **Acta Scientiarum. Agronomy**, v.27, p.315-319, 2005.

SILVA, E. C.; SOUZA, R. J.; SANTOS, W. S. Efeitos do tempo de frigidificação em cultivares de alho (*Allium sativum* L.) provenientes de cultura de meristemas. **Ciência e Agrotecnologia**, v.24, n.4, p.939-946, 2000.

SINGH, D. The relative importance of characters affecting genetic divergence. **Indian Journal Genetics & Plant Breeding**, v.41, p.237-245, 1981.

SOKAL, R. R.; ROHLF, F. J. The comparison of dendrograms by objective methods. **Taxon**, v.11, p.30-40, 1962.

SOUZA, F. F.; QUEIRÓZ, M. A.; DIAS, R. S. C. Divergência genética em linhagens de melancia. **Horticultura Brasileira**, v.23, p179-183, 2005.

SOUZA, R. J.; MACÊDO, F. S.; YURI, J. E. Alho (*Allium sativum* L.). In: PAULA JÚNIOR, T.J.; VENZON, M. (Eds.). **101 Culturas: Manual de tecnologias agrícolas**. Belo Horizonte: EPAMIG, 2007. cap. 8, p.75-84.

STAHLSCHMIDT, O.; CAVAGNARO, J. B.; BORGIO, R. Influencia del área foliar y la partición de asimilados en tres cultivares de ajo. **Actas XX Reunión Arg. Fisiol. Vegetal**. p.194-195, 1993.

STEEL, R. G. D.; TORRIE, J. H. **Principles and procedures of statistics**. 2 ed. New York: McGraw-Hill, 1980. 633p.

TRANI, P. E.; PASSOS, F. A.; FOLTRAN, D. E.; TIVELLI, S.W.; RIBEIRO, L. J. A. Avaliação dos acessos de alho do Instituto Agrônômico de Campinas. **Horticultura Brasileira**, v.24, p.156-160, 2005.

TRANI, P. E.; TAVARES, M.; SIQUEIRA, W. J.; SANTOS, R. R. Cultura do alho: recomendações para o seu cultivo no estado de São Paulo. Campinas: Instituto Agrônômico, 1997. 39p. (Boletim Técnico, 170).

VASCONCELOS, E. S.; CRUZ, C. D.; BHERING, L. L.; RESENDE JÚNIOR, M. F. R. Método alternativo para análise de agrupamento. **Pesq. agropec. bras.**, v.42, n.10, p.1421-1428, 2007.

VIEIRA, E. A. *et al.* Divergência genética entre acessos açucarados e não açucarados de mandioca. **Pesq. agropec. bras.**, v.43, n.12, p.1707-1715, 2008.

WARD, J. Hierarchical grouping to optimize an objective function. **Journal of American Statistical Association**, v.58, p.236-244, 1963.

APÊNDICE

A - Análise Discriminante

Tabela 22A. Coeficientes das funções discriminantes lineares estimados de acordo com o agrupamento de 89 acessos de alho obtido com cada método. UFV, Viçosa (MG), 2012.

Variável	LD1	LD2	LD3	LD4	LD5
<i>Tocher</i>					
DIAM	-0,39	0,11	-0,86	0,29	-0,40
COMP	0,12	-0,06	0,21	0,79	0,32
PMB	-27,68	4,17	183,95	58,64	-235,68
NbB	0,21	-0,10	-0,06	0,06	-0,05
PT	86,76	-12,23	-573,64	-184,75	736,85
Proporção	0,63	0,19	0,09	0,05	0,03
<i>Tocher modificado</i>					
DIAM	-0,26	-0,24	0,72	-0,42	-0,17
COMP	0,07	0,03	0,41	0,68	0,04
PMB	-21,25	-64,37	-55,59	27,09	-241,40
NbB	0,18	0,08	0,08	-0,01	-0,04
PT	66,57	200,65	171,94	-84,78	754,72
Proporção	0,64	0,28	0,06	0,02	0,00
<i>UPGMA</i>					
DIAM	0,45	0,01	-0,47	-0,60	-0,03
COMP	-0,30	-0,51	0,33	-0,23	0,53
PMB	46,13	-7,38	-185,07	119,60	133,42
NbB	-0,15	-0,15	-0,06	-0,04	0,00
PT	-144,63	24,57	578,81	-372,55	-417,42
Proporção	0,70	0,28	0,01	0,01	0,00
<i>Ward</i>					
DIAM	0,37	0,15	-0,41	0,67	
COMP	-0,14	-0,15	1,00	0,16	
PMB	58,22	-22,30	44,57	98,01	
NbB	-0,18	0,11	0,03	0,07	
PT	-182,31	68,81	-140,00	-307,59	
Proporção	0,69	0,18	0,13	0,01	

DIAM: diâmetro do bulbo (mm); COMP: comprimento do bulbo (mm); PMB: peso médio do bulbo (g); NbB: número de bulbilhos por bulbo; PT: produtividade ($t\ ha^{-1}$).

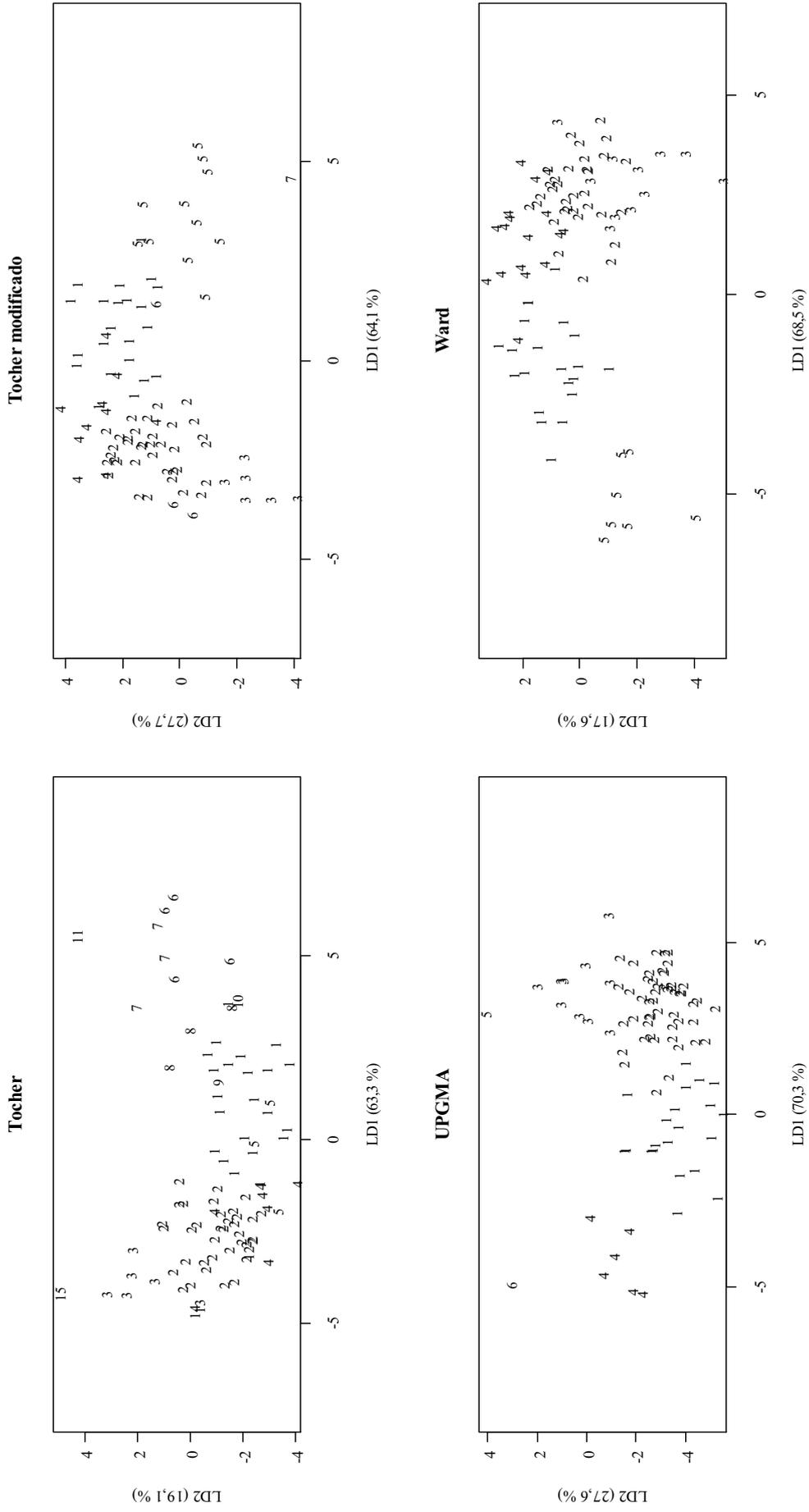


Figura 7A. Escores das duas primeiras funções discriminantes (LD1 e LD2) estimadas de acordo com o agrupamento de 89 acessos de alho obtido com cada método. UFV, Viçosa (MG), 2012.

B - Script das Análises no *Software R*

```

# *****
# 1. Leitura dos dados
#
# *****

X=read.table("D:\\UFV\\Dissertação\\Análises\\DADOS.dat",header=F)

# *****
# 2. ANOVA em blocos casualizados
#
# *****

Trat=as.factor(X[,1])
Blocos=as.factor(X[,2])
Modelo1=aov(X[,3]~Blocos+Trat) #DIAM
summary(Modelo1)
Modelo2=aov(X[,4]~Blocos+Trat) #COMP
summary(Modelo2)
Modelo3=aov(X[,5]~Blocos+Trat) #PMB
summary(Modelo3)
Modelo4=aov(X[,6]~Blocos+Trat) #NbB
summary(Modelo4)
Modelo5=aov(X[,7]~Blocos+Trat) #PT
summary(Modelo5)

# *****
# 3. Distância generalizada de Mahalanobis
# *****

mah=read.table("D:\\UFV\\Dissertação\\Análises\\mahalanobis.dat",header=F)
D2=as.matrix(mah)
matrizD2=as.dist(m = D2, diag = FALSE, upper = FALSE)

# *****
# 4. Agrupamentos hierárquicos
# *****

UPGMA=hclust(matrizD2,method="average")
ward=hclust(matrizD2,method="ward")

# *****
# 5. Pontos de corte dos dendrogramas (Método de Mojena, 1977)
# *****

k=1.25 # Milligan e Cooper (1985)

PCUPGMA=mean(UPGMA$height)+k*sd(UPGMA$height)
PCUPGMA

PCward=mean(ward$height)+k*sd(ward$height)
PCward

par(family="serif",las=1)
plot(UPGMA, hang = -1,cex=0.8,ylab="Dissimilaridade")
abline(h=PCUPGMA,v=NULL,col=4,lty=2)

```

```

G=rect.hclust(UPGMA,h=PCUPGMA,which=c(1:6),border=1:6) #Identificar os
grupos
G

plot(ward, hang = -1,cex=0.8,ylab="Dissimilaridade")
  abline(h=PCward,v=NULL,col=4,lty=2)
  G=rect.hclust(ward,h=PCward,which=c(1:5),border=1:5) #Identificar os grupos
  G

# *****
# 6. Coeficiente de Correlação Cofenética (r)
# *****

cofUPGMA=cophenetic(UPGMA)
cor(matrizD2,cofUPGMA)

cofward=cophenetic(ward)
cor(matrizD2,cofward)

cof_UPGMA=read.table("D:/UFV/Dissertação/Análises/cof_UPGMA.txt",head=T)
cof_ward=read.table("D:/UFV/Dissertação/Análises/cof_ward.txt",head=T)

par(family="serif",las=1,mfrow=c(1,2))
plot(cof_UPGMA$D2,cof_UPGMA$cof_UPGMA,xlab="Matriz
  D²",ylab="Matriz_Cof_UPGMA")
  text(25,2,"r = 0.76**")

plot(cof_ward$D2,cof_ward$cof_ward,xlab="Matriz D²",ylab="Matriz_Cof_Ward")
  text(25,25,"r = 0.56**")

# *****
# 7. Teste de Mantel
# *****

require(ape)

mantel.test(D2,as.matrix(cophenetic(UPGMA)),
  nperm=1000,graph=T,xlab="z-statistic",ylab="Density",
  main="Mantel Test (UPGMA)")
mantel.test(D2,as.matrix(cophenetic(ward)),
  nperm=1000,graph=T,xlab="z-statistic",ylab="Density",
  main="Mantel Test (Ward)")

# *****
# 8. Análise Discriminante de Fisher
# *****

grupos=read.table("D:\\UFV\\Dissertação\\Análises\\grupos_discrim.txt",heade
  r=T)
  grupos

require(MASS)

ADtocher=lda(grupos$Tocher ~ grupos$DIAM + grupos$COMP + grupos$PMB +
  grupos$NbB + grupos$PT)
par(family="serif")
plot(ADtocher,dimen=2)
classesTocher=predict(ADtocher)

```

```

cbind(grupos$Acesso,grupos$Tocher,classesTocher$class)
cont=0
for(i in 1:length(grupos$Tocher))
{
if (grupos$Tocher[i]==classesTocher$class[i]){cont=cont+1}else{cont=cont}
}
(TEA=1-cont/89)

ADtocherM=lda(grupos$Tocher_Modif ~ grupos$DIAM + grupos$COMP + grupos$PMB +
grupos$NbB + grupos$PT)
plot(ADtocherM,dimen=2)
classesTocherM=predict(ADtocherM)
cbind(grupos$Acesso,grupos$Tocher_Modif,classesTocherM$class)
cont=0
for(i in 1:length(grupos$Tocher_Modif))
{
if
(grupos$Tocher_Modif[i]==classesTocherM$class[i]){cont=cont+1}else{cont=cont}
}
}
(TEA=1-cont/89)

ADupgma=lda(grupos$UPGMA ~ grupos$DIAM + grupos$COMP + grupos$PMB + grupos$NbB
+ grupos$PT)
plot(ADupgma,dimen=2)
classesUPGMA=predict(ADupgma)
cont=0
for(i in 1:length(grupos$UPGMA))
{
if (grupos$UPGMA[i]==classesUPGMA$class[i]){cont=cont+1}else{cont=cont}
}
}
(TEA=1-cont/89)

ADward=lda(grupos$Ward ~ grupos$DIAM + grupos$COMP + grupos$PMB + grupos$NbB +
grupos$PT)
plot(ADward,dimen=2)
classesWard=predict(ADward)
cbind(grupos$Acesso,grupos$Ward,classesWard$class)
cont=0
for(i in 1:length(grupos$Ward))
{
if (grupos$Ward[i]==classesWard$class[i]){cont=cont+1}else{cont=cont}
}
}
(TEA=1-cont/89)

#####

```