

AUTARQUIA ASSOCIADA À UNIVERSIDADE DE SÃO PAULO

Estudo comparativo dos algoritmos hierárquicos de análise de agrupamentos em resultados experimentais

PRISCILLA RAMOS CARVALHO

Dissertação apresentada como parte dos requisitos para obtenção do Grau de Mestre em Ciências na Área de Tecnologia Nuclear - Aplicações

Orientador:

Prof. Dr. Casimiro Sepúlveda Munita

São Paulo 2018

INSTITUTO DE PESQUISAS ENERGÉTICAS E NUCLEARES Autarquia associada à Universidade de São Paulo

Estudo comparativo dos algoritmos hierárquicos de análise de agrupamentos em resultados experimentais

PRISCILLA RAMOS CARVALHO

Dissertação apresentada como parte dos requisitos para obtenção do Grau de Mestre em Ciências na Área de Tecnologia Nuclear - Aplicações

Orientador:

Prof. Dr. Casimiro Sepúlveda Munita

Versão Corrigida Versão Original disponível no IPEN

> São Paulo 2018

Fonte de Financiamento: CAPES

Autorizo a reprodução e divulgação total ou parcial deste trabalho, para fins de estudo e pesquisa, desde que citada a fonte

Como citar:

CARVALHO, P. R. *Estudo comparativo dos algoritmos hierárquicos de análise de agrupamentos em resultados experimentais*. 2018. 140 p. Dissertação (Mestrado em Tecnologia Nuclear), Instituto de Pesquisas Energéticas e Nucleares, IPEN CNEN/SP, São Paulo. Disponível em: <www.teses.usp.br> (data de consulta no formato: dd/mm/aaaa)

Ficha catalográfica elaborada pelo Sistema de geração automática da Biblioteca IPEN/USP, com os dados fornecidos pelo(a) autor(a)

Carvalho, Priscilla Ramos
Estudo comparativo dos algoritmos hierárquicos de análise
de agrupamentos em resultados experimentais / Priscilla Ramos
Carvalho; orientador Casimiro Sepúlveda Munita. -- São Paulo,
2018.
140 p.

Dissertação (Mestrado) - Programa de Pós-Graduação em Tecnologia Nuclear (Aplicações) -- Instituto de Pesquisas Energéticas e Nucleares, São Paulo, 2018.

1. Análise de Agrupamentos. 2. Métodos Hierárquicos. 3. Coeficiente de Correlação Cofenética. 4. Arqueometria. 5. INAA. I. Munita, Casimiro Sepúlveda, orient. II. Título.



Universidade de São Paulo



ATA DE DEFESA

Aluno: 85131 - 6910097 - 2 / Página 1 de 1

Resultado

Aprovado

Aprovado

Ata de defesa pública de Dissertação do(a) Senhor(a) Príscilla Ramos Carvalho no Programa: Tecnologia Nuclear, do(a) Instituto de Pesquisas Energéticas e Nucleares da Universidade de São Paulo.

Aos 11 dias do mês de setembro de 2018, no(a) Auditório Dr. Alcidio Abrão (1 realizou-se a Defesa da Dissertação do(a) Senhor(a) Priscilla Ramos Carvalho, apresentada para a obtenção do título de Mestra intitulada:

"Estudo comparativo dos algoritmos hierárquicos de análise de agrupamentos em resultados experimentais"

Função

Titular

Sigla da CPG

Após declarada aberta a sessão, o(a) Sr(a) Presidente passa a palavra ao candidato para exposição e a seguir aos examinadores para as devidas arguições que se desenvolvem nos termos regimentais. Em seguida, a Comissão Julgadora proclama o resultado:

Casimiro Jaime Alfredo Sepúlveda Munita Presidente IPEN(IPEN)

Nome dos Participantes da Banca

Lucia Pereira Barroso

Andreza Portella Ribeiro	Titular	UNINOVE - Externo	Aprovado
Resultado Final: Aprovado			
Eu, Paulo Albino Balan Junior Senhores(as) examinadores. São Paulo, aos 11 dias		e ata , que assino junta embro de 2018.	mente com os(as)
Lucia Pereira Barroso		Andreza Portell	. ← a-Ribeiro

Casimiro Jaime Alfredo Sepúlveda Munita Presidente da Comissão Julgadora

Presidente da Comissão de Pós-Graduação

À minha família, em especial aos meus pais Jorge e Elizabeth, por sempre me incentivarem, apoiarem e darem força para seguir em busca dos meus ideais.

AGRADECIMENTOS

Agradeço aos meus pais, Jorge e Elizabeth, e a minha irmã, Patricia, pelo incentivo em buscar e não desistir dos meus sonhos, e por me apoiarem e auxiliarem durante todos esses anos. Vocês me deram forças para persistir e tiveram paciência ao longo desta caminhada.

Ao meu orientador Prof. Dr. Casimiro Sepúlveda Munita pela dedicação, disponibilidade, confiança e orientação na execução deste trabalho.

Agradeço aos Professores Doutores André Luis Lapolli e Mário Olímpio de Menezes, pela ajuda com o programa estatístico R.

Ao meu namorado Eduardo pelo permanente apoio e confiança, ajudando a superar os obstáculos e dificuldades com toda a paciência e disponibilidade.

Aos amigos do CRPq pela ajuda, amizade e alegrias vivenciadas durante todo este período.

Agradeço à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES), Código de Financiamento 001, pela concessão da bolsa de estudo e o suporte financeiro para a execução deste trabalho.

"Nada na vida deve ser temido, somente compreendido. Agora é a hora de compreender mais, para que possamos temer menos."

RESUMO

CARVALHO, Priscilla R. *Estudo comparativo dos algoritmos hierárquicos de análise de agrupamentos em resultados experimentais.* 2018. 140 p. Dissertação (Mestrado em Tecnologia Nuclear) – Instituto de Pesquisas Energéticas e Nucleares – IPEN – CNEN/SP. São Paulo.

Objetivou-se, com este trabalho, estudar os métodos hierárquicos de análise de agrupamentos (ligação simples, ligação completa, ligação média, centróide e de Ward com base nas distâncias Euclidiana, Euclidiana ao quadrado, Manhattan e Mahalanobis), de modo a identificar qual é o mais adequado para uma base de dados arqueológicos. Utilizou-se uma base de dados fornecida pelo Grupo de Estudos Arqueométricos do IPEN - CNEN/SP, na qual foram analisadas 146 amostras de fragmentos cerâmicos de três sítios arqueológicos por análise por ativação com nêutrons instrumental, sendo determinadas as frações de massa de 24 elementos químicos: As, Ba, Ce, Co, Cr, Cs, Eu, Fe, Hf, K, La, Lu, Na, Nd, Rb, Sb, Sc, Sm, Ta, Tb, Th, U, Yb e Zn. Para a determinação do melhor método, foram avaliados os dendrogramas conjuntamente com o valor dos coeficientes de correlação cofenética (CCC), obtidos para cada método. O método da ligação média mostrou-se mais coerente na formação dos agrupamentos, apresentando também os maiores valores do CCC. Por último, um script com funções do programa estatístico R foi desenvolvido para calcular o CCC, com o intuito de auxiliar os pesquisadores a encontrar o método de agrupamento mais apropriado para sua base de dados.

Palavras-chave: Análise de Agrupamentos; Métodos Hierárquicos; Coeficiente de Correlação Cofenética; Arqueometria; INAA.

ABSTRACT

CARVALHO, Priscilla R. *Comparative study of hierarchical algorithms of cluster analysis in experimental results*. 2018. 140 p. Dissertação (Mestrado em Tecnologia Nuclear) – Instituto de Pesquisas Energéticas e Nucleares – IPEN – CNEN/SP. São Paulo.

The objective of this work was to study the hierarchical methods of cluster analysis (single linkage, complete linkage, average linkage, centroid and Ward with base in the Euclidean distance, Squared Euclidean distance, Manhattan and Mahalanobis) to identify which is the most appropriate in archaeological data set. It was used a data set provided by the Archaeometry Studies Group from IPEN – CNEN/SP in which 146 samples of ceramic fragments from three archaeological sites analyzed by instrumental neutron activation analysis, being determined the mass fractions of 24 chemical elements: As, Ba, Ce, Co, Cr, Cs, Eu, Fe, Hf, K, La, Lu, Na, Nd, Rb, Sb, Sc, Sm, Ta, Tb, Th, U, Yb and Zn. For the determination of the best method, it was evaluated the dendrograms together with value of the cophenetic correlation coefficient (CCC) obtained for each method. The average linkage method was more coherent in formation of the clusters, and with higher CCC values. Finally, a script with functions of the R statistical software was developed to calculate the CCC with aim to be a tool for researchers find the most appropriate method of cluster analysis for their data set.

Key words: Cluster Analysis; Hierarchical Methods; Cophenetic Correlation Coefficient; Archaeometry; INAA.

LISTA DE TABELAS

Página
Tabela 5.1 - Resultados da fração de massa das amostras de cerâmica do sítio arqueológico Prado, dados em mg/kg, n = 3454
Tabela 5.2 - Resultados da fração de massa das amostras de cerâmica do sítio arqueológico Água Limpa, dados em mg/kg, n = 8155
Tabela 5.3 - Resultados da fração de massa das amostras de cerâmica do sítio arqueológico Rezende, dados em mg/kg, n = 3157
Tabela 5.4 - Agrupamentos gerados pelo método de ligação simples com base na distância Euclidiana59
Tabela 5.5 - Agrupamentos gerados pelo método de ligação simples com base na distância Euclidiana ao quadrado
Tabela 5.6 - Agrupamentos gerados pelo método de ligação simples com base na distância <i>Manhattan</i> 66
Tabela 5.7 - Agrupamentos gerados pelo método de ligação simples com base na distância <i>Mahalanobis</i> 70
Tabela 5.8 - Agrupamentos gerados pelo método de ligação completa com base na distância Euclidiana73
Tabela 5.9 - Agrupamentos gerados pelo método de ligação simples com base na distância Euclidiana ao quadrado76
Tabela 5.10 - Agrupamentos gerados pelo método de ligação completa com base na distância <i>Manhattan</i> 79
Tabela 5.11 - Agrupamentos gerados pelo método de ligação completa com base na distância <i>Mahalanobis</i> 81
Tabela 5.12 - Agrupamentos gerados pelo método de ligação média com base na distância Euclidiana85
Tabela 5.13 - Agrupamentos gerados pelo método de ligação média com base na distância Euclidiana ao quadrado88
Tabela 5.14 - Agrupamentos gerados pelo método de ligação média com base na distância <i>Manhattan</i> 89
Tabela 5.15 - Agrupamentos gerados pelo método de ligação média com base na

Tabela 5.16 - Agrupamentos gerados pe distância Euclidiana	lo método do centróide com base na97
Tabela 5.17 - Agrupamentos gerados pe distância Euclidiana ao qua	lo método do centróide com base na adrado100
Tabela 5.18 - Agrupamentos gerados pe distância <i>Manhattan</i>	lo método do centróide com base na103
Tabela 5.19 - Agrupamentos gerados pe distância <i>Mahalanobis</i>	lo método do centróide com base na106
Tabela 5.20 - Agrupamentos gerados pe distância Euclidiana	lo método de Ward com base na 108
Tabela 5.21 - Agrupamentos gerados pe distância Euclidiana ao qua	lo método de Ward com base na idrado111
Tabela 5.22 - Agrupamentos gerados pe distância <i>Manhattan</i>	lo método de Ward com base na 114
Tabela 5.23 - Agrupamentos gerados pe distância <i>Mahalanobis</i>	lo método de Ward com base na 117
de dissimilaridades obtidas	ntre as matrizes cofenéticas e as matrizes conforme método de agrupamento122

LISTA DE FIGURAS

Página
Figura 1.1 - Infográfico da Arqueometria19
Figura 1.2 - Esquema das métricas de distâncias e métodos de agrupamentos estudados
Figura 3.1 - Diagrama de agrupamento que mostra a variação entre e dentro do agrupamento29
igura 3.2 - Diagrama das passagens da análise de agrupamentos31
Figura 3.3 - Princípio dos métodos hierárquicos aglomerativos e divisivos35
Figura 3.4 - Mapeamento do método de ligação simples36
Figura 3.5 - Mapeamento do método de ligação completa37
igura 3.6 - Mapeamento do método de ligação média38
Figura 4.1 - Exemplo de dendrograma obtido em uma análise de agrupamentos.
Figura 5.1 - Gráfico da ligação em função da distância do agrupamento das amostras de cerâmica, utilizando o método de ligação simples e com base na distância Euclidiana59
Figura 5.2 - Dendrograma do agrupamento das amostras de cerâmica, utilizando o método de ligação simples e com base na distância Euclidiana60
Figura 5.3 - Resultado do pacote <i>NbClust</i> para o método de ligação simples com base na distância Euclidiana61
Figura 5.4 - Gráfico da ligação em função da distância do agrupamento das amostras de cerâmica, utilizando o método de ligação simples e com base na distância Euclidiana ao quadrado62
Figura 5.5 - Dendrograma do agrupamento das amostras de cerâmica, utilizando o método de ligação simples e com base na distância Euclidiana ao quadrado63
Figura 5.6 - Resultado do pacote <i>NbClust</i> para o método de ligação simples com base na distância Euclidiana ao quadrado64
Figura 5.7 - Gráfico da ligação em função da distância do agrupamento das amostras de cerâmica, utilizando o método de ligação simples e com base na distância <i>Manhattan</i> 64

•	ograma do agrupamento das amostras de cerâmica, utilizando do de ligação simples e com base na distância <i>Manhattan</i> 65
_	ado do pacote <i>NbClust</i> para o método de ligação simples com a distância <i>Manhattan</i> 66
amos	co da ligação em função da distância do agrupamento das tras de cerâmica, utilizando o método de ligação simples e com na distância <i>Mahalanobi</i> s
	ltado do pacote <i>NbClust</i> para o método de ligação simples com na distância <i>Mahalanobis</i> 68
o mét	lrograma do agrupamento das amostras de cerâmica, utilizando codo de ligação simples e com base na distância <i>Mahalanobis</i>
amos	co da ligação em função da distância do agrupamento das tras de cerâmica, utilizando o método de ligação completa e pase na distância Euclidiana71
•	lrograma do agrupamento das amostras de cerâmica, utilizando codo de ligação completa e com base na distância Euclidiana.72
	ltado do pacote <i>NbClust</i> para o método de ligação completa pase na distância Euclidiana73
amos	co da ligação em função da distância do agrupamento das tras de cerâmica, utilizando o método de ligação completa e pase na distância Euclidiana ao quadrado74
o mét	lrograma do agrupamento das amostras de cerâmica, utilizando odo de ligação completa e com base na distância Euclidiana ao rado
	Itado do pacote <i>NbClust</i> para o método de ligação completo pase na distância Euclidiana ao quadrado76
amos	co da ligação em função da distância do agrupamento das tras de cerâmica, utilizando o método de ligação completa e pase na distância <i>Manhattan</i> 77
	lrograma do agrupamento das amostras de cerâmica, utilizando codo de ligação completa e com base na distância <i>Manhattan</i> .78
•	Itado do pacote <i>NbClust</i> para o método de ligação completa pase na distância <i>Manhattan</i> 79

Figura 5.22	 Gráfico da ligação em função da distância do agrupamento das amostras de cerâmica, utilizando o método de ligação completa e com base na distância <i>Mahalanobis</i>80
Figura 5.23	- Resultado do pacote <i>NbClust</i> para o método de ligação completa com base na distância <i>Mahalanobis</i> 81
Figura 5.24	- Dendrograma do agrupamento das amostras de cerâmica, utilizando o método de ligação completa e com base na distância <i>Mahalanobis</i> .
Figura 5.25	- Gráfico da ligação em função da distância do agrupamento das amostras de cerâmica, utilizando o método de ligação média e com base na distância Euclidiana83
Figura 5.26	 Dendrograma do agrupamento das amostras de cerâmica, utilizando o método de ligação média e com base na distância Euclidiana84
Figura 5.27	- Resultado do pacote <i>NbClust</i> para o método de ligação média com base na distância Euclidiana85
Figura 5.28	- Gráfico da ligação em função da distância do agrupamento das amostras de cerâmica, utilizando o método de ligação média e com base na distância Euclidiana ao quadrado86
Figura 5.29	- Dendrograma do agrupamento das amostras de cerâmica, utilizando o método de ligação média e com base na distância Euclidiana ao quadrado87
Figura 5.30	- Resultado do pacote <i>NbClust</i> para o método de ligação média com base na distância Euclidiana ao quadrado88
Figura 5.31	- Gráfico da ligação em função da distância do agrupamento das amostras de cerâmica, utilizando o método de ligação média e com base na distância <i>Manhattan</i> 89
Figura 5.32	 Dendrograma do agrupamento das amostras de cerâmica, utilizando o método de ligação média e com base na distância Manhattan90
Figura 5.33	- Resultado do pacote <i>NbClust</i> para o método de ligação média com base na distância <i>Manhattan</i> 91
Figura 5.34	- Gráfico da ligação em função da distância do agrupamento das amostras de cerâmica, utilizando o método de ligação média e com base na distância <i>Mahalanobis</i>
Figura 5.35	- Resultado do pacote <i>NbClust</i> para o método de ligação média com base na distância <i>Mahalanobis</i> 92

Figura 5.36	 Dendrograma do agrupamento das amostras de cerâmica, utilizando o método de ligação média e com base na distância Mahalanobis. 94
Figura 5.37	- Gráfico da ligação em função da distância do agrupamento das amostras de cerâmica, utilizando o método do centróide e com base na distância Euclidiana95
Figura 5.38	- Dendrograma do agrupamento das amostras de cerâmica, utilizando o método do centróide e com base na distância Euclidiana96
Figura 5.39	Resultado do pacote <i>NbClust</i> para o método do centróide com base na distância Euclidiana97
Figura 5.40	- Gráfico da ligação em função da distância do agrupamento das amostras de cerâmica, utilizando o método do centróide e com base na distância Euclidiana ao quadrado98
Figura 5.41	- Dendrograma do agrupamento das amostras de cerâmica, utilizando o método do centróide e com base na distância Euclidiana ao quadrado99
Figura 5.42	- Resultado do pacote <i>NbClust</i> para o método do centróide com base na distância Euclidiana ao quadrado100
Figura 5.43	- Gráfico da ligação em função da distância do agrupamento das amostras de cerâmica, utilizando o método do centróide e com base na distância <i>Manhattan</i>
Figura 5.44	- Dendrograma do agrupamento das amostras de cerâmica, utilizando o método do centróide e com base na distância <i>Manhattan</i> 102
	Resultado do pacote <i>NbClust</i> para o método do centróide com base na distância <i>Manhattan</i> 103
Figura 5.46	- Gráfico da ligação em função da distância do agrupamento das amostras de cerâmica, utilizando o método do centróide e com base na distância <i>Mahalanobi</i> s
Figura 5.47	- Dendrograma do agrupamento das amostras de cerâmica, utilizando o método do centróide e com base na distância <i>Mahalanobis</i> 105
Figura 5.48	- Resultado do pacote <i>NbClust</i> para o método do centróide com base na distância <i>Mahalanobis</i> 106
Figura 5.49	- Gráfico da ligação em função da distância do agrupamento das amostras de cerâmica, utilizando o método de Ward e com base na distância Euclidiana107
Figura 5.50	- Dendrograma do agrupamento das amostras de cerâmica, utilizando o método de Ward com base na distância Euclidiana109

_	Resultado do pacote <i>NbClust</i> para o método de Ward com base na distância Euclidiana110
_	Gráfico da ligação em função da distância do agrupamento das amostras de cerâmica, utilizando o método de Ward e com base na distância Euclidiana ao quadrado110
	Dendrograma do agrupamento das amostras de cerâmica, utilizando o método de Ward e com base na distância Euclidiana ao quadrado.
_	Resultado do pacote <i>NbClust</i> para o método de Ward com base na distância Euclidiana ao quadrado113
_	Gráfico da ligação em função da distância do agrupamento das amostras de cerâmica, utilizando o método de Ward e com base na distância <i>Manhattan</i> 113
•	Dendrograma do agrupamento das amostras de cerâmica, utilizando o método de Ward e com base na distância <i>Manhattan</i> 115
_	Resultado do pacote <i>NbClust</i> para o método de Ward com base na distância <i>Manhattan</i> 116
_	Gráfico da ligação em função da distância do agrupamento das amostras de cerâmica, utilizando o método de Ward e com base na distância <i>Mahalanobis</i> 116
	Dendrograma do agrupamento das amostras de cerâmica, utilizando o método de Ward e com base na distância <i>Mahalanobis</i> 118
•	Resultado do pacote <i>NbClust</i> para o método de Ward com base na distância <i>Mahalanobis</i> 119
	Gráfico da função discriminante 1 <i>v</i> s função discriminante 2. A elipse representa um nível de confiança de 95%121
Figura 5.62 –	- Gráfico de perfis124
Figura 5.63 -	Tabela gerada pelo <i>script</i> desenvolvido126

SUMÁRIO

1 INTRODUÇÃO	18
1.1 Arqueometria	18
1.2 Técnicas arqueométricas para obtenção e análise de dado	s em
cerâmicas	21
1.3 Aplicação da Estatística em estudos arqueométricos	22
2 OBJETIVOS	27
2.1 Objetivos específicos	27
3 ANÁLISE DE AGRUPAMENTOS	28
3.1 Análise das variáveis e das amostras a serem agrupadas	32
3.2 Medidas de similaridade/dissimilaridade	32
3.3 Algoritmos de agrupamentos	34
3.3.1 Método de ligação simples	35
3.3.2 Método de ligação completa	36
3.3.3 Método de ligação média	37
3.3.4 Método do centróide	38
3.3.5 Método de Ward	39
3.4 Escolha da quantidade de agrupamentos formados	40
3.5 Validação e interpretação de agrupamentos	
4 MATERIAL E MÉTODOS	42
4.1 Base de dados	42
4.2 Suporte computacional	43
4.3 Métodos estatísticos	44
4.3.1 Medidas de distâncias	44
4.3.2 Métodos hierárquicos aglomerativos	45
4.3.3 Dendrogramas	46
4.3.4 Determinação do número de grupos	47
4.3.5 Análise discriminante	48
4.3.6 Coeficiente de correlação cofenética	49
5 RESULTADOS E DISCUSSÃO	51
5.1 Seleção de variáveis	51
5.2 Transformação de dados	52
5.3 Análise de dados discrepantes (outliers)	52

5.4 Análise de agrupamentos	58
5.4.1 Determinação do número de grupos	58
5.4.2 Comparação dos agrupamentos gerados pelos métodos utilizados	120
5.5 Análise discriminante	121
5.6 Coeficiente de Correlação Cofenética	121
5.7 Interpretação dos grupos gerados pela análise de agrupamentos	123
5.8 Script do R	125
CONCLUSÃO	127
REFERÊNCIAS BIBLIOGRÁFICAS	129
APÊNDICE	139

1 INTRODUÇÃO

A Arqueologia tem por objetivo a reconstituição da história das culturas humanas, a partir de teorias, métodos e técnicas específicas, baseada nos artefatos arqueológicos e nas características locais. Sendo definida como uma parte das ciências históricas, que estuda o passado da humanidade, por meio da cultura material (JONES, 2004).

A cultura material é um dos vestígios mais antigos da vida humana e, nesse sentido, subestimar a sua relevância seria subestimar grande parte da experiência humana (MARTINS, 2011). Sendo assim, o estudo de artefatos arqueológicos é extremamente importante na Arqueologia, uma vez que, esse estudo é essencial no processo de reconstrução da pré-história humana, especialmente, para o estabelecimento das relações sociais e culturais entre as comunidades (GLIOZZO et al., 2008; SCARLETT et al., 2007).

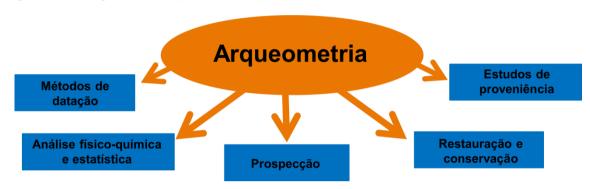
Nos últimos 50 anos, a comunicação e troca de conhecimentos entre arqueólogos, historiadores, físicos, químicos, matemáticos e estatísticos, acerca do estudo de materiais arqueológicos, mostram o quanto a interdisciplinaridade entre estas áreas tornou-se importante para classificação destes artefatos (HAYASHIDA, 2003). Nesta perspectiva, muitos trabalhos têm sido realizados gerando uma série de informações que, constantemente, se complementam caracterizando uma importante área da pesquisa, a Arqueometria.

1.1 Arqueometria

A Arqueometria é uma área de pesquisa interdisciplinar que envolve o desenvolvimento e uso de métodos científicos físico-químicos, cujo objetivo é responder a questões específicas que contribuam para o conhecimento dos objetos produzidos pelas diferentes sociedades. Permitindo revelar e identificar os materiais e tecnologias utilizadas no passado para melhorar o entendimento da história, dos processos migratórios, das características culturais, assim como, ter

parâmetros mais embasados para que haja preservação do patrimônio cultural (RIZZUTTO, 2015), dando subsídios para interpretações dos povos estudados. Na Figura 1.1, apresenta-se o infográfico da Arqueometria.

Figura 1.1 - Infográfico da Arqueometria



Fonte: autora da dissertação.

Os primeiros indícios de estudos arqueométricos surgem em 1786, quando George Pearson publica o texto: *Observations on some ancient metallic arms and utensil with experiments to determine their composition.* Entretanto, apenas no século XIX, a Arqueometria teve grande evolução nas técnicas e no desenvolvimento de métodos (LEUTE, 1987).

A partir do século XX, surge o interesse em abordar, de forma interdisciplinar, o estudo dos objetos arqueológicos, envolvendo historiadores de arte, curadores, físicos e químicos (MUNITA e CARVALHO, 2015). Um grande incentivador da interdisciplinaridade nesse período foi Edward W. Forbes, que, em 1931, criou o *Department for Conservation and Tecnical Research*, no *Fogg Art Museum* da Universidade de Harvard, hoje chamado *Straus Center for Conservation and Technical Studies*. Os trabalhos realizados nesta instituição chamaram atenção para a investigação de materiais arqueológicos e técnicas de arte, bem como para questões relacionadas à procedência de produções artísticas originais (AINSWORTH, 2005).

O grande marco da interdisciplinaridade entre a Química, a Física, a Matemática, a Estatística e a Arqueologia, foi em 1958, com o surgimento da revista *Archaeometry* na Universidade de Oxford em Londres (MUNITA, 2010), sendo a mesma impressa até os dias atuais. Além disso, estabeleceu-se o termo

"Arqueometria", introduzindo-o como um novo ramo científico, que une as ciências exatas à Arqueologia, nos estudos de objetos antigos.

De modo geral, os principais aspectos estudados em Arqueometria são a composição química, as técnicas de manufatura e as cronologias dos vestígios resgatados (FELICÍSSIMO et al., 2004). O que confirma, mais uma vez, que a interdisciplinaridade entre as ciências exatas e humanas vem contribuindo, significativamente, na obtenção de melhores resultados.

No Brasil, a inserção da Química e da Física na análise de materiais de origem arqueológica ocorreu juntamente com o desenvolvimento das pesquisas arqueológicas. Entre os primeiros trabalhos, um dos mais relevantes foi desenvolvido na década de 60 por Watanabe, que realizou datações por termoluminescência de fragmentos cerâmicos de vasos e urnas funerárias, encontradas no interior do estado de São Paulo (WATANABE, 1969).

Por um longo período, no país, os estudos arqueométricos restringiram-se às datações. A partir da década de 90, diversos laboratórios têm desenvolvido projetos para o estudo da composição química e mineralógica de diversos materiais de origem arqueológica e de objetos de arte. Atualmente, várias técnicas analíticas estão disponíveis para este tipo de estudo, direcionadas, principalmente, para estudos de proveniência, identificação de pigmentos e estudos para conservação de obras de arte.

O uso de métodos analíticos para estudos dos registros arqueológicos intensificou-se desde suas origens no século XIX até os tempos atuais, auxiliando, assim no desenvolvimento de uma ampla gama de novas técnicas e metodologias interdisciplinares, as quais envolvem as ciências exatas e humanas.

Essa abordagem interdisciplinar vem contribuindo, com muita relevância, para as pesquisas arqueométricas e para a história da arte (AINSWORTH, 2005). Como resultado disso, atualmente, é intenso o uso das propriedades físicas e químicas nos estudos de caracterização dos registros arqueológicos (FELICÍSSIMO et al., 2004), podendo ser observado pelo crescente número de artigos publicados em revistas que tratam deste assunto.

A caracterização físico-química de materiais arqueológicos é, frequentemente, empregada para tratar de questões referentes à origem da matéria-prima, processos de fabricação dos objetos e existência de interconexões

entre grupos humanos na forma, por exemplo, de redes de comunicação e rotas de comércio (PAIVA, 2006).

1.2 Técnicas arqueométricas para obtenção e análise de dados em cerâmicas

Um dos principais focos da Arqueometria, desde a década de 1950, tem sido o estudo de cerâmicas, que por sua resistência a diferentes condições ambientais, são encontradas em quantidades significativas na maioria dos sítios. Nas últimas décadas, tem havido um aumento significativo dos estudos sobre cerâmicas, visando a compreender fatores como a sua produção (COSTIN, 2000; D'ERCOLE et al., 2015), tecnologia (HEGMON, 2000), uso dos artefatos manufaturados, assim como de seu intercâmbio sociocultural, entre outros fenômenos sociais (KALENTZIDOU, 2000; YAN et al., 2015).

As características macroscópicas das cerâmicas, tais como, decoração da superfície e forma, são, frequentemente, utilizadas como indicadores culturais e cronológicos. Entretanto, as propriedades microscópicas, tais como: textura, composição química e mineralógica, podem ser usadas para estudo da tecnologia de fabricação e proveniência desses materiais (BARIA et al., 2015; GUIRAO et al., 2014).

Sob uma ótica estritamente geológica, as cerâmicas são formadas de rocha sedimentares, cujo componente principal é a argila. Os principais componentes da argila são Al₂O₃ e SiO₂, que estão presentes em quantidades superiores a 10%. Impurezas menores, tais como os óxidos de Na, Mg, K, Ca, Ti e Fe encontram-se em concentrações que variam de 0,1 a 5%. Por outro lado, os elementos ao nível do traço, menores que 0,1%, são os que, normalmente, proporcionam a melhor informação em estudos arqueométricos. característica torna estatisticamente improvável, que a concentração de diferentes elementos traço argila de uma determinada localidade quantitativamente, com a argila de objetos cerâmicos provenientes de outras localidades (GLASCOCK, 1992). Essa última afirmação constitui um postulado denominado Postulado de Proveniência (WEIGAND et al. 1977, p. 24).

De acordo com esse postulado, o estudo da composição química elementar permite determinar a origem da matéria-prima que foi usada na fabricação do artefato cerâmico, e, consequentemente, podem ser testadas as hipóteses de interações socioeconômicas entre comunidades ceramistas.

Atualmente, existem vários exemplos da aplicação de técnicas utilizadas em pesquisas arqueométricas envolvendo material cerâmico, direcionadas, principalmente, para estudos de proveniência, identificação de pigmentos e estudos para conservação de obras de arte. Dentre os diversos métodos analíticos disponíveis, podemos citar os mais comumente utilizados pelos arqueometristas: análise por ativação com nêutrons instrumental (INAA), fluorescência de raios X (XRF), espectrometria de plasma indutivamente acoplado (ICP) (FUNTUA et al., 2012; SHACKLEY, 2008), difração de raios X (XRD), datação por termoluminescência (TL) e radioacarbônicas (TUDELA et al., 2012).

Tendo em vista o crescente avanço das técnicas físico-químicas em estudos arqueométricos e a facilidade para armazenar essas informações, a quantidade de dados gerados (resultados) tem aumentado significativamente. Com isso, bancos de dados com muitas variáveis tornam-se comuns.

Para construir agrupamentos com bancos de dados de uma ou duas dimensões, geralmente, empregam-se gráficos como o de dispersão e o histograma. Entretanto, com bancos de dados com muitas variáveis, torna-se difícil o emprego de estratégias puramente visuais como estas. Faz-se necessário, então, o uso de métodos analítico-computacionais, que consideram essa estrutura multivariada, ou seja, o uso de uma análise estatística multivariada.

1.3 Aplicação da Estatística em estudos arqueométricos

A aplicação da Estatística aos dados arqueológicos já tinha sido sugerida em 1975. O emprego da Estatística, assim como a da similaridade como forma de diferenciar grupos de artefatos, faz com que a quantidade, aos poucos, se transforme em qualidade (CARANDINI, 1975). Os principais objetivos de aplicarem-se métodos estatísticos em dados arqueométricos são a exploração de resultados, geração de hipóteses, verificação da hipótese e o resumo dos resultados (MUNITA, 2005).

Dessa forma, as técnicas estatísticas multivariadas representam ferramentas poderosas para a interpretação de dados arqueológicos, sendo amplamente usadas em Arqueometria (BAXTER et al. 2008). Tais técnicas, não supervisionadas e supervisionadas, como são conhecidas, vêm acompanhadas de modernos programas computacionais, fornecendo visualização e interpretação bastante eficiente, no que se refere ao tratamento de dados arqueológicos. Por sua vez, permitem avaliar um conjunto de amostras, levando em consideração as correlações existentes entre as variáveis (FERREIRA, 2008; JOHNSON e WICHERN, 1992).

De uma forma geral, cada amostra analisada pode ser representada como um ponto no espaço multidimensional, onde cada dimensão do hiperespaço corresponde a eixos determinados pela composição química das amostras. Com o objetivo de agrupar as amostras, conforme sua similaridade/dissimilaridade, devem-se formar grupos de amostras de acordo com alguns critérios estatísticos. Os resultados podem ser organizados dentro de uma matriz de dados X_{np} , sendo n o número de amostras e p o número de variáveis (MUCHA e BARTEL, 2015).

Diversas técnicas estatísticas multivariadas têm sido utilizadas, tais como, a análise de agrupamentos, análise discriminante, análise de componentes principais, análise por meio de redes neurais, entre outras. Entretanto, a mais utilizada em Arqueometria é a análise de agrupamentos (PAPAGEORGIOU e BAXTER, 2001), que tem como objetivo dividir um conjunto de *n* amostras em *k* grupos, de modo que exista homogeneidade dentro de cada grupo e heterogeneidade entre eles (HAIR JR. et al., 2005).

O processo de agrupamento envolve basicamente duas etapas: a primeira relaciona-se com a estimação de uma medida de similaridade (ou dissimilaridade); e a segunda, com adoção de uma técnica de agrupamento para a formação dos grupos.

As distâncias são as medidas de dissimilaridade mais utilizadas no estudo de bancos de dados com variáveis quantitativas. Um grande número de medidas de dissimilaridade tem sido proposto e utilizado em análise de agrupamentos (BUSSAB et al., 1990; EVERITT et al., 2011; HAIR JR. et al., 2005; MINGOTI, 2017).

Na literatura são apresentados diversos métodos para particionar a base de dados (BARROSO e ARTES, 2003; EVERITT et al., 2011; HAIR JR. et al., 2005; MINGOTI, 2017; MURTAGH e CONTRERAS, 2012; SAXENA et al., 2017; TREBUNA e HALCINOVÁ, 2013). Esses métodos se distinguem pelo tipo de resultado a ser fornecido e pelos diferentes modos de definir a proximidade entre as amostras e os grupos formados.

Na análise de agrupamentos, os métodos mais populares usados são classificados como hierárquicos e não-hierárquicos. Por não serem excludentes, podem ser empregados separadamente ou, também, como combinação de ambos os métodos. Dentre esses métodos, os hierárquicos têm sido muito utilizados em estudos arqueométricos. KULEFF et al. (1985) e SCHUBERT (1986) utilizaram o método de ligação simples; MIRTI et al. (1990) e RAURET et al. (1987) o método de ligação completa; o método de ligação média foi utilizado por BLASIUS et al. (1983), DJINGOVA e KULEFF (1992) e HAZENFRATZ et al (2016); e o método de Ward utilizado por BARIA et al (2015), NUNES et al. (2013) e STEWART et al. (1990).

Segundo MINGOTI (2017), as várias combinações de métodos em diferentes medidas de dissimilaridade podem levar a diferentes padrões de agrupamento e a falsas interpretações, tendo o pesquisador que tomar a decisão de qual é o mais adequado para o seu propósito.

Nesse sentido, têm-se estudado os métodos de agrupamentos com o intuito de encontrar o mais adequado para a base de dados utilizada. SARAÇLI et al. (2013) cita que este problema de comparar métodos de agrupamentos para identificar o mais adequado, não é novo.

BAKER (1974) investigou o impacto de erros de observação nos dendrogramas produzidos pelas técnicas de agrupamento hierárquico de ligação completa e de ligação simples, utilizando o coeficiente gamma Goodman-Kruskal para medir a qualidade do ajuste dos dendrogramas.

KUIPER e FISHER (1975) compararam seis procedimentos hierárquicos de agrupamento (ligação simples, ligação completa, mediana, ligação média, centróide e método de Ward) para dados normais multivariados, assumindo que o verdadeiro número de clusters era conhecido. Os autores usaram o índice Rand, que fornece uma proporção de agrupamentos corretos, para comparar os métodos de agrupamentos.

BLASHFIELD (1976) comparou quatro tipos de métodos hierárquicos de agrupamento (ligação simples, ligação completa, ligação média e o método de Ward) para precisão na recuperação de aglomerados populacionais originais, usando a estatística de Cohen para medir a precisão dos métodos de agrupamento.

SARAÇLI et al. (2013) utilizaram sete métodos hierárquicos de agrupamentos sob condições diferentes: tamanho de amostra, número de variáveis e medidas de distâncias, com o objetivo de investigar o melhor método de agrupamento. E, para a comparação foi utilizado o coeficiente de correlação cofenética.

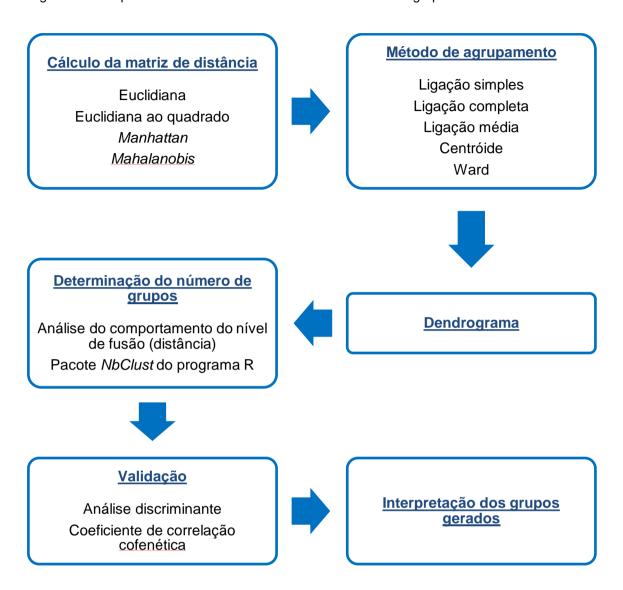
De modo geral, um método é melhor do que outro, quando o dendrograma fornece uma imagem menos distorcida da realidade (VALENTIN, 2012). Pode-se avaliar o grau de deformação provocado pela construção do dendrograma, calculando-se o coeficiente de correlação cofenética, que tem sido amplamente utilizado em estudos, tanto como uma medida do grau de adequação de uma classificação de um conjunto de dados, como um critério para avaliar a eficiência dos vários métodos de agrupamento (SARAÇLI et al., 2013).

Outra questão importante, que o pesquisador enfrenta, está relacionada em como se deve proceder para escolher o número final de grupos que define a partição do conjunto de dados analisado, ou, em qual passo o algoritmo de agrupamento deve ser interrompido.

Neste trabalho, será feito um estudo comparativo dos métodos hierárquicos aglomerativos de análise de agrupamento utilizando diferentes métricas de distância, com o intuito de identificar qual é o mais adequado em uma base de dados arqueológicos. Um esquema dos estudos realizados está representado na Figura 1.2.

O estudo foi implementado usando uma base de dados do Grupo de Estudos Arqueométricos do IPEN, de 146 amostras de fragmentos cerâmicos de três sítios arqueológicos, que foram determinados, por análise de ativação com nêutrons instrumental, 24 elementos químicos: As, Ba, Ce, Co, Cr, Cs, Eu, Fe, Hf, K, La, Lu, Na, Nd, Rb, Sb, Sc, Sm, Ta, Tb, Th, U, Yb e Zn.

Figura 1.2 - Esquema das métricas de distâncias e métodos de agrupamentos estudados.



Fonte: autora da dissertação.

2 OBJETIVOS

Uma vez que diferentes métodos de agrupamento dão origem a diferentes partições, separações de grupos, realizar-se-á um estudo comparativo para encontrar o método de partição mais adequado para a base de dados.

2.1 Objetivos específicos

- Comparar métricas de distâncias.
- Determinar o número ideal de grupos gerados pelo método de agrupamento.
- Validar os agrupamentos gerados utilizando a análise discriminante e o coeficiente de correlação cofenética.
- Desenvolver um script com funções do programa R para facilitar a detecção do método mais apropriado para uma base de dados qualquer.

3 ANÁLISE DE AGRUPAMENTOS

A análise de agrupamentos, também conhecida como análise de conglomerados ou *Cluster Analysis*, é uma técnica estatística multivariada que foi originalmente desenvolvida para classificação biológica (SARAÇLI et al., 2013). Para ALDENDERFER e BLASHFIELD (1984), uma importante contribuição ao desenvolvimento das técnicas de agrupamentos foi feita a partir do livro "*Principles of Numerical Taxonomy*" dos biólogos SOKAL e SNEATH (1963). Atualmente, é amplamente utilizada em diversas áreas do conhecimento tais como: bioinformática, meteorologia, medicina, marketing, administração, arqueometria, entre outras (OLIVEIRA et al., 2015; SAXENA et al., 2017).

Essa técnica tem como propósito primário reunir amostras, baseandose nas características das mesmas, ou seja, obter grupos que apresentem padrões (características) semelhantes e que possam refletir a forma como os dados são estruturados (SEMAAN et al., 2013).

Para HAIR JR. et al. (2005), a análise de agrupamentos agrupa amostras em grupos, de modo que, as amostras em um mesmo grupo são mais parecidas entre si do que em relação a outros grupos. É nesse sentido que o principal objetivo da análise de agrupamentos é agrupar amostras com base na similaridade ou dissimilaridade (TREBUNA e HALCINOVÁ, 2013) de variáveis predeterminadas. O grupo resultante dessa classificação deve, então, exibir um alto grau de homogeneidade interna e alta heterogeneidade externa (HAIR JR. et al., 2014), ou seja, os grupos são formados de modo a maximizar a homogeneidade dentro dos grupos, ao mesmo tempo em que se maximiza a heterogeneidade entre os grupos, conforme Figura 3.1.

Variação externa
---- Variação interna

Figura 3.1 - Diagrama de agrupamento que mostra a variação entre e dentro do agrupamento.

Fonte: HAIR JR. et al., 2014.

Convencionalmente, o processo de agrupamento pode ser definido como segue: dada uma matriz de dados A de dimensão $n \times p$, sendo n o número de amostras e p o número de variáveis, deve-se obter um conjunto de k grupos, $G = \{G_1, G_2, G_3, ..., G_k\}$, de forma a garantir que as amostras de cada grupo sejam homogêneas, segundo alguma medida de similaridade (ROUSSEEUW, 1987). Além disso, devem ser respeitadas as restrições concernentes a cada problema particular abordado. Estas restrições, definidas pelas Equações 3.1, 3.2. e 3.3, determinam, respectivamente, que: o conjunto A corresponde à união das amostras dos grupos; cada amostra pertence exatamente a um grupo; todos os grupos possuem pelo menos uma amostra (SAXENA et al., 2017; SEMAAN et al., 2013; TREBUNA e HALCINOVÁ, 2013).

$$\bigcup_{i=1}^{k} G_i = A \tag{3.1}$$

$$G_i \cap G_j = \emptyset, \ \forall \ i \neq j, \ 1 \leq i, \ j \leq k$$
 (3.2)

$$G_i \neq \emptyset, \ 1 \le i \le k \tag{3.3}$$

Desse modo, se a classificação for bem sucedida, as amostras dentro dos agrupamentos estarão próximas quando representadas graficamente e em diferentes agrupamentos, estarão distantes.

Para isso, as amostras são inicialmente tratadas de maneira individual e, em seguida, são analisadas em uma matriz de correlação ou matriz de similaridade/dissimilaridade das amostras, onde são calculadas distâncias amostra-amostra, amostra-grupo e grupo-grupo, sucessivamente, até a formação de um único grupo. De uma forma geral, quanto menor for a distância entre as amostras, maiores serão suas similaridades.

Sendo assim, basicamente, a análise de agrupamentos pode ser dividida nas seguintes etapas, que estão representadas na Figura 3.2.

Análise das variáveis e das amostras a serem agrupadas Seleção de variáveis Padronização Identificação de dados discrepantes Seleção de uma medida de similaridade/dissimilaridade As variáveis de agrupamento são métricas ou não-métricas? Dados métricos Dados não-métricos Medidas de distância Medidas de associação Medidas de correlação Seleção do algoritmo de agrupamento Usar um método hierárquico, não-hierárquico ou uma combinação dos dois? Métodos hierárquicos Combinação Ligação simples Usar um método hierárquico Ligação completa Métodos não-hierárquicos para especificar pontos Ligação média K-means sementes de agrupamento Centróide para um método Ward não-hierárquico Escolha da quantidade de agrupamentos formados Examinar aumentos no coeficiente de aglomeração Examinar dendrograma e gráficos verticais Considerações conceituais Validação e interpretação de agrupamentos

Figura 3.2 - Diagrama das passagens da análise de agrupamentos.

Fonte: adaptado de HAIR JR. et al., 2005.

3.1 Análise das variáveis e das amostras a serem agrupadas

Em qualquer aplicação, os objetivos da análise de agrupamentos, sejam eles exploratórios ou confirmatórios, não podem ser separados da escolha das variáveis usadas para caracterizar as amostras a serem agrupadas. Os grupos derivados da análise refletem a estrutura inerente dos dados, conforme definidos pelas variáveis. Por esse motivo, a escolha das variáveis deve ser feita com extremo cuidado.

Outro aspecto importante a ser considerado pelo pesquisador ao aplicar esta técnica está na utilização de variáveis em unidades diferentes, pois pode causar distorção na estrutura do agrupamento.

A maioria das medidas de distância sofre influência das diferentes escalas ou magnitudes entre as variáveis. Em geral, variáveis com maior dispersão, ou seja, maior desvio padrão, têm maior impacto sobre o valor de similaridade. Sendo, esse problema contornado com a padronização dos dados, uma vez que será atribuído o mesmo peso para cada variável (HAIR JR. et al., 2005).

Por último, a análise de agrupamentos é altamente sensível à inclusão de variáveis irrelevantes, assim como à presença de dados discrepantes (SAXENA et al., 2017; EVERITT et al.,2011). Os dados discrepantes podem ser definidos como observações que fogem do padrão esperado em cada variável, isto é, referem-se a observações com características muito destoantes dos demais membros da população, podendo assim prejudicar a qualidade dos resultados (HAIR JR. et al., 2005). Assim, antes de efetuar a análise de agrupamentos, é recomendável verificar a existência de dados discrepantes.

3.2 Medidas de similaridade/dissimilaridade

Para agrupar amostras, é necessário determinar uma medida de similaridade ou dissimilaridade que quantifique o grau de semelhança entre elas. Com base nessa medida, as amostras similares são agrupadas e as demais são colocadas em grupos separados (ALBUQUERQUE, 2005).

O conceito de similaridade de amostras dentro de um grupo desempenha o papel mais importante no processo da análise de agrupamentos (SAXENA et al., 2017), uma vez que a identificação de grupos de amostras, ou variáveis, só é possível com adoção de alguma medida de semelhança que permita à comparação objetiva entre as amostras (FÁVERO et al., 2009).

A similaridade entre objetos pode ser mensurada de várias maneiras, mas três métodos são dominantes nas aplicações de análise de agrupamentos: medidas correlacionais, medidas de distância e medidas de associação. Cada um desses métodos representa uma perspectiva particular da similaridade, dependendo de seus objetivos e do tipo de dados. As medidas correlacionais e as medidas de distância requerem dados quantitativos (métricos), enquanto as medidas de associação são destinadas ao tratamento de dados qualitativos (não métricos) (HAIR JR. et al., 2005).

Pode-se utilizar a expressão coeficiente de parecença para se referir ao critério que mede a distância entre duas amostras ou que quantifique o quanto elas são parecidas, dividindo-o, segundo BARROSO e ARTES (2003), em duas categorias: medidas de similaridade (quanto maior o valor, maior a semelhança entre as amostras) e de dissimilaridade (quanto maior o valor, menor a semelhança entre as amostras).

De um modo geral, as medidas de similaridade e de dissimilaridade são interrelacionadas e, facilmente, transformáveis entre si (BUSSAB et al., 1990).

Cabe ressaltar que apesar das medidas correlacionais terem um apelo intuitivo e serem usadas em muitas outras técnicas multivariadas, elas não são o coeficiente de parecença mais comumente empregado em análise de agrupamentos (HAIR JR. et al., 2005). As distâncias, que são medidas de dissimilaridade, são o método mais frequentemente utilizado no estudo de banco de dados com variáveis quantitativas. Um grande número de medidas de dissimilaridade tem sido proposto e utilizado em análise de agrupamentos (BUSSAB et al., 1990; MINGOTI, 2017).

Uma medida d_{ij} representa uma distância entre as amostras i e j se:

a) $d_{ij} \ge 0$ para qualquer escolha de $i \in j$;

b) $d_{ii} = 0$;

- c) $d_{ii} = d_{ii}$;
- d) $d_{ij} \leq d_{im} + d_{mk}$ (MURTAGH e CONTRERAS, 2012).

Escolhida a métrica, as distâncias d_{ij} são calculadas para todas as amostras, duas a duas, e armazenadas em uma matriz $\mathbf{D}_{(n \times n)}$, chamada de matriz de distâncias, ou de parecença, que é simétrica e com zeros na diagonal principal. Agora, o próximo passo, é escolher qual algoritmo de agrupamento será utilizado para formação dos grupos.

3.3 Algoritmos de agrupamentos

Uma vez escolhido o coeficiente de parecença, é necessário determinar o algoritmo que fará o processo de agrupamento, isto é, a formação dos grupos decorre do critério de distância entre os vetores de dados e do método de agregação escolhido.

Os algoritmos mais comumente usados podem ser classificados em duas categorias gerais: hierárquico e não-hierárquico (HAIR JR. et al., 2005).

Nos algoritmos de agrupamentos hierárquicos, conhecidos como SAHN ("Sequencial, Agglomerative, Hierarquic, Nonoverlapping Clustering Methods") (ALBUQUERQUE, 2005), os agrupamentos são formados a partir de uma matriz de parecença para identificar o par de amostras que são mais semelhantes. Esse par de amostras é agrupado e será considerado como um grupo. Isto requer que se defina uma nova matriz de parecença. Seguido da identificação do par mais semelhante que formarão um novo grupo.

De acordo com a forma de criar decomposições (Figura 3.3), os métodos hierárquicos de agrupamentos podem ser subdivididos em dois grupos (SAXENA et al., 2017; TREBUNA e HALCINOVÁ, 2013):

aglomerativos, onde cada amostra inicia-se representando um grupo, a cada passo, um grupo ou amostra é ligado a outro de acordo com sua similaridade, até o último passo, onde é formado um único grupo com todas as amostras; e divisivos, que trabalham na direção oposta dos métodos aglomerativos, ou seja, partem de um único grupo, e por divisões sucessivas vão sendo obtidos 2, 3, etc. grupos.

x₁, x₂, x₃, x₄, x₅

x₁, x₂, x₃

x₂, x₃

x₄, x₅

Figura 3.3 - Princípio dos métodos hierárquicos aglomerativos e divisivos.

Fonte: TREBUNA e HALCINOVÁ, 2013.

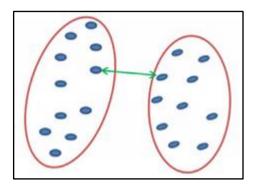
Dentre os métodos hierárquicos, os aglomerativos são os mais usados, pois os métodos divisivos exigem uma maior capacidade computacional. E nos métodos hierárquicos aglomerativos, os mais utilizados são: ligação simples, ligação completa, ligação média, centróide e Ward.

3.3.1 Método de ligação simples

O método de ligação simples é um dos métodos mais antigos, teve sua origem na década de 1950 por pesquisadores poloneses (MURTAGH e CONTRERAS, 2012). Foi primeiro descrito por FLOREK et al. (1951) e mais tarde por SNEATH (1957) e JOHNSON (1967). Segundo EVERITT et al. (2011), a distância entre dois grupos (G_1) e (G_2) é definida como distância mínima entre qualquer amostra em um agrupamento e qualquer amostra de outro (Figura 3.4) e pode ser obtida por:

$$d(G_1, G_2) = \min_{i \in G_1, i \in G_2} d_{ij}$$
(3.4)

Figura 3.4 - Mapeamento do método de ligação simples.



Fonte: SAXENA et al., 2017.

ANDERBERG (1973) cita as seguintes características desse método:

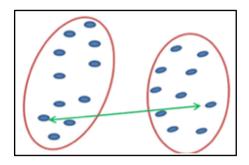
- Em geral, grupos muito próximos podem não ser identificados;
- Permite detectar grupos de formas não-elípticas;
- Apresenta pouca tolerância a ruído, pois tem tendência a incorporar os ruídos em um grupo já existente;
- Apresenta bons resultados tanto para distância Euclidiana quanto para outras distâncias;

3.3.2 Método de ligação completa

O método de ligação completa é semelhante ao método de ligação simples, exceto no caso em que a distância entre dois grupos (G₁) e (G₂) agora é definida como a maior distância entre qualquer amostra em um agrupamento e qualquer amostra de outro e não a menor (Figura 3.5) e, pode ser obtida por (EVERITT et al., 2011):

$$d(G_1, G_2) = \max_{i \in G_1} d_{ij}$$
 (3.5)

Figura 3.5 - Mapeamento do método de ligação completa.



Fonte: SAXENA et al., 2017.

Este método, geralmente, leva a grupos compactos e discretos, tendo os seus valores de dissimilaridade relativamente grande (ALBUQUERQUE, 2005).

KAUFMAN e ROUSSEEUW (1990) citam as seguintes características desse método:

- Apresenta bons resultados tanto para a distância Euclidiana quanto para outras distâncias;
- Possui tendência a formar grupos compactos;
- Os ruídos da matriz de dados demoram a serem incorporados ao grupo.

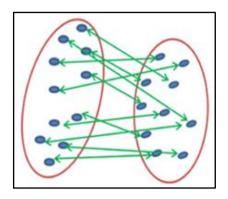
3.3.3 Método de ligação média

No método de ligação média, a distância entre dois grupos (G_1) e (G_2) é a distância média entre as amostras de G_1 e as de G_2 (Figura 3.6) e, pode ser obtida por (EVERITT et al., 2011):

$$d(G_1, G_2) = \sum_{i \in G_1} \sum_{j \in G_2} \frac{d_{ij}}{n_1 n_2}$$
(3.6)

sendo n_1 e n_2 o número de amostras nos grupos G_1 e G_2 , respectivamente.

Figura 3.6 - Mapeamento do método de ligação média.



Fonte: SAXENA et al., 2017.

Conforme ALBUQUERQUE (2005), este método tem a vantagem de evitar valores extremos e de levar em consideração toda a informação dos grupos. Um grupo passa a ser definido como um conjunto de amostras no qual cada uma delas têm mais semelhanças, em média, com todos os membros do mesmo grupo do que com todas as amostras de qualquer outro grupo.

KAUFMAN e ROUSSEEUW (1990) destacam as seguintes características desse método:

- Menor sensibilidade a ruídos que os métodos de ligação simples e completa;
- Apresenta bons resultados tanto para a distância Euclidiana quanto para outras distâncias;
- Tendência a formar grupos com número de elementos similares.

3.3.4 Método do centróide

O método do centróide foi proposto por SOKAL e MICHENER (1958). Neste método, a distância entre um grupo (G_1) e um grupo $(G_2 \cup G_3)$, formado pela fusão de dois grupos $(G_2 \in G_3)$, é definida como a distância entre os seus centróides. Cada grupo é representado pela média de suas amostras, chamado de centróide. Segundo LANCE e WILLIANS (1967), a distância entre dois grupos pode ser obtida através da seguinte expressão:

$$d(G_1, G_2 \cup G_3) = \frac{n_2}{n_2 + n_3} d(G_1, G_2) + \frac{n_3}{n_2 + n_3} d(G_1, G_3) - \frac{n_2 n_3}{(n_2 + n_3)^2} d(G_2, G_3)$$
(3.7)

sendo n_1 , n_2 e n_3 o número de amostras nos grupos G_1 , G_2 e G_3 , respectivamente.

HAIR JR. et al. (2014) citam as seguintes características desse método:

- Robustez à presença de ruídos;
- > Fenômeno da inversão.

O fenômeno da inversão ocorre quando a distância entre centróides é menor do que a distância entre grupos já formados, isso fará com que os novos grupos sejam formados em um nível inferior aos grupos já existentes, tornando o dendrograma confuso (HAIR JR. et al., 2014).

3.3.5 Método de Ward

O método de *Ward* foi proposto por WARD em 1963 (WARD, 1963) e também é chamado de "Variação Mínima" (MINGOTI, 2017). Neste método, a formação dos grupos se dá pela maximização da homogeneidade dentro dos grupos. A soma de quadrados dentro dos grupos é usada como medida de homogeneidade. Isto é, o método de *Ward* tenta minimizar a soma de quadrados dentro do grupo. Segundo LANCE e WILLIANS (1967), a distância entre um grupo (G_1) e um grupo ($G_2 \cup G_3$), formado pela fusão de dois grupos ($G_2 \in G_3$), pode ser obtida através da seguinte expressão:

$$d(G_1, G_2 \cup G_3) = \frac{(n_1 + n_2) \cdot d(G_1, G_2) + (n_1 + n_3) \cdot d(G_1, G_3) - n_1 \cdot d(G_2, G_3)}{n_1 + n_2 + n_3}$$
(3.8)

sendo n_1 , n_2 e n_3 o número de amostras nos grupos G_1 , G_2 e G_3 , respectivamente.

HAIR JR. et al. (2014) destacam que este método tende a produzir grupos com aproximadamente o mesmo número de amostras. ROMESBURG (1984) cita que o método de Ward é sensível à presença de dados discrepantes.

3.4 Escolha da quantidade de agrupamentos formados

Segundo HAIR JR. et al. (2014), uma questão de grande importância é: como se deve proceder para escolher o número final k de grupos que define a partição do conjunto de dados analisado? Ou de outra forma, em qual etapa o algoritmo de agrupamento deve ser interrompido? Não existe uma resposta exata para essa pergunta. Entretanto, existem alguns critérios que podem auxiliar na decisão final, como: análise visual das ramificações do dendrograma, sendo o corte estabelecido em pontos onde há mudança abrupta da ramificação; critérios estatísticos, como por exemplo: desvio padrão médio e coeficiente de determinação; e análise do comportamento do nível de fusão (distância). Outros métodos são mostrados em MINGOTI (2017).

3.5 Validação e interpretação de agrupamentos

Segundo BARROSO e ARTES (2003), validar o agrupamento significa certificar-se de que os grupos, realmente, diferem. Nesta etapa da análise, podem ser empregados vários testes estatísticos desde os univariados para a comparação de médias até testes multivariados. Dentre os multivariados pode-se citar a MANOVA, usada para verificar se há diferença estatisticamente significante entre os vetores de médias de cada grupo, e a análise discriminante, usada quando já são conhecidos os grupos e os seus componentes.

O coeficiente de correlação cofenética (CCC) é outra medida de validação principalmente utilizada nos métodos de agrupamento hierárquicos. Desde a sua introdução por SOKAL e ROHLF (1962), o CCC tem sido

amplamente utilizado em estudos, tanto como uma medida do grau de adequação de uma classificação de um conjunto de dados, como um critério para avaliar a eficiência das várias técnicas de agrupamento (SARAÇLI et al., 2013).

Certificando-se de que o agrupamento escolhido é adequado, deve-se interpretar os resultados. O estágio de interpretação envolve o exame de cada grupo, tendo em vista o conjunto de variáveis eleitas, a fim de que se denomine ou atribua uma identificação que descreva, assim, adequadamente a natureza dos grupos (HAIR JR. et al., 2014).

4 MATERIAL E MÉTODOS

4.1 Base de dados

No presente trabalho, foi utilizada uma base de dados de 146 amostras de fragmentos cerâmicos fornecida pelo Grupo de Estudos Arqueométricos do IPEN - CNEN/SP. As amostras são de cerâmica pré-histórica brasileira de três sítios arqueológicos (MUNITA et al., 2003):

- Prado, localizado na Fazenda Engenho Velho, na cidade de Perdizes, estado de Minas Gerais, Brasil, 19º14´25´´ LS-47º16´00´´ LW. Os vestígios arqueológicos coletados são representados por dois tipos de documentos: cerâmica e lítica (polida e não polida). Os poucos vasos cerâmicos inteiros coletados e os parcialmente reconstruídos em campo ou laboratório são lisos, sem decoração plástica ou pintura, com predomínio de médio a grande granularidade, com má seleção de grãos. Eles foram produzidos com objetivos utilitários e funerários, de acordo com a evidência arqueológica. Suas formas mais expressivas são vasos e potes globulares, tigelas esféricas e urnas;
- Água Limpa, localizado na confluência de três pequenas fazendas, na cidade do Monte Alto no norte do estado de São Paulo, Brasil, 21º15´40´´ S-48º29´47´´ W. As cerâmicas coletadas desse sítio são de dois tipos: lisas e pintadas. A pintura apresenta-se em vermelho e branco e a seleção de grãos é boa com predominância de grãos finos e médios;

➢ Rezende, localizado na fazenda Paiolão, em Piedade, Vale do Paranaíba, 7 km do município de Centralina, no estado de Minas Gerais, Brasil, 18º33´ LS, 49º13´ LW. A cerâmica produzida era simples, utilitária e funerária. O mais antigo é pré-cerâmico ocupação (ou pura lítica) de 90/130 cm de profundidade e datada de 7.300 ± 80 anos B.P.

Essas cerâmicas foram analisadas por meio da análise por ativação com nêutrons instrumental (INAA) para determinar as frações de massas de 24 elementos químicos: As, Ba, Ce, Co, Cr, Cs, Eu, Fe, Hf, K, La, Lu, Na, Nd, Rb, Sb, Sc, Sm, Ta, Tb, Th, U, Yb e Zn. Os detalhes sobre a preparação da amostra e o método analítico, utilizados pelo Grupo de Estudos Arqueométricos do IPEN - CNEN/SP, foram publicados em outro trabalho (MUNITA et al., 2003).

4.2 Suporte computacional

As técnicas multivariadas foram implementadas utilizando o ambiente R (versão 3.4.2). O R é uma linguagem de programação que permite manipular dados, fazer cálculos e construir gráficos estatísticos (DALGAARD, 2002; MURRELL, 2006; R, 2018; VENABLES et al., 2018). Caracteriza-se como um sistema completamente planejado e coerente, não apenas um conjunto ampliado de ferramentas muito específicas e inflexíveis, como são frequentes em outros programas de análise de dados, como o SAS, MINITAB, etc.

O software pode ser obtido pelo site do CRAN (R, 2018). No CRAN é possível baixar o pacote principal do R e pacotes opcionais chamados de contribuídos (em inglês *Contributed Packages*) e uma série de manuais. O software é público e gratuito, de modo que o R foi amplamente aceito por pesquisadores de todo o mundo. No entanto, por usar a linguagem de programação, R, exige ao usuário um breve conhecimento de programação.

4.3 Métodos estatísticos

Com proposta de metodologia estatística para análise de agrupamentos considerou-se a seguinte ordem: cálculo da matriz de distância, considerando-se as distâncias: Euclidiana, Euclidiana ao quadrado, *Manhattan* e *Mahalanobis*; utilização dos algoritmos: ligação simples, ligação completa, ligação média, centróide e de Ward; dendrogramas; análise discriminante; e, por último, coeficiente de correlação cofenética.

4.3.1 Medidas de distâncias

As distâncias são as medidas de dissimilaridade mais utilizadas no estudo de banco de dados com variáveis quantitativas. Um grande número de métricas de distância tem sido proposto e utilizado em análise de agrupamentos (BUSSAB et al., 1990; MINGOTI, 2017). Entre essas, as escolhidas para realizar o trabalho foram as distâncias: Euclidiana (Eq. 4.1), Euclidiana ao quadrado (Eq. 4.2), *Manhattan* (ou *City-Block*) (Eq. 4.3) e *Mahalanobis* (Eq. 4.4).

$$d_{ij} = \sqrt{\sum_{z=1}^{p} (x_{iz} - x_{jz})^2}$$
 (4.1)

$$d_{ij} = \sum_{z=1}^{p} (x_{iz} - x_{jz})^2$$
 (4.2)

$$d_{ij} = \sum_{z=1}^{p} |x_{iz} - x_{jz}|$$
 (4.3)

$$d_{ij} = \sqrt{(x_i - x_j)^T S^{-1}(x_i - x_j)}$$
 (4.4)

Sendo:

- d_{ij} a distância entre a amostra i e a amostra j, com i, j = 1, 2, 3, ..., n;
- \triangleright x_i e x_i os valores observados para as amostras *i* e *j*;
- > S a matriz de covariância para os dados do grupo.

As distâncias d_{ij} foram calculadas para todas as amostras, duas a duas, usando cada uma das métricas acima listadas, e armazenadas numa matriz $\mathbf{D}_{(nxn)}$. Totalizando, dessa forma, quatro matrizes. Para fazer o cálculo das distâncias foi utilizado o comando *vegdist* do programa R.

4.3.2 Métodos hierárquicos aglomerativos

Neste trabalho, foram utilizados os métodos: ligação simples, ligação completa, ligação média, centróide e de Ward. De forma geral, podemos descrever, com algoritmo a seguir, o processo utilizado para gerar os agrupamentos hierárquicos aglomerativos com *n* amostras:

- 1) Inicia-se com k grupos, cada um com uma única amostra, e com uma matriz simétrica de ordem n de distâncias $D = (d_{ij})$;
- Busca-se na matriz D o par de grupos mais similar (menor distância);
- Fundir os grupos u e v e nomeá-lo por (uv). Recalcular e rearranjar as distâncias na matriz D utilizando o método aglomerativo escolhido.
- Repetir os passos 2 e 3 num total de (n − 1) vezes, até que todos os objetos estejam agrupados em único grupo.

Como se pode observar, os métodos hierárquicos aglomerativos se diferenciam no passo 3, conforme BARROSO e ARTES (2003) ressaltam. O que diferencia os métodos aglomerativos é a regra de redefinição da matriz de parecença a cada união de pares de amostras. Cada método tem uma métrica diferente para calcular a distância entre as amostras. Essas métricas estão representadas nas Equações 3.4 a 3.8.

Sendo assim, para cada matriz de distância **D** calculada, utilizou-se cada um dos métodos acima para formar os agrupamentos. Para isso, foi utilizado o comando *hclust* do programa R.

4.3.3 Dendrogramas

Foram construídos 20 dendrogramas, gerados a partir da combinação das métricas de distâncias e dos métodos utilizados. Conforme HAIR JR. et al. (2005), um dendrograma é uma representação gráfica, diagrama em árvore, dos resultados de um procedimento hierárquico, no qual o eixo das abscissas representa as amostras e o eixo das ordenadas representa os níveis (distâncias) em que as amostras foram agrupadas. Começando com cada amostra representada como um agrupamento separado, o dendrograma ilustra graficamente como os agrupamentos são combinados, em cada passo do procedimento, até que todos estejam contidos em um único agrupamento (Figura 4.1).

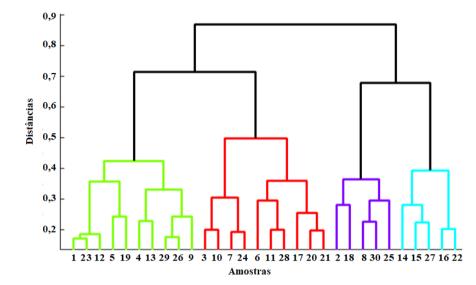


Figura 4.1 - Exemplo de dendrograma obtido em uma análise de agrupamentos.

Fonte: HAZENFRATZ-MARKS, 2014.

4.3.4 Determinação do número de grupos

Dentre os critérios apresentados por MINGOTI (2017) para auxiliar a escolha da quantidade de agrupamentos formados, adotou-se o critério da análise do comportamento do nível de fusão (distância).

À medida que se avança no algoritmo de agrupamento, ou seja, passase da etapa x para a etapa x + 1, a similaridade entre as amostras que estão sendo combinadas nas respectivas etapas vai decrescendo. Consequentemente, a distância entre elas vai aumentando. Desse modo, foi feito um gráfico da ligação em função da distância (nível de fusão) do agrupamento de cada etapa do processo, com objetivo de visualizar se há "ponto de salto" relativamente grande em relação aos demais valores de distância. Esses pontos indicam o momento ideal de parada do algoritmo, isto é, o número de grupos final e a sua composição. Foi feito um gráfico para cada combinação de medida de distância e método hierárquico aglomerativo. Gerando, dessa forma, 20 gráficos.

Além dessa análise, utilizou-se o pacote *NbClust* do programa R para auxiliar na decisão do número ideal de grupos. Na literatura, uma ampla variedade de índices tem sido proposta para encontrar o número ideal de grupos em um particionamento de um conjunto de dados, durante o processo da análise

de agrupamentos (DUNN, 1974; HALKIDI et al., 2000; HALKIDI e VAZIRGIANNIS, 2001; KAUFMAN e ROUSSEEUW, 1990; LEBART et al. 2000; MILLIGAN e COOPER, 1985; ROUSSEEUW, 1987; TIBSHIRANI et al., 2001). No entanto, para a maioria dos índices propostos na literatura, os programas não estão disponíveis para testá-los e compará-los. O pacote *NbClust* do programa R foi desenvolvido para esse propósito (CHARRAD et al., 2014; *R*, 2018).

O pacote *NbClust* fornece 30 índices para determinar o número apropriado de grupos definidos pela partição do conjunto de dados analisado. Os índices utilizados neste pacote são: Calinski-Harabasz, Duda, estatística Pseudo T², Cindex, Gamma, Beale, Cubic Clustering Criterion, Ptbiserial, G-plus, Davies-Bouldin (DB), Frey, Hartigan, Tau, Ratkowsky, Scott, Marriot, Ball, Trcovw, Tracew, Friedman, McClain, Rubin, KL, Silhouette, Gap, Dindex, índice de Dunn, estatística de Hubert, SD e S-Dbw (CHARRAD et al., 2014).

Este pacote permite que o usuário varie, simultaneamente, o método de agrupamento, a medida de distância e os índices para decidir a forma mais adequada de agrupar as amostras em seu conjunto de dados ou comparar todos os índices utilizados a fim de determinar o número mais apropriado de grupos.

4.3.5 Análise discriminante

Análise discriminante é uma técnica estatística multivariada utilizada para diferenciar grupos e/ou classificar amostras em grupos previamente definidos.

A aplicação desta técnica pode estar relacionada ao interesse do pesquisador em realizar predições precisas de pertinência de objetos amostrais a grupos já existentes e conhecidos, ou uma descrição da natureza das diferenças entre esses grupos, ou de ambos os tipos de informação (WARNER, 2013).

Segundo BARROSO e ARTES (2003), os principais objetivos da análise discriminante são:

discriminação: que consiste em encontrar funções das variáveis observadas (funções discriminantes) que são responsáveis ou que possam explicar as diferenças entre os k grupos; classificação ou alocação: que consiste em determinar funções das variáveis observadas que permitem classificar novas amostras em um dos k grupos.

Dessa forma, a técnica consiste em obter funções matemáticas capazes de classificar uma amostra em um dos grupos G_i (i = 1, 2, ..., k), com base em medidas de um número p de características (variáveis), buscando minimizar a probabilidade de má classificação, isto é, minimizar a probabilidade de classificar erroneamente uma amostra em um grupo G_i , quando realmente pertence ao grupo G_i ($i \neq j$) com i, j = 1, 2, ..., k (MINGOTI, 2017).

Diferentemente da análise de agrupamentos, a análise discriminante é uma técnica supervisionada, pois neste tipo de análise há necessidade do conhecimento, *a priori*, dos grupos aos quais pertencem as amostras.

Segundo MINGOTI (2017), uma aplicação interessante da análise discriminante surge com a análise de agrupamentos. Supondo-se que o usuário tenha obtido sua partição final dos dados amostrais em *k* grupos através de algum dos métodos, ele poderá entrar com estes grupos na análise discriminante para avaliar as taxas de classificações incorretas e validar ou não sua partição. Assim sendo, a análise discriminante, comporta-se como uma técnica confirmatória da análise de agrupamentos.

Neste trabalho, a análise discriminante foi utilizada para validar os grupos produzidos pela análise de agrupamentos.

4.3.6 Coeficiente de correlação cofenética

Depois de aplicar o método escolhido para a formação de grupos, utiliza-se o coeficiente de correlação cofenética (CCC) para verificar a qualidade do agrupamento.

$$CCC = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (c_{ij} - \bar{c}) (d_{ij} - \bar{d})}{\sqrt{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (c_{ij} - \bar{c})^{2}} \sqrt{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (d_{ij} - \bar{d})^{2}}}$$
(4.5)

Sendo:

- c_{ij} = valor de dissimilaridade entre as amostras i e j, obtidos a partir da matriz cofenética;
- d_{ij} = valor de dissimilaridade entre as amostras i e j, obtidos a partir da matriz de dissimilaridade.

$$\bar{c} = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} c_{ij}$$
 (4.6)

$$\bar{d} = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} d_{ij}$$
 (4.7)

A correlação cofenética consiste em realizar uma comparação entre as distâncias efetivamente observadas entre as amostras e as distâncias previstas a partir de um processo de agrupamento (ALBUQUERQUE, 2005; BARROSO e ARTES, 2003), medindo o grau de ajuste entre a matriz de dissimilaridade original, matriz D, e a matriz resultante da simplificação proporcionada pelo método de agrupamento, matriz C (ALBUQUERQUE, 2005).

Para os diversos métodos de agrupamento utilizados foram obtidas as respectivas matrizes cofenéticas, resultantes da simplificação proporcionada pelo método. A matriz cofenética foi obtida após a construção do dendrograma. Com base nas matrizes de dissimilaridade original e cofenética, foi obtida a correlação cofenética conforme a Equação 4.6.

Nota-se que essa correlação equivale à correlação de Pearson entre a matriz de dissimilaridade original e aquela obtida após a construção do dendrograma. Assim, quanto mais próximo de um, menor será a distorção provocada pelo agrupamento das amostras com os métodos, ou seja, melhor será a representação produzida pelo dendrograma e, então, melhor será a qualidade do agrupamento. Por outro lado, quanto mais próximo de zero, pior será a representação (BUSSAB et al. 1990).

5 RESULTADOS E DISCUSSÃO

Neste capítulo, são apresentados os resultados das análises realizadas neste trabalho. Preferiu-se apresentar a seção Resultados e Discussão de cada aplicação, separadamente, para facilitar a compreensão.

5.1 Seleção de variáveis

Em estudos composicionais em que são determinados valores de concentrações maiores, menores e traços, é de fundamental importância que o método de análise empregado tenha uma boa precisão. A precisão reflete a adequabilidade do método e é um parâmetro importante para a confiabilidade do mesmo (HAZENFRATZ-MARKS, 2014). Elementos químicos que são determinados com baixo nível de precisão podem influenciar, significativamente, na interpretação dos resultados.

Um critério utilizado é o de selecionar, como adequados, para análises estatísticas multivariadas, aqueles elementos químicos cujo desvio-padrão relativo não ultrapasse 10% em relação ao material de referência utilizado no controle de qualidade (IAEA, 2003; MUNITA et al., 2004).

Esta análise foi feita por MUNITA et al. (2003), que selecionaram como variáveis 13 elementos químicos: As, Ce, Cr, Eu, Fe, Hf, La, Na, Nd, Sc, Sm, Th e U. E, devido a confiabilidade, para os resultados deste trabalho, foram selecionados os mesmos elementos químicos como variáveis.

5.2 Transformação de dados

Inicialmente, as variáveis selecionadas para a análise de agrupamentos tiveram suas frações de massas transformadas em logaritmos de base decimal (log10), procedimento usual em estudos arqueométricos (GLASCOCK e NEFF, 2003). Um dos objetivos desta transformação é prevenir a atribuição de pesos excessivos a elementos químicos que estejam presentes em altas concentrações, em relação aos elementos presentes em nível traço (HAZENFRATZ et al., 2016).

A distribuição de muitos elementos-traço é, frequentemente, considerada como lognormal, como pode ser visto em BEIER e MOMMSEN, (1994), GLASCOCK (1992), KOCH e LINK (2002) e SAYRE (1975). Sendo assim, as frações de massas também foram transformadas para melhorar a condição de normalidade.

A transformação em logaritmo de base decimal, também, pode ser vista como uma semipadronização mais "absoluta", apresentando vantagens sobre procedimentos clássicos de padronização que utilizam a média e o desvio-padrão dos dados, uma vez que para novos dados introduzidos na base, as coordenadas padronizadas de uma determinada amostra deveriam ser atualizadas, devido à modificação da média e do desvio (HARBOTTLE, 1976). Adicionalmente, a transformação em logaritmos de base decimal é um tipo de transformação não linear que traz potenciais dados discrepantes da extremidade das frações de massa, para mais perto da média, reduzindo seu impacto nas análises multivariadas (WARNER, 2013).

5.3 Análise de dados discrepantes (outliers)

Um conjunto de medidas pode apresentar valores que incorporam erros sistemáticos, dentre os quais, os mais prováveis ocorrem devido aos erros oriundos ao equipamento de medida e aos erros durante a preparação da amostra (MUNITA, 2005). A presença desses valores, dados discrepantes, distorce a verdadeira estrutura e tornam os agrupamentos obtidos não representativos da verdadeira estrutura da população (HAIR JR. et al, 2005).

Dessa forma, faz-se necessária a sua identificação e eliminação antes da aplicação da análise de agrupamentos.

A análise de dados discrepantes foi realizada, utilizando-se a distância de *Mahalanobi*s, calculada pela equação 5.1, e o critério de lambda *Wilks* (D_{crítico}) como valor crítico (OLIVEIRA e MUNITA, 2003).

$$D = \sqrt{(x - \bar{x})^{T} S^{-1}(x - \bar{x})}$$
 (5.1)

na qual x é o vetor-coluna das variáveis medidas de uma amostra, \bar{x} é o vetor-coluna das médias para o grupo em questão, e S é a matriz de covariância para os dados do grupo.

Considerando a base de dados com n amostras e p variáveis medidas, o $D_{\text{crítico}}$ é calculado por:

$$D_{crítico} = \frac{p (n-1)^2 F_{p, n-p-1, \alpha/n}}{n (n-p-1+pF_{p, n-p-1, \alpha/n})}$$
(5.2)

sendo F a distribuição de *Fisher* e α o nível de significância (0,05).

Quando o valor calculado para a distância de *Mahalanobis* for superior ao valor crítico de *Wilks*, a amostra é considerada *outlier*. Dessa forma, foram calculadas a distância de *Mahalanobis* para cada amostra, última coluna da Tabela 5.1 e as duas últimas colunas das Tabelas 5.2 e 5.3, e, o valor crítico, calculado pela Equação 5.2, é apresentado nas últimas linhas das Tabelas 5.1, 5.2 e 5.3. As amostras que ultrapassaram o valor crítico foram eliminadas e o mesmo procedimento foi repetido para as amostras restantes. O procedimento terminou quando as amostras apresentaram valores de distância de *Mahalanobis*, inferiores ao valor crítico.

Tabela 5.1 - Resultados da fração de massa das amostras de cerâmica do sítio arqueológico Prado, dados em mg/kg, n = 34.

Amostra	As	Се	Cr	Eu	Fe	Hf	La	Na	Nd	Sc	Sm	Th	U	D ₁
A01	1,80	117,50	175,00	1,01	17300,00	10,00	38,50	786,00	57,00	26,69	7,75	19,20	4,50	13,59
A02	1,54	103,80	110,00	1,43	30600,00	9,60	36,90	731,00	48,00	27,44	7,92	16,70	3,70	10,80
A03	2,25	106,10	186,00	0,97	20900,00	10,10	31,77	678,00	52,00	27,15	6,73	18,20	4,20	14,00
A04	1,60	137,20	186,00	1,28	17200,00	11,00	38,90	727,00	45,00	26,96	8,07	19,50	4,70	21,96
A05	1,08	119,60	130,00	1,45	27400,00	8,50	35,80	526,00	51,00	28,76	8,12	18,00	3,90	5,80
A06	2,50	113,40	123,00	1,51	38100,00	8,80	31,50	302,00	35,00	31,51	7,74	17,80	4,60	15,37
A07	1,50	108,20	109,00	1,47	32600,00	9,50	41,10	766,00	40,00	26,11	7,81	18,00	4,30	14,27
A08	1,30	110,00	138,00	1,36	28900,00	8,40	30,70	552,00	34,00	29,46	6,81	17,70	1,80	16,97
A09	1,80	105,40	142,00	1,16	26600,00	9,30	27,20	543,00	26,00	27,91	6,35	16,40	3,30	11,54
A10	1,80	108,20	157,00	1,26	30700,00	9,20	29,30	552,00	36,00	31,40	6,75	17,90	6,30	12,58
A11	1,30	129,70	136,00	1,42	28000,00	9,00	34,30	533,00	38,00	28,75	7,30	17,80	5,30	7,81
A12	1,80	117,60	156,00	1,40	29800,00	8,80	33,00	590,00	32,00	30,16	7,43	18,70	3,50	6,26
A13	1,60	113,50	164,00	1,36	29900,00	9,10	29,50	555,00	26,00	31,43	6,82	17,60	4,10	7,98
A14	1,40	120,90	152,00	1,42	29600,00	9,00	33,50	621,00	39,00	30,37	7,76	18,50	5,40	8,37
A15	1,80	113,50	170,00	1,27	29900,00	9,50	30,00	635,00	27,00	31,29	7,00	17,20	4,30	11,11
A16	1,60	119,40	151,00	1,37	28300,00	8,20	33,20	590,00	34,00	28,94	7,49	18,00	5,00	8,08
A17	0,67	117,40	129,00	1,62	20400,00	8,30	32,80	535,00	40,00	33,17	8,45	17,10	4,50	13,37
A18	0,90	137,70	126,00	1,60	23400,00	7,60	45,60	532,00	44,00	30,31	9,26	20,00	5,40	18,67
A19	0,98	116,10	134,00	1,46	22400,00	9,00	30,10	476,00	49,00	33,56	8,03	16,90	4,60	12,32
A20	1,40	102,90	114,00	1,36	36100,00	8,70	40,40	644,00	38,00	27,64	7,84	17,00	4,30	14,26
A21	2,00	86,50	116,00	1,24	34700,00	9,40	32,90	643,00	37,00	27,30	7,74	16,50	3,40	14,43
A22	1,20	113,20	138,00	1,33	28000,00	8,50	31,40	557,00	29,00	28,62	7,02	15,80	4,80	11,33
A23	1,48	115,70	136,00	1,37	28700,00	9,10	32,40	548,00	41,00	29,45	7,39	16,70	4,60	3,61
A24	1,46	104,00	136,00	1,30	26300,00	8,40	29,33	579,00	38,00	27,63	6,83	16,00	3,50	8,97
A25	1,30	121,50	135,00	1,50	21700,00	8,40	33,00	516,00	36,00	33,88	8,08	18,00	4,10	9,63
A26	1,80	116,70	134,00	1,52	33500,00	8,70	33,70	484,00	37,00	30,64	7,91	17,20	3,70	5,30
A27	1,60	115,40	124,00	1,68	38400,00	8,40	30,40	328,00	43,00	32,48	7,43	17,70	3,90	12,19
A28	2,10	111,70	117,00	1,43	33300,00	8,20	29,70	500,00	32,00	30,23	7,27	16,70	3,50	9,56
A29	1,70	120,30	115,00	1,70	36000,00	9,00	32,60	377,00	40,00	30,72	8,09	16,60	4,90	10,27
A30	1,54	108,80	115,00	1,39	21600,00	7,60	28,10	507,00	29,00	29,93	6,82	16,40	3,30	16,96
A31	2,10	121,00	121,00	1,61	37300,00	9,10	33,50	493,00	34,00	31,80	6,63	17,60	5,20	15,06
A32	1,54	117,60	144,00	1,40	28200,00	8,30	32,10	656,00	31,00	29,33	7,27	17,30	4,20	5,51
A33	1,80	131,00	140,00	1,64	26500,00	8,90	35,30	593,00	46,00	29,07	6,50	16,50	5,00	21,65
A34	1,40	112,50	140,00	1,34	27900,00	8,30	31,60	558,00	36,00	28,45	7,15	17,00	2,60	6,45
													D _{crítico}	23,79

A partir da Tabela 5.1, pode-se observar que para as 34 amostras o valor da distância de *Mahalanobis* foi inferior ao valor crítico. Sendo assim, para o sítio de Prado não foi identificado nenhuma amostra discrepante.

Tabela 5.2 - Resultados da fração de massa das amostras de cerâmica do sítio arqueológico Água Limpa, dados em mg/kg, n = 81 (continua).

Amostra	As	Се	Cr	Eu	Fe	Hf	La	Na	Nd	Sc	Sm	Th	U	D ₁	D ₂
B01	1,50	108,30	134,20	2,52	32000,00	7,82	64,10	1961,00	63,00	12,87	8,89	9,81	1,30	11,74	11,70
B02	4,40	133,50	148,00	2,77	45500,00	8,40	86,60	2487,00	66,00	14,91	11,66	12,80	1,90	9,18	8,97
B03	1,40	110,60	156,00	2,30	33900,00	9,60	71,40	2938,00	60,00	14,52	9,79	12,00	2,00	9,21	9,44
B04	4,60	124,40	141,70	2,70	56100,00	7,18	79,60	2042,00	82,00	15,35	9,74	11,36	1,20	14,65	14,70
B05	2,40	98,80	128,00	1,98	39500,00	7,50	56,30	1065,00	52,00	13,13	7,77	10,70	0,99	16,18	16,56
B06	1,50	180,60	275,00	3,54	22500,00	8,20	91,40	700,00	80,00	21,91	14,35	19,10	1,80	36,79	
B07	2,20	84,90	125,00	1,86	31200,00	6,80	59,60	3151,00	49,00	10,16	7,54	9,06	1,40	18,77	18,48
B08	1,50	113,20	202,00	2,99	42100,00	8,40	78,40	3075,00	85,00	19,80	10,29	12,12	1,30	19,48	19,18
B09	4,60	102,80	107,40	2,42	43700,00	6,80	77,00	1647,00	48,00	13,25	9,79	10,53	1,04	11,16	13,62
B10	3,60	99,90	139,00	2,18	35000,00	9,30	66,80	2514,00	47,00	12,75	8,93	10,80	1,30	6,21	6,60
B11	2,70	122,30	133,00	2,57	38600,00	6,30	83,40	1487,00	64,00	15,23	10,14	12,60	0,99	12,33	12,02
B12	2,50	133,60	182,00	2,28	32800,00	9,80	70,70	1699,00	57,00	18,15	9,83	17,20	1,60	10,03	10,23
B13	1,80	102,60	118,00	2,23	34900,00	5,70	72,90	2260,00	46,00	12,60	9,24	9,71	1,20	13,69	14,24
B14	2,00	111,90	138,00	2,31	37800,00	8,40	62,70	2254,00	49,00	12,60	8,43	12,10	0,90	13,07	13,13
B15	2,00	107,40	122,00	2,58	33500,00	5,40	76,00	2129,00	63,00	13,59	10,08	9,70	1,10	15,85	16,12
B16	1,10	137,50	172,60	2,60	25500,00	9,60	73,70	1638,00	80,00	17,19	9,97	14,08	1,30	11,85	12,62
B17	1,60	99,20	148,00	2,28	35100,00	7,80	67,60	1982,00	52,00	12,58	8,94	10,20	1,20	7,22	9,10
B18	2,20	127,10	143,00	2,39	34400,00	8,00	71,40	2052,00	62,00	14,78	9,28	12,40	1,20	3,13	3,72
B19	2,20	116,40	130,00	2,24	28000,00	9,00	70,10	1660,00	60,00	12,20	9,68	10,50	1,50	12,45	13,63
B20	1,20	125,60	150,00	2,67	34400,00	9,30	83,40	1617,00	51,00	17,24	11,34	13,50	1,30	12,45	12,55
B21	2,40	132,50	148,00	3,06	37800,00	8,10	80,90	2957,00	64,00	15,27	11,72	10,50	1,70	17,51	17,17
B22	2,20	110,90	155,00	2,70	44500,00	7,90	75,70	1940,00	69,00	14,71	10,26	10,90	1,30	5,82	6,89
B23	0,90	137,90	195,00	2,24	21400,00	8,10	54,00	1928,00	39,00	18,31	7,34	13,20	1,60	24,38	24,74
B24	4,40	120,10	159,00	2,36	42000,00	9,00	70,20	2535,00	58,00	15,16	10,25	12,40	1,20	9,43	9,52
B25	2,10	123,90	141,80	2,62	38800,00	8,30	73,00	2414,00	66,00	14,78	9,65	12,06	1,20	4,77	5,87
B26	3,90	123,80	175,00	2,65	43900,00	9,10	72,50	2254,00	63,00	16,78	10,17	15,00	1,30	6,57	6,80
B27	1,80	151,50	150,00	2,56	29000,00	7,60	79,30	2188,00	59,00	14,61	11,08	12,00	1,10	11,99	13,08
B28	2,40	158,80	215,00	2,63	20600,00	9,10	72,50	1384,00	55,00	18,23	10,50	18,10	2,00	13,15	20,95
B29	1,90				25600,00	8,80	66,80	1941,00	64,00	20,26	10,02	15,50	0,99	15,61	16,96
B30	2,20	100,00	145,00		37200,00	9,90	58,20	1919,00	•	-	8,14	11,40	1,20	8,58	8,43
B31	2,40		-	1,98		8,20	61,00	1496,00	•	,	7,51	11,80	1,80	13,07	13,29
B32	2,50	160,30	183,00	3,79		7,60	96,80	2613,00	-		13,10	14,20	1,20	16,48	18,40
B33	2,10	98,20	130,00		32600,00	7,80	61,10				8,07	9,90	1,40	9,56	11,00
B34	1,70	227,00	275,00	•	22900,00	9,80	116,60	1187,00	•	•	•	14,00	1,60	46,57	
B35	2,10	126,50	•		35100,00	8,50	77,30	1526,00	62,00	-	•	15,40	0,80	12,30	12,23
B36	2,20	141,70			45700,00	8,30	95,80	1304,00					•	10,18	14,25
B37	2,20	97,80	130,00	,	39200,00	7,30	67,20	2765,00	•	-	8,69	10,80	1,10	14,21	14,03
B38	0,90	133,00	188,00		19700,00	7,20	57,80	1729,00	•	-	7,88	12,90	1,20	12,26	12,32
B39	1,50				25500,00	8,00	58,40	1625,00	•	-	7,86	13,40	0,99	9,29	11,34
B40	1,60	148,70			27100,00	7,80	70,30	1851,00			10,46		1,10	18,92	
B41	1,60	113,80	•		30400,00	10,00	54,90	1604,00	•	-	7,03	12,20	1,40	13,53	14,17
B42	1,80	142,70	168,00	2,52	26700,00	8,20	78,70	1322,00	53,00	16,73	10,60	14,60	1,33	5,69	7,66

Tabela 5.2 - Resultados da fração de massa das amostras de cerâmica do sítio arqueológico Água Limpa, dados em mg/kg, n = 81 (continuação).

Amostra	As	Се	Cr	Eu	Fe	Hf	La	Na	Nd	Sc	Sm	Th	U	D ₁	D ₂
B43	3,30	123,40	151,00	2,61	40800,00	7,80	66,80	1702,00	54,00	16,26	9,04	14,00	0,99	10,01	10,32
B44	2,70	115,20	145,00	2,48	31500,00	7,40	70,00	2166,00	61,00	13,86	9,36	11,60	1,50	4,23	4,17
B45	1,20	137,20	144,00	2,56	28300,00	8,40	72,60	1706,00	59,00	15,02	10,05	12,70	1,10	6,12	6,29
B46	1,50	104,60	135,00	2,12	24500,00	9,20	60,70	1015,00	46,00	14,87	8,16	13,70	1,30	10,99	11,44
B47	4,50	148,20	173,00	2,37	44900,00	9,00	68,00	2435,00	66,00	16,62	9,42	12,70	1,20	16,02	15,76
B48	2,10	146,30	242,00	3,10	37900,00	10,20	84,70	4068,00	81,00	20,37	12,09	14,40	1,30	18,50	19,53
B49	0,99	127,40	171,00	2,10	18500,00	6,50	57,90	841,00	54,00	17,19	8,10	12,50	1,20	20,07	22,83
B50	2,00	116,80	183,00	2,27	26800,00	8,10	61,40	1690,00	59,00	17,70	8,21	14,00	1,40	6,34	6,35
B51	5,80	124,80	177,00	2,73	48300,00	8,80	74,20	2156,00	68,00	17,35	10,30	13,10	1,30	5,83	5,98
B52	2,30	105,10	142,50	2,09	22300,00	8,50	62,50	1250,00	61,00	14,44	8,83	15,00	1,60	20,09	20,38
B53	1,10	119,50	184,00	2,57	27200,00	9,60	68,40	1835,00	50,00	16,46	9,47	13,80	1,40	4,94	5,61
B54	3,30	109,10	127,00	2,08	49700,00	8,00	64,50	2340,00	55,00	10,94	8,46	10,23	1,50	12,07	12,28
B55	1,60	104,50	150,00	2,42	30900,00	7,70	61,80	2437,00	47,00	12,82	8,73	11,00	1,28	10,05	10,34
B56	2,30	104,70	161,00	2,19	29100,00	9,00	63,00	2499,00	50,00	14,95	8,20	12,80	1,20	6,50	7,47
B57	2,70	104,00	129,20	2,35	40100,00	8,60	60,70	2310,00	60,00	13,40	9,10	11,63	1,80	13,97	16,36
B58	0,99	120,90	141,00	2,84	32600,00	7,00	87,10	1408,00	59,00	14,87	11,19	12,20	1,50	8,58	9,80
B59	2,70	115,10	155,00	3,04	35800,00	7,60	79,20	1725,00	62,00	15,74	10,71	12,00	1,30	6,51	7,12
B60	1,90	85,50	147,00	2,33	28800,00	10,40	61,50	1480,00	44,00	14,02	9,28	11,70	1,60	20,81	21,28
B61	2,65	117,30	187,00	2,18	27400,00	10,50	67,30	2627,00	57,00	16,14	9,05	14,90	2,36	14,89	14,62
B62	2,70	123,10	186,00	2,72	33200,00	8,60	71,60	2367,00	59,00	17,57	8,95	13,00	1,50	9,65	10,23
B63	2,10	126,80	166,00	2,45	35900,00	8,20	65,60	1693,00	59,00	16,32	9,62	12,70	1,50	6,12	7,11
B64	2,40	120,10	141,00	2,19	33200,00	7,30	59,90	1665,00	52,00	15,04	8,76	12,90	2,00	14,33	17,16
B65	2,00	117,50	184,00	2,45	33900,00	9,20	69,50	2151,00	57,00	16,98	9,81	14,60	2,00	6,59	7,24
B66	1,80	121,60	160,00	2,55	29300,00	8,60	72,40	1712,00	63,00	16,41	9,88	11,10	1,20	8,42	8,35
B67	3,10	96,00	145,00	2,24	45700,00	7,80	61,20	2599,00	49,00	13,04	8,00	11,60	1,10	11,13	11,04
B68	1,30	152,10	158,00	2,57	24800,00	7,40	80,70	1985,00	68,00	15,28	10,05	12,60	1,10	11,47	14,47
B69	1,10	125,50	182,00	2,21	18100,00	9,80	68,90	1284,00	50,00	17,47	9,31	14,80	1,50	11,19	11,49
B70	1,80	138,50	192,00	2,67	32100,00	9,30	78,20	2183,00	57,00	19,71	10,54	15,50	1,70	7,53	7,61
B71	1,20	125,20	158,00	2,79	29800,00	9,20	71,30	1176,00	58,00	17,68	9,89	13,00	1,40	10,49	10,24
B72	2,00	131,90	169,00	2,98	34900,00	9,30	77,60	1037,00	60,00	17,77	10,34	14,40	1,70	13,26	13,89
B73	2,00		152,00			8,70	89,00	2119,00	64,00	15,81	10,76	12,40	1,70	11,16	11,64
B74	1,40	115,10	147,00	2,34	27500,00	7,70	65,30	2181,00	47,00	14,60	8,90	11,50	1,20	3,89	4,14
B75	3,00	127,30	166,00	2,63	41000,00	9,90	80,90	2223,00	72,00	16,99	11,16	14,00	1,20	7,48	7,28
B76	1,10	116,30	130,00	2,13	26000,00	7,80	66,50	1353,00	44,00	12,67	8,15	11,20	1,20	8,86	10,25
B77	1,40	112,70	-	-	31600,00	8,10	69,40	1536,00	50,00	13,08	8,99	11,60	1,50	3,56	4,00
B78	3,60	111,00	156,00	2,33	34500,00	8,40	64,90	2140,00	48,00	14,72	8,56	12,20	1,30	2,59	2,68
B79	2,00	142,00	166,00	3,07	41300,00	8,30	86,40	2376,00	72,00	16,85	11,62	13,90	1,40	4,60	4,70
B80	4,80	138,80	222,00	2,50	24900,00	8,70	64,70	1500,00	57,00	19,88	9,03	15,70	1,70	7,44	9,27
B81	0,70	105,70	184,00	2,48	29900,00	8,90	73,90	1512,00	60,00	19,73	9,84	12,80	1,20	18,39	18,44
													$\textbf{D}_{\text{crítico}}$	30,96	30,79

Tabela 5.3 - Resultados da fração de massa das amostras de cerâmica do sítio arqueológico Rezende, dados em mg/kg, n = 31.

Amostra	As	Се	Cr	Eu	Fe	Hf	La	Na	Nd	Sc	Sm	Th	U	D ₁	D ₂
C01	2,60	67,80	212,00	2,94	11270,00	10,80	31,80	132,00	41,00	39,90	9,43	6,40	1,30	7,11	6,86
C02	1,70	75,80	205,00	2,94	8550,00	12,50	31,80	121,00	45,00	41,75	8,98	6,90	1,60	9,98	10,59
C03	1,90	61,10	215,00	2,91	10320,00	10,90	30,80	176,00	45,00	46,20	9,12	7,30	1,50	8,25	8,02
C04	1,60	56,40	183,00	2,39	8160,00	10,80	28,00	120,00	35,00	43,40	7,45	6,40	1,50	15,57	15,21
C05	2,80	68,60	215,00	2,95	11660,00	11,80	34,00	145,00	48,00	45,00	9,26	6,30	1,40	2,94	3,82
C06	2,00	61,70	212,00	2,98	8660,00	10,80	34,00	125,00	48,00	47,40	9,17	6,90	1,70	15,95	15,48
C07	2,20	62,50	195,00	2,82	9130,00	11,30	29,30	92,00	46,00	42,46	9,21	7,10	1,30	8,87	9,72
C08	1,60	82,00	187,00	3,20	11170,00	10,80	37,20	260,00	47,00	37,17	9,80	4,80	1,20	18,94	18,29
C09	1,50	90,80	303,00	3,20	12120,00	11,00	39,50	266,00	52,00	41,72	10,21	5,60	1,10	16,47	16,99
C10	2,40	85,20	214,00	3,30	16240,00	10,80	37,60	155,00	53,00	43,90	10,75	5,20	1,20	11,59	15,88
C11	1,80	101,50	230,00	3,40	13960,00	11,70	45,50	144,00	51,00	45,00	11,43	7,70	1,30	14,69	14,17
C12	1,40	95,20	245,00	3,50	11450,00	12,10	44,00	187,00	57,00	43,00	11,34	5,80	1,40	4,38	6,10
C13	2,00	67,50	205,00	2,80	10580,00	10,60	36,60	93,00	42,00	43,50	10,05	7,00	1,90	15,88	15,33
C14	1,20	63,40	183,00	2,85	9830,00	10,50	33,90	130,00	44,00	40,71	9,57	6,70	1,70	6,34	6,16
C15	3,00	65,30	212,00	2,85	12910,00	10,50	33,45	138,00	50,00	42,60	9,73	6,80	1,60	7,13	6,99
C16	2,70	67,80	236,00	3,02	11000,00	11,00	33,80	139,00	55,00	41,16	9,99	6,30	1,40	4,41	5,30
C17	1,90	52,50	195,00	2,68	9240,00	11,60	26,20	136,00	43,00	43,20	8,46	7,30	1,40	9,80	9,89
C18	1,90	109,70	218,00	3,29	7580,00	11,70	37,80	181,00	60,00	39,36	10,31	5,20	1,10	9,93	15,06
C19	1,70	87,80	241,00	3,30	11390,00	10,90	40,80	200,00	71,00	45,60	11,02	7,00	1,30	15,03	14,87
C20	1,60	78,90	230,00	3,20	8600,00	10,90	41,10	189,00	69,00	40,01	11,33	5,10	1,10	15,80	15,34
C21	1,20	68,90	204,00	2,92	8430,00	11,40	32,80	191,00	51,00	44,30	10,15	6,80	1,60	13,82	13,49
C22	2,50	54,50	203,00	2,95	12590,00	10,90	34,10	138,00	44,00	44,70	9,61	6,79	1,20	9,51	10,56
C23	1,10	245,00	162,00	4,70	7060,00	11,60	53,50	172,00	72,00	48,40	15,55	5,30	0,99	25,44	
C24	1,40	70,90	192,00	3,00	8320,00	11,90	36,10	117,00	61,00	46,10	10,31	7,40	1,50	10,65	10,30
C25	1,40	93,20	243,00	3,44	12570,00	12,80	40,90	189,00	54,00	45,80	11,41	6,10	1,20	6,48	6,55
C26	1,60	110,00	260,00	3,80	12630,00	12,30	48,30	159,00	59,00	44,10	13,23	5,80	0,90	13,00	13,18
C27	1,70	95,20	204,00	3,42	13460,00	12,50	43,50	192,00	48,00	50,10	11,05	6,75	1,20	12,72	13,21
C28	1,30	89,20	249,00	3,40	14460,00	12,30	39,50	165,00	62,00	48,90	11,10	5,70	1,40	12,10	11,74
C29	2,40	123,20	224,00	4,31	9160,00	12,80	51,50	176,00	58,00	47,80	14,04	7,40	1,60	17,61	19,31
C30	1,80	97,50	238,00	3,27	8030,00	11,90	38,00	167,00	52,00	42,30	10,36	6,20	1,80	9,85	9,52
C31	1,80	92,70	253,00	3,60	14940,00	12,80	44,20	125,00	63,00	48,30	11,70	6,40	1,20	9,76	10,10
													D _{crítico}	22,82	22,47

Para o conjunto das 81 amostras do sítio Água Limpa, analisadas com o valor crítico de *Wilks* calculado no nível de significância de 0,05, duas amostras (indicadas em negrito na Tabela 5.2) foram consideradas como *outliers*. E para as 31 amostras do sítio Rezende, somente uma amostra (indicadas em negrito na Tabela 5.3) obteve valor da distância de *Mahalanobis* maior que o valor crítico de

Wilks, sendo assim, classificada como *outlier*. Dessa forma, essas amostras foram removidas da base de dados.

5.4 Análise de agrupamentos

Após a eliminação das amostras discrepantes (*outliers*), os 143 resultados das amostras de cerâmicas foram submetidos à análise de agrupamentos.

Para cada matriz de distância, foram utilizados todos os métodos de análise de agrupamentos. Gerando, dessa forma, 20 dendrogramas, apresentados a seguir nas Figuras 5.2, 5.5, 5.8, 5.12, 5.14, 5.17, 5.20, 5.24, 5.26, 5.29, 5.32, 5.36, 5.38, 5.41, 5.44, 5.47, 5.50, 5.53, 5.56, 5.59.

5.4.1 Determinação do número de grupos

Para cada método associado a uma distância, foi feito um gráfico da ligação em função da distância (nível de fusão) do agrupamento de cada etapa do processo, com o objetivo de localizar se há "pontos de salto" relativamente grandes em relação aos demais valores de distância. Esses pontos indicam o momento ideal de parada do algoritmo, isto é, o número final de grupos e a composição desses grupos. Na Figura 5.1, pode-se observar um maior salto entre as distâncias 0,35 e 0,59, então, um corte adequado a ser realizado no dendrograma, para determinar o número de grupos, é na distância 0,35. Feito o corte nesta distância, formaram-se dessa forma, três grupos, conforme se observa no dendrograma da Figura 5.2. O valor da distância, utilizado para fazer o corte no dendrograma, é o valor do extremo mínimo do intervalo do salto observado, no gráfico da ligação em função da distância (Figura 5.1).

Pistgucia 0,35 – 0,59 0,00 40 60 80 100 120 140 Ligação

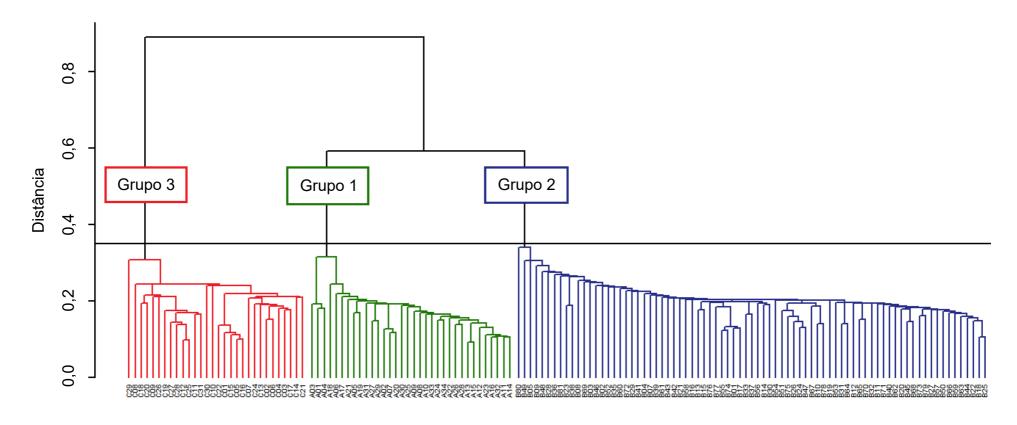
Figura 5.1 - Gráfico da ligação em função da distância do agrupamento das amostras de cerâmica, utilizando o método de ligação simples e com base na distância Euclidiana.

Os grupos formados foram chamados de Grupo 1 representado pela cor verde, Grupo 2 representado pela cor azul e Grupo 3 representado pela cor vermelha, sendo o Grupo 1 formado por 34 amostras, o Grupo 2 formado por 79 amostras e Grupo 3 formado por 30 amostras, conforme apresentado na Tabela 5.4.

Tabela 5.4 - Agrupamentos gerados pelo método de ligação simples com base na distância Euclidiana.

Grupo	Amostras
1	A03, A01, A04, A18, A06, A17, A21, A05, A19, A31, A27, A29, A02, A07,
	A20, A30, A25, A09, A08, A10, A33, A24, A34, A22, A26, A28, A13, A15,
	A12, A23, A16, A32, A11, A14.
2	B80, B49, B05, B09, B48, B28, B36, B81, B23, B38, B08, B69, B03, B46,
	B02, B52, B35, B60, B72, B29, B41, B04, B07, B39, B61, B43, B42, B21,
	B58, B16, B13, B15, B76, B77, B55, B74, B01, B17, B33, B37, B56, B14,
۷	B30, B54, B51, B75, B26, B24, B47, B67, B10, B78, B19, B53, B31, B64,
	B12, B65, B70, B32, B11, B71, B40, B62, B20, B45, B68, B73, B79, B27,
	B57, B50, B66, B59, B63, B44, B22, B18, B25.
3	C29, C08, C18, C20, C09, C26, C19, C27, C28, C12, C25, C11, C31, C30,
	C10, C22, C01, C15, C05, C16, C07, C24, C13, C02, C06, C04, C03, C17,
	C14, C21.

Figura 5.2 - Dendrograma do agrupamento das amostras de cerâmica, utilizando o método de ligação simples e com base na distância Euclidiana.



Lednépucia entre todos os judices os propositions os superiores de la companya del companya del companya de la companya del companya de la companya de la companya del companya de la comp

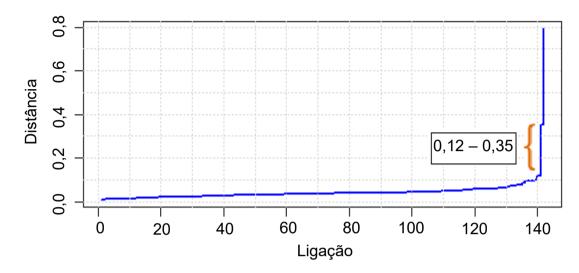
Figura 5.3 - Resultado do pacote *NbClust* para o método de ligação simples com base na distância Euclidiana.

Na Figura 5.3 é mostrado o gráfico extraído após a realização da inspeção do número ideal de grupos utilizando o pacote *NbClust*. O gráfico relaciona a frequência entre os múltiplos índices em função de um determinado número de grupo. Os índices utilizados estão descritos no seção 4.3.4.

Ao observar a Figura 5.3, nota-se que a maior frequência entre os índices calculados está para a definição de três grupos, indicando que este seria o número de grupos ideal para a partição do conjunto de dados analisados, corroborando, assim, com a análise do comportamento do nível de fusão (distância).

Na Figura 5.4, pode-se observar um maior salto entre as distâncias 0,12 e 0,35, então, um corte adequado a ser realizado no dendrograma, para determinar o número de grupos, é na distância 0,12. Feito o corte nesta distância, formaram-se, dessa forma, três grupos, conforme se observa no dendrograma da Figura 5.5.

Figura 5.4 - Gráfico da ligação em função da distância do agrupamento das amostras de cerâmica, utilizando o método de ligação simples e com base na distância Euclidiana ao quadrado.

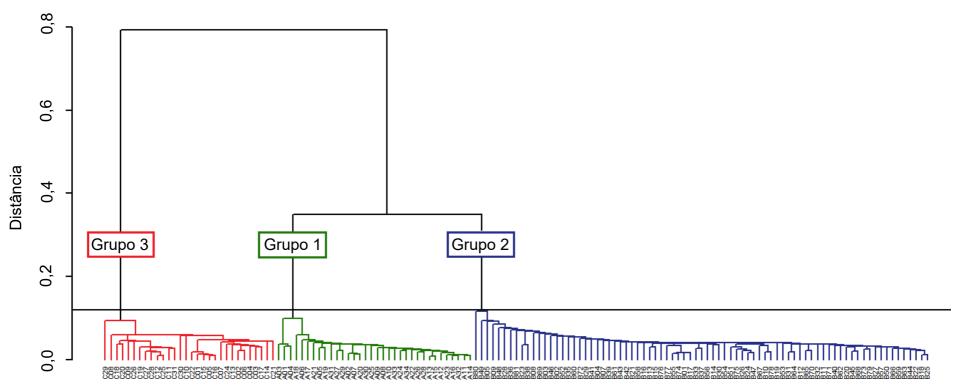


Os grupos formados foram denominados de Grupo 1 formado por 34 amostras, Grupo 2 formado por 79 amostras e Grupo 3 formado por 30 amostras, conforme é apresentado na Tabela 5.5.

Tabela 5.5 - Agrupamentos gerados pelo método de ligação simples com base na distância Euclidiana ao quadrado.

Grupo	Amostras											
1	A03, A01, A04, A18, A06, A17, A21, A05, A19, A31, A27, A29, A02, A07,											
	A20, A30, A25, A09, A08, A10, A33, A24, A34, A22, A26, A28, A13, A15,											
	A12, A23, A16, A32, A11, A14.											
2	B80, B49, B05, B09, B48, B28, B36, B81, B23, B38, B08, B69, B03, B46,											
	B02, B52, B35, B60, B72, B29, B41, B04, B07, B39, B61, B43, B42, B21,											
	B58, B16, B13, B15, B76, B77, B55, B74, B01, B17, B33, B37, B56, B14,											
2	B30, B54, B51, B75, B26, B24, B47, B67, B10, B78, B19, B53, B31, B64,											
	B12, B65, B70, B32, B11, B71, B40, B62, B20, B45, B68, B73, B79, B27,											
	B57, B50, B66, B59, B63, B44, B22, B18, B25.											
	C29, C08, C18, C20, C09, C26, C19, C27, C28, C12, C25, C11, C31, C30,											
3	C10, C22, C01, C15, C05, C16, C07, C24, C13, C02, C06, C04, C03, C17,											
	C14, C21.											

Figura 5.5 - Dendrograma do agrupamento das amostras de cerâmica, utilizando o método de ligação simples e com base na distância Euclidiana ao quadrado.



Lednêucia entre todos os fudices of fudices

Figura 5.6 - Resultado do pacote *NbClust* para o método de ligação simples com base na distância Euclidiana ao quadrado.

Ao observar a Figura 5.6, nota-se que a maior frequência entre os índices calculados está na definição de três grupos, confirmando o resultado obtido na Figura 5.4.

Na Figura 5.7, pode-se observar um maior salto entre as distâncias 0,97 e 1,59. Sendo assim, foi feito o corte na distância 0,97, formando, dessa forma, três grupos, conforme se observa no dendrograma da Figura 5.8.

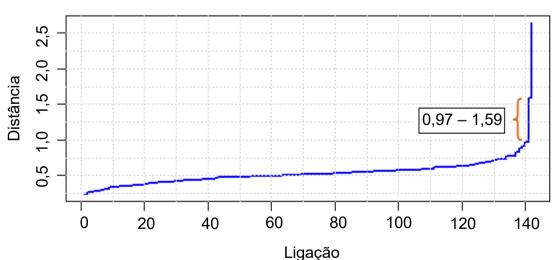
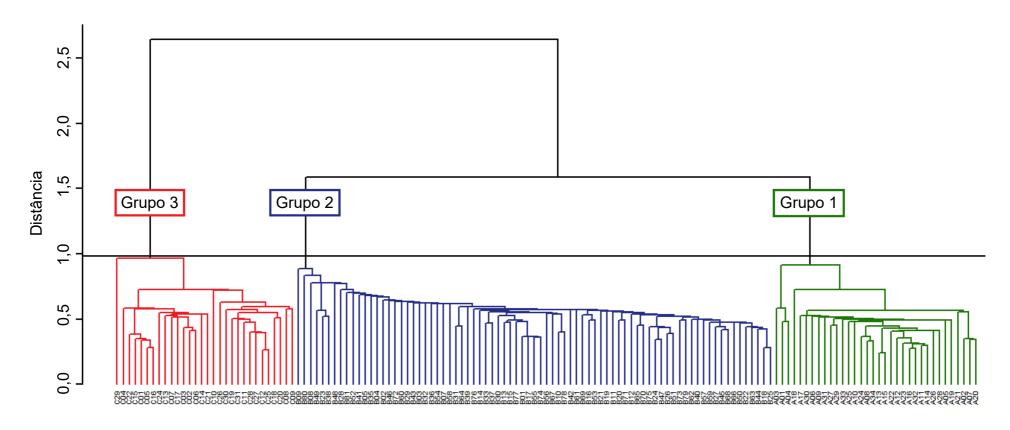


Figura 5.7 - Gráfico da ligação em função da distância do agrupamento das amostras de cerâmica, utilizando o método de ligação simples e com base na distância *Manhattan*.

Figura 5.8 - Dendrograma do agrupamento das amostras de cerâmica, utilizando o método de ligação simples e com base na distância *Manhattan*.



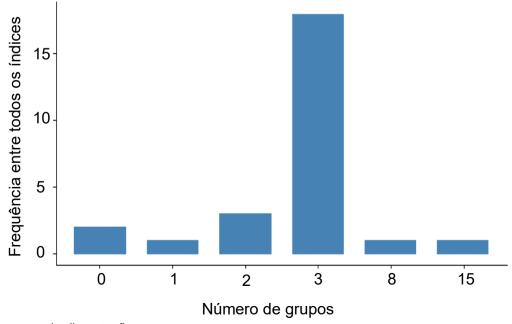
Os grupos formados foram denominados de Grupo 1, formado por 34 amostras, Grupo 2, formado por 79 amostras e Grupo 3, formado por 30 amostras, conforme é apresentado na Tabela 5.6.

Tabela 5.6 - Agrupamentos gerados pelo método de ligação simples com base na distância *Manhattan*.

Grupo	Amostras										
1	A03, A01, A04, A18, A17, A30, A06, A09, A31, A27, A29, A33, A25, A10,										
	A24, A08, A34, A13, A15, A22, A12, A23, A16, A32, A11, A14, A26, A28,										
	A05, A19, A21, A02, A07, A20.										
-	B09, B80, B08, B49, B23, B38, B48, B28, B81, B52, B41, B05, B35, B04,										
	B02, B46, B72, B60, B29, B43, B03, B32, B36, B54, B07, B58, B31, B64,										
	B39, B76, B14, B33, B37, B30, B13, B15, B77, B01, B17, B55, B74, B56,										
2	B67, B10, B78, B42, B61, B69, B16, B53, B21, B19, B11, B20, B71, B12,										
	B65, B70, B75, B24, B47, B26, B51, B73, B79, B62, B40, B57, B59, B27,										
	B45, B68, B66, B50, B22, B63, B44, B18, B25.										
3	C29, C04, C22, C15, C01, C05, C16, C24, C13, C07, C17, C03, C02, C06,										
	C14, C21, C10, C26, C30, C19, C31, C11, C28, C27, C12, C25, C18, C20,										
	C08, C09.										

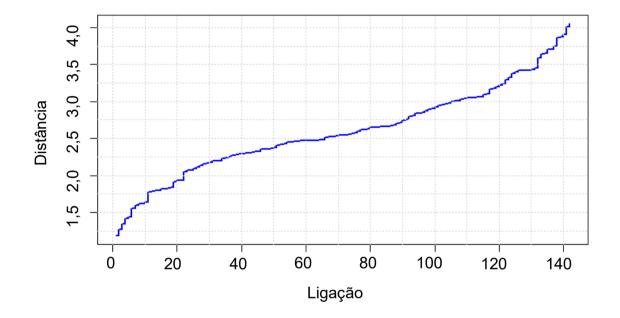
Fonte: autora da dissertação.

Figura 5.9 – Resultado do pacote *NbClust* para o método de ligação simples com base na distância *Manhattan*.



Ao observar a Figura 5.9, nota-se que a maior frequência entre os índices calculados está na definição de três grupos, corroborando, assim, com o resultado da Figura 5.7.

Figura 5.10 - Gráfico da ligação em função da distância do agrupamento das amostras de cerâmica, utilizando o método de ligação simples e com base na distância *Mahalanobis*.



Fonte: autora da dissertação.

Na Figura 5.10, fica difícil analisar entre quais distâncias houve um maior salto. Neste caso, utilizou-se o gráfico gerado pela análise do *NbClust* (Figura 5.11), a fim de determinar o número ideal de grupos gerados pelo agrupamento.

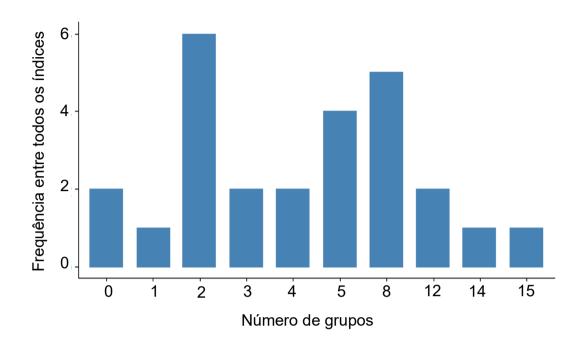


Figura 5.11 - Resultado do pacote *NbClust* para o método de ligação simples com base na distância *Mahalanobis*.

A partir da análise da Figura 5.11, o número ideal de grupos seria dois, pois a maioria dos índices comparados no pacote *NbClust* propõe dois como o melhor número de grupos. Segundo CHARRAD et al. (2014), o uso do pacote *NbClust* é uma das opções de lidar com o problema de determinação do número ideal de grupos gerados pela análise de agrupamentos.

Sendo assim, foi feito um corte no dendrograma (Figura 5.12), de modo que gerasse dois grupos. Os grupos formados foram denominados de Grupo 1, formado por 1 amostra e Grupo 2, formado por 142 amostras, descritos na Tabela 5.7.

Figura 5.12 - Dendrograma do agrupamento das amostras de cerâmica, utilizando o método de ligação simples e com base na distância *Mahalanobis*.

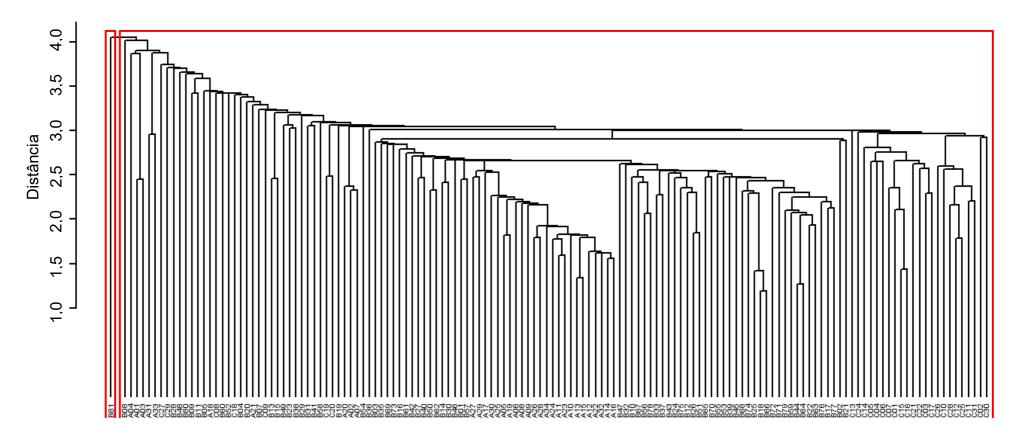


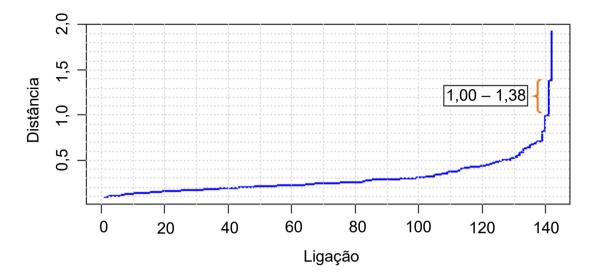
Tabela 5.7 - Agrupamentos gerados pelo método de ligação simples com base na distância *Mahalanobis*.

Grupo	Amostras
1	B81.
	B08, A04, A01, A03, A31, A33, C27, C29, B28, B48, B80, B09, B11, B05,
	A18, C08, B60, B52, C18, B04, B20, A21, B07, C09, B13, B15, B49, B23,
	B38, B29, B31, B41, B58, C19, C20, B19, A20, A02, A07, B54, B36, B03,
	B35, B69, B73, B16, B61, B42, B27, B40, B50, B62, B14, B30, B46, B01,
	B57, A27, A29, A17, A30, A25, A05, A19, A06, A08, A09, A26, A28, A34,
2	A24, A11, A23, A10, A13, A15, A12, A22, A32, A14, A16, B47, B32, B10,
	B67, B56, B78, B33, B37, B43, B24, B75, B12, B26, B51, B65, B70, B55,
	B53, B39, B45, B68, B74, B25, B18, B66, B72, B71, B79, B59, B44, B64,
	B22, B63, B76, B17, B77, B02, B21, C13, C24, C14, C05, C04, C06, C07,
	C01, C15, C16, C21, C22, C03, C17, C26, C10, C28, C12, C25, C11, C31,
	C02, C30.

A partir da análise dos dendrogramas do agrupamento, utilizando o método de ligação simples com base nas distâncias Euclidiana, Euclidiana ao quadrado e *Manhattan* (Figuras 5.2, 5.5, 5.8) percebe-se a formação de três grupos bem definidos. Além disso, analisando as Tabelas 5.4, 5.5 e 5.6, os grupos formados são os mesmos e compostos por amostras de mesmo sítio arqueológico.

Todavia, o mesmo não acontece com o método associado à distância de *Mahalanobis*, no qual, fica impossível a determinação de grupos apenas com a inspeção visual do dendrograma. Para este caso, utilizou-se a análise do pacote *NbClust* gerando, dessa forma, dois grupos: um formado por uma única amostra e o outro formado por amostras de sítios arqueológicos diferentes.

Figura 5.13 - Gráfico da ligação em função da distância do agrupamento das amostras de cerâmica, utilizando o método de ligação completa e com base na distância Euclidiana.



Na Figura 5.13, pode-se observar um maior salto entre as distâncias 1,00 e 1,38, então, um corte adequado a ser realizado no dendrograma, para determinar o número de grupos, é na distância 1,00. Feito o corte nesta distância, formaram-se, dessa forma, três grupos, conforme se observa no dendrograma da Figura 5.14.

Os grupos formados foram chamados de Grupo 1 representado pela cor verde, Grupo 2 representado pela cor azul e Grupo 3 representado pela cor vermelha, sendo o Grupo 1 formado por 34 amostras, o Grupo 2 formado por 79 amostras e Grupo 3 formado por 30 amostras, conforme se observa na Tabela 5.8.

Figura 5.14 - Dendrograma do agrupamento das amostras de cerâmica, utilizando o método de ligação completa e com base na distância Euclidiana.

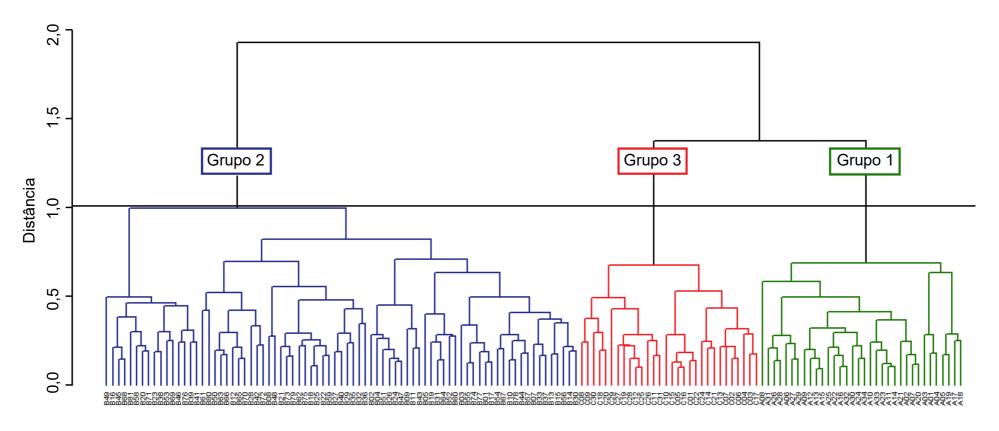
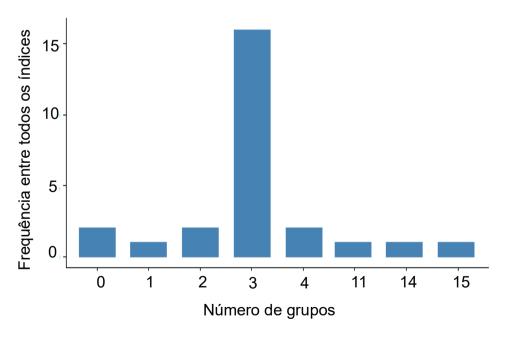


Tabela 5.8 - Agrupamentos gerados pelo método de ligação completa com base na distância Euclidiana.

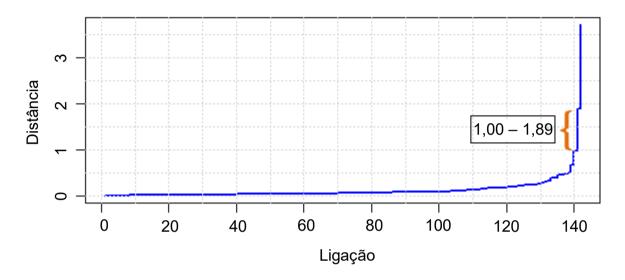
Grupo	Amostras
1	A08, A31, A26, A28, A06, A27, A29, A09, A12, A13, A15, A25, A22, A16,
	A32, A30, A24, A34, A10, A33, A23, A11, A14, A21, A02, A07, A20, A03,
	A01, A04, A05, A19, A17, A18.
2	B49, B16, B45, B68, B81, B58, B20, B71, B23, B38, B53, B69, B46, B76,
	B39, B41, B61, B80, B50, B63, B66, B12, B65, B70, B28, B42, B72, B08,
	B48, B21, B73, B79, B62, B75, B18, B25, B22, B59, B27, B40, B29, B35,
	B32, B36, B02, B04, B51, B26, B24, B47, B09, B11, B43, B05, B19, B31,
	B64, B52, B60, B03, B55, B74, B77, B01, B17, B54, B67, B10, B78, B44,
	B57, B07, B33, B37, B13, B15, B56, B14, B30.
3	C08, C09, C30, C18, C20, C29, C27, C19, C28, C12, C25, C26, C11, C31,
	C10, C15, C05, C16, C01, C22, C24, C14, C21, C13, C07, C02, C06, C04,
	C03, C17.

Figura 5.15 – Resultado do pacote *NbClust* para o método de ligação completa com base na distância Euclidiana.



Ao observar a Figura 5.15, nota-se que a maior frequência entre os índices calculados está na definição de três grupos, confirmando o resultado da Figura 5.13.

Figura 5.16 - Gráfico da ligação em função da distância do agrupamento das amostras de cerâmica, utilizando o método de ligação completa e com base na distância Euclidiana ao quadrado.



Fonte: autora da dissertação.

Na Figura 5.16, pode-se observar um maior salto entre as distâncias 1,00 e 1,89, portanto, um corte adequado a ser realizado no dendrograma, para determinar o número de grupos, é na distância 1,89. Feito o corte nesta distância, formaram-se, dessa forma, três grupos, conforme se observa no dendrograma da Figura 5.17.

Os grupos formados foram chamados de Grupo 1 representado pela cor verde, Grupo 2 representado pela cor azul e Grupo 3 representado pela cor vermelha, sendo o Grupo 1 formado por 34 amostras, o Grupo 2 formado por 79 amostras e Grupo 3 formado por 30 amostras, conforme Tabela 5.9. Os grupos formados são compostos por amostras do mesmo sítio arqueológico.

Figura 5.17 - Dendrograma do agrupamento das amostras de cerâmica, utilizando o método de ligação completa e com base na distância Euclidiana ao quadrado.

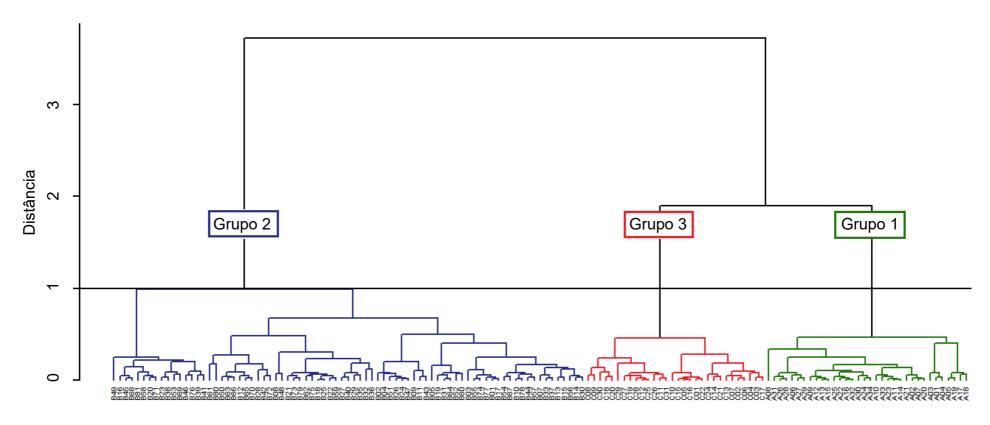
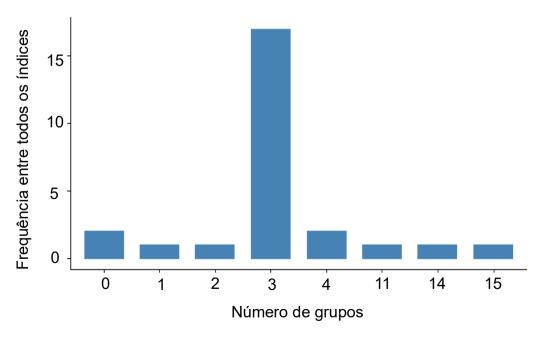


Tabela 5.9 - Agrupamentos gerados pelo método de ligação simples com base na distância Euclidiana ao quadrado.

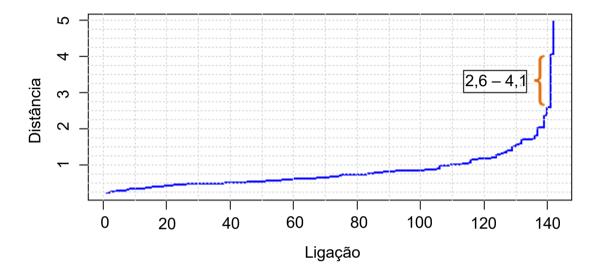
Grupo	Amostras
1	A08, A31, A26, A28, A06, A27, A29, A09, A12, A13, A15, A25, A22, A16,
	A32, A30, A24, A34, A10, A33, A23, A11, A14, A21, A02, A07, A20, A03,
	A01, A04, A05, A19, A17, A18.
	B49, B16, B45, B68, B81, B58, B20, B71, B23, B38, B53, B69, B46, B76,
2	B39, B41, B61, B80, B50, B63, B66, B12, B65, B70, B28, B42, B72, B08,
	B48, B21, B73, B79, B62, B75, B18, B25, B22, B59, B27, B40, B29, B35,
	B32, B36, B02, B04, B51, B26, B24, B47, B09, B11, B43, B05, B19, B31,
	B64, B52, B60, B03, B55, B74, B77, B01, B17, B54, B67, B10, B78, B44,
	B57, B07, B33, B37, B13, B15, B56, B14, B30.
3	C08, C09, C30, C18, C20, C29, C27, C19, C28, C12, C25, C26, C11, C31,
	C10, C15, C05, C16, C01, C22, C24, C14, C21, C13, C07, C02, C06, C04,
	C03, C17.

Figura 5.18 - Resultado do pacote *NbClust* para o método de ligação completo com base na distância Euclidiana ao quadrado.



Ao observar a Figura 5.18, nota-se que a maior frequência entre os índices calculados está na definição de três grupos, confirmando o resultado obtido com a análise do comportamento do nível de fusão (Figura 5.16).

Figura 5.19 - Gráfico da ligação em função da distância do agrupamento das amostras de cerâmica, utilizando o método de ligação completa e com base na distância *Manhattan*.



Fonte: autora da dissertação.

Na Figura 5.19, pode-se observar um maior salto entre as distâncias 2,6 e 4,1, portanto, um corte adequado a ser realizado no dendrograma, para determinar o número de grupos, é na distância 2,6. Feito o corte nesta distância, formaram-se, dessa forma, três grupos, conforme se observa no dendrograma da Figura 5.20.

Os grupos formados foram chamados de Grupo 1 representado pela cor verde, Grupo 2 representado pela cor azul e Grupo 3 representado pela cor vermelha, sendo o Grupo 1 formado por 34 amostras, o Grupo 2 formado por 79 amostras e Grupo 3 formado por 30 amostras, conforme se observa na Tabela 5.10.

Figura 5.20 - Dendrograma do agrupamento das amostras de cerâmica, utilizando o método de ligação completa e com base na distância *Manhattan*.

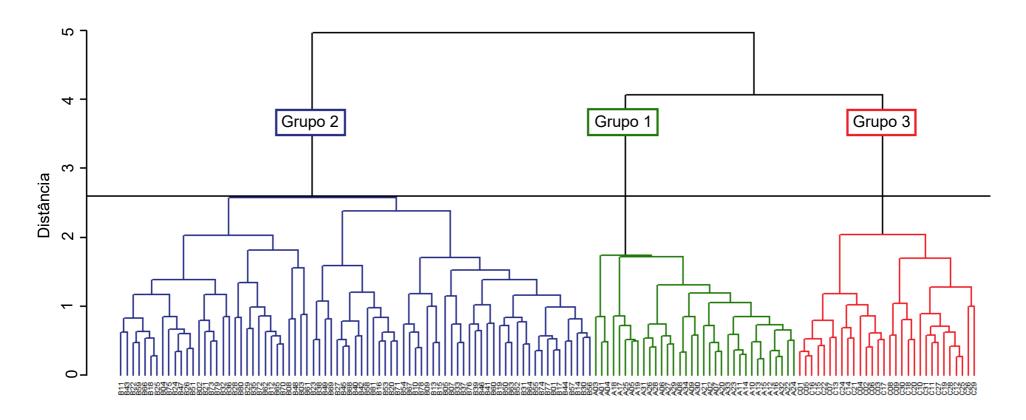
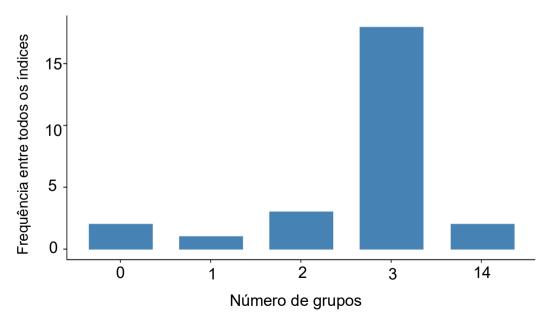


Tabela 5.10 - Agrupamentos gerados pelo método de ligação completa com base na distância *Manhattan*.

Grupo	Amostras
1	A03, A01, A04, A18, A17, A25, A05, A19, A31, A26, A28, A06, A27, A29,
	A08, A34, A09, A30, A21, A02, A07, A20, A33, A23, A11, A14, A10, A13,
	A15, A12, A16, A32, A22, A24.
2	B49, B16, B45, B68, B81, B58, B20, B71, B23, B38, B53, B69, B46, B76,
	B39, B41, B61, B80, B50, B63, B66, B12, B65, B70, B28, B42, B72, B08,
	B48, B21, B73, B79, B62, B75, B18, B25, B22, B59, B27, B40, B29, B35,
	B32, B36, B02, B04, B51, B26, B24, B47, B09, B11, B43, B05, B19, B31,
	B64, B52, B60, B03, B55, B74, B77, B01, B17, B54, B67, B10, B78, B44,
	B57, B07, B33, B37, B13, B15, B56, B14, B30.
3	C01, C05, C16, C15, C22, C07, C13, C24, C14, C21, C04, C02, C06, C03,
	C17, C08, C09, C30, C18, C20, C10, C31, C11, C27, C19, C28, C12, C25,
	C26, C29.

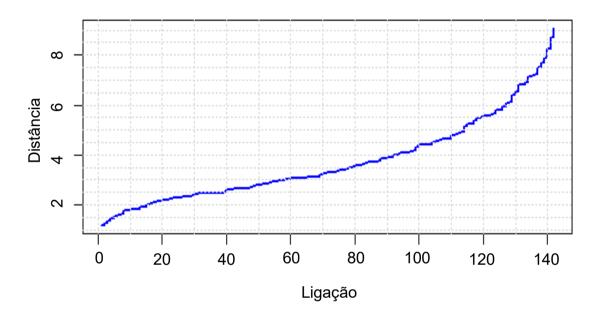
Com o dendrograma da Figura 5.20 e a Tabela 5.8, observa-se que os grupos formados são compostos por amostra de mesmo sítio arqueológico.

Figura 5.21 – Resultado do pacote *NbClust* para o método de ligação completa com base na distância *Manhattan*.



Ao observar a Figura 5.21, nota-se que a maior frequência entre os índices calculados está na definição de três grupos, confirmando o resultado da Figura 5.19.

Figura 5.22 - Gráfico da ligação em função da distância do agrupamento das amostras de cerâmica, utilizando o método de ligação completa e com base na distância *Mahalanobis*.



Fonte: autora da dissertação.

Na Figura 5.22, fica difícil de analisar entre quais distâncias houve um maior salto. Neste caso, utilizou-se o gráfico gerado pela análise do *NbClust* (Figura 5.23) para determinar o número ideal de grupos gerados pelo agrupamento.

A partir da análise da Figura 5.23, observa-se que a maioria dos índices comparados no pacote *NbClust*, propõe dois como o melhor número de grupos. Sendo assim, foi feito um corte no dendrograma (Figura 5.24) de modo a gerar dois grupos. Os grupos formados foram denominados de Grupo 1 constituído por 3 amostras e Grupo 2 constituído por 140 amostras, descritos na Tabela 5.11.

Figura 5.23 - Resultado do pacote *NbClust* para o método de ligação completa com base na distância *Mahalanobis*.

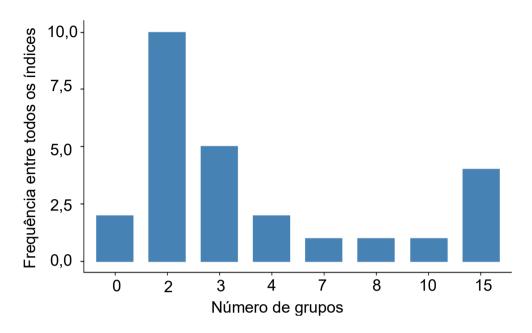
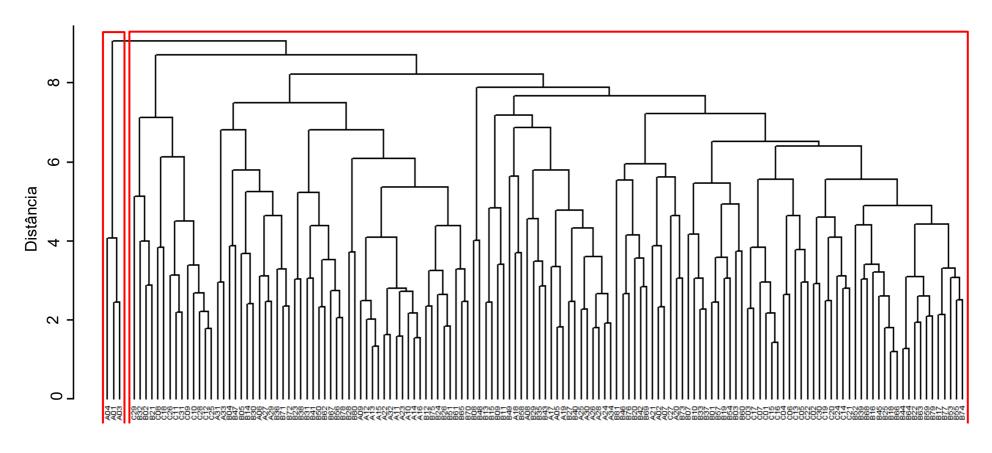


Tabela 5.11 - Agrupamentos gerados pelo método de ligação completa com base na distância *Mahalanobis*.

Grupo	Amostras
1	A04, A01, A03.
	C29, B32, B02, B21, C08, C18, C26, C11, C31, C09, C10, C28, C12, C25,
	A31, A33, B04, B47, B05, B14, B30, A06, A27, A29, B36, B71, B72, B23,
	B38, B31, B41, B50, B62, B67, B56, B78, B28, B80, A09, A12, A13, A15,
	A22, A32, A11, A23, A10, A14, A16, B12, B75, B24, B26, B51, B61, B65,
2	B70, B08, B48, B13, B15, B09, B11, B49, A18, B58, A08, B29, B35, B43,
2	A17, A05, A19, B27, B40, A25, A30, A26, A28, A24, A34, B81, B46, B76,
	B20, B42, B69, A21, A02, A07, C27, A20, B73, B07, B10, B33, B37, B01,
	B57, B19, B54, B03, B60, C03, C17, C07, C01, C15, C16, C04, C06, C13,
	C05, C22, C02, C30, C19, C20, C24, C14, C21, B52, B39, B68, B16, B45,
	B25, B18, B66, B44, B64, B22, B63, B59, B79, B17, B77, B53, B55, B74.

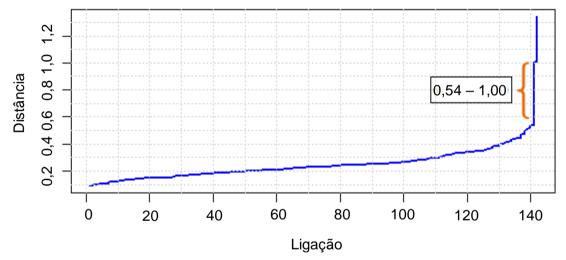
Figura 5.24 - Dendrograma do agrupamento das amostras de cerâmica, utilizando o método de ligação completa e com base na distância *Mahalanobis*.



A partir da análise dos dendrogramas do agrupamento utilizando o método de ligação completa com base nas distâncias Euclidiana, Euclidiana ao quadrado e *Manhattan* (Figuras 5.14, 5.17, 5.20) percebe-se a formação de três grupos bem definidos. Além disso, os grupos formados são os mesmos e compostos por amostras de mesmo sítio arqueológico.

Entretanto, o mesmo não acontece com o método associado à distância de *Mahalanobis* (Figura 5.24), no qual, formaram-se dois grupos: um formado por três amostras do mesmo sítio e o outro formado por 140 amostras de sítios arqueológicos diferentes.

Figura 5.25 - Gráfico da ligação em função da distância do agrupamento das amostras de cerâmica, utilizando o método de ligação média e com base na distância Euclidiana.



Fonte: autora da dissertação.

Na Figura 5.25, pode-se observar um maior salto entre as distâncias 0,54 e 1,00. Sendo assim, foi feito o corte no dendrograma na distância 0,54, formando-se, dessa forma, três grupos (Figura 5.26).

Os grupos formados foram denominados de Grupo 1 constituído por 34 amostras, Grupo 2 constituído por 79 amostras e Grupo 3 constituído por 30 amostras, conforme observa-se na Tabela 5.12.

Figura 5.26 - Dendrograma do agrupamento das amostras de cerâmica, utilizando o método de ligação média e com base na distância Euclidiana.

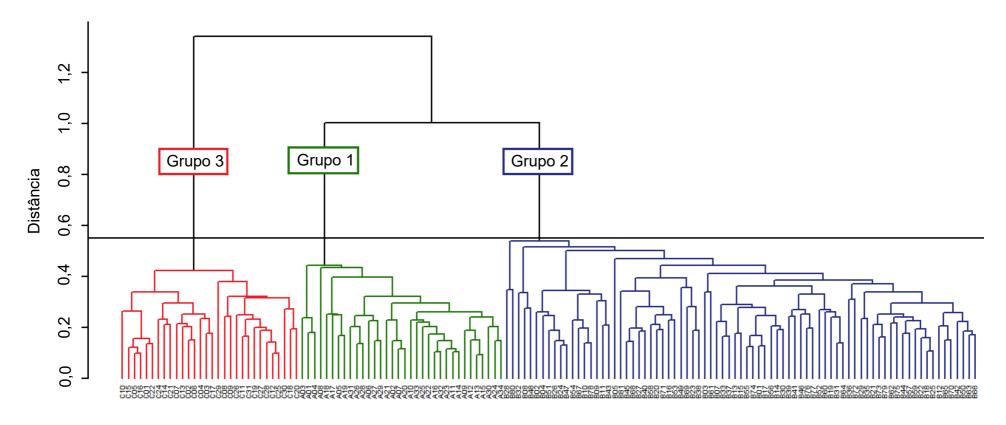
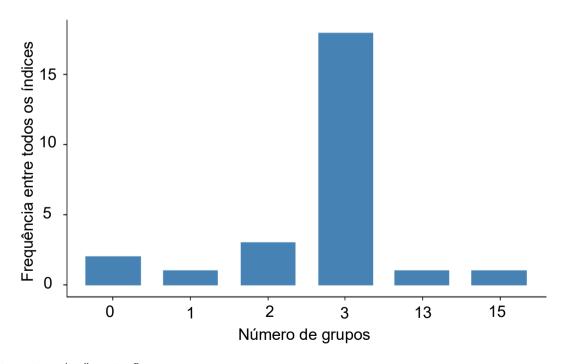


Tabela 5.12 - Agrupamentos gerados pelo método de ligação média com base na distância Euclidiana.

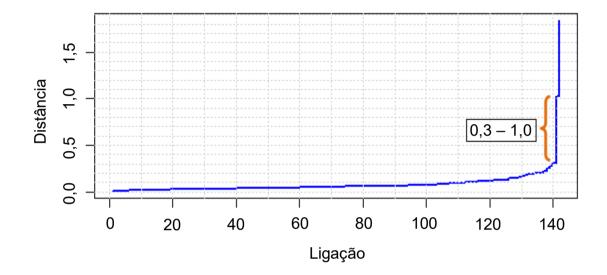
Grupo	Amostras
1	A03, A01, A04, A18, A18, A17, A05, A19, A31, A26, A28, A06, A27, A29,
	A21, A02, A07, A20, A10, A33, A25, A22, A16, A32, A23, A11, A14, A09,
	A12, A13, A15, A30, A24, A34.
	B28, B80, B32, B08, B48, B02, B04, B51, B26, B24, B47, B54, B67, B10,
2	B78, B09, B11, B43, B05, B81, B45, B68, B27, B40, B58, B20, B71, B16,
	B53, B49, B69, B23, B38, B03, B61, B07, B33, B37, B13, B15, B55, B74,
	B01, B17, B56, B14, B30, B39, B41, B46, B76, B77, B52, B60, B19, B31,
	B64, B36, B72, B29, B35, B21, B73, B79, B62, B75, B44, B57, B59, B22,
	B18, B25, B12, B65, B70, B42, B50, B63, B66.
3	C10, C15, C05, C16, C01, C22, C24, C14, C21, C07, C13, C02, C06, C04,
	C03, C17, C29, C08, C09, C26, C11, C31, C19, C27, C28, C12, C25, C30,
	C18, C20.

Figura 5.27 – Resultado do pacote *NbClust* para o método de ligação média com base na distância Euclidiana.



Ao observar a Figura 5.27, nota-se que a maior frequência entre os índices calculados está na definição de três grupos, confirmando o resultado obtido com a análise de comportamento do nível de fusão (Figura 5.25).

Figura 5.28 - Gráfico da ligação em função da distância do agrupamento das amostras de cerâmica, utilizando o método de ligação média e com base na distância Euclidiana ao quadrado.



Fonte: autora da dissertação.

Na Figura 5.28, pode-se observar um maior salto entre as distâncias 0,3 e 1,0. Sendo assim, foi feito o corte na distância 0,3, formando, dessa forma, três grupos bem definidos, conforme se observa no dendrograma da Figura 5.29.

Os grupos formados foram chamados de Grupo 1 representado pela cor verde, Grupo 2 representado pela cor azul e Grupo 3 representado pela cor vermelha, sendo o Grupo 1 formado por 34 amostras, o Grupo 2 formado por 79 amostras e Grupo 3 formado por 30 amostras, cada grupo é composto por amostra de mesmo sítio arqueológico, conforme Tabela 5.13.

Figura 5.29 - Dendrograma do agrupamento das amostras de cerâmica, utilizando o método de ligação média e com base na distância Euclidiana ao quadrado.

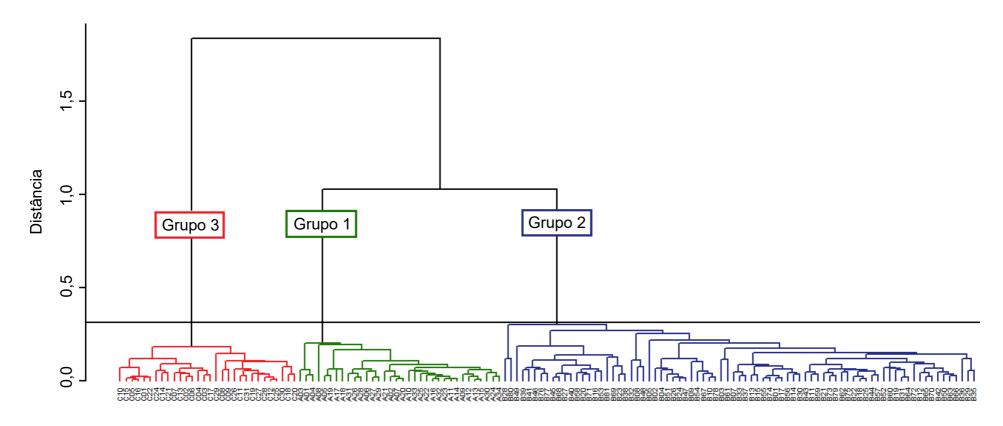
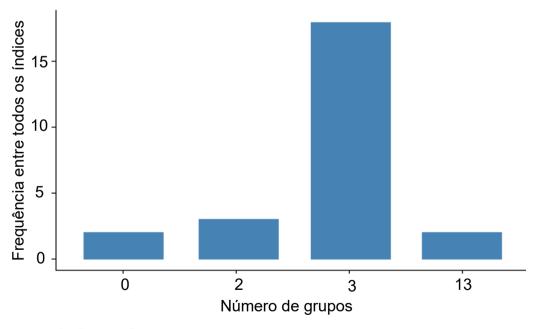


Tabela 5.13 - Agrupamentos gerados pelo método de ligação média com base na distância Euclidiana ao quadrado.

Grupo	Amostras
1	A03, A01, A04, A08, A05, A19, A17, A18, A31, A26, A28, A06, A27, A29,
	A21, A02, A07, A20, A10, A33, A25, A22, A16, A32, A23, A11, A14, A09,
	A12, A13, A15, A30, A24, A34.
	B28, B80, B49, B39, B41, B46, B76, B77, B45, B68, B27, B40, B58, B20,
2	B71, B16, B53, B81, B69, B23, B38, B32, B08, B48, B05, B02, B04, B51,
	B26, B24, B47, B09, B54, B67, B10, B78, B03, B61, B07, B33, B37, B13,
	B15, B55, B74, B01, B17, B56, B14, B30, B43, B11, B59, B21, B73, B79,
	B62, B75, B22, B18, B25, B44, B57, B52, B60, B19, B31, B64, B72, B12,
	B65, B70, B42, B50, B63, B66, B36, B29, B35.
3	C10, C15, C05, C16, C01, C22, C24, C14, C21, C07, C13, C02, C06, C04,
	C03, C17, C29, C08, C09, C26, C11, C31, C19, C27, C28, C12, C25, C30,
	C18, C20.

Figura 5.30 – Resultado do pacote *NbClust* para o método de ligação média com base na distância Euclidiana ao quadrado.



Fonte: autora da dissertação.

Ao observar a Figura 5.30, nota-se que a maior frequência entre os índices calculados está para a definição de três grupos, corroborando, assim, com o resultado da Figura 5.28.

Distância
0,5 1,0 1,5 2,0 2,5 3,0 3,5
1,0 1,5 2,0 2,5 3,0 3,5
1,0 1,5 2,0 2,5 3,0 3,5

60

40

80

Ligação

100

120

140

Figura 5.31 - Gráfico da ligação em função da distância do agrupamento das amostras de cerâmica, utilizando o método de ligação média e com base na distância *Manhattan*.

Fonte: autora da dissertação.

0

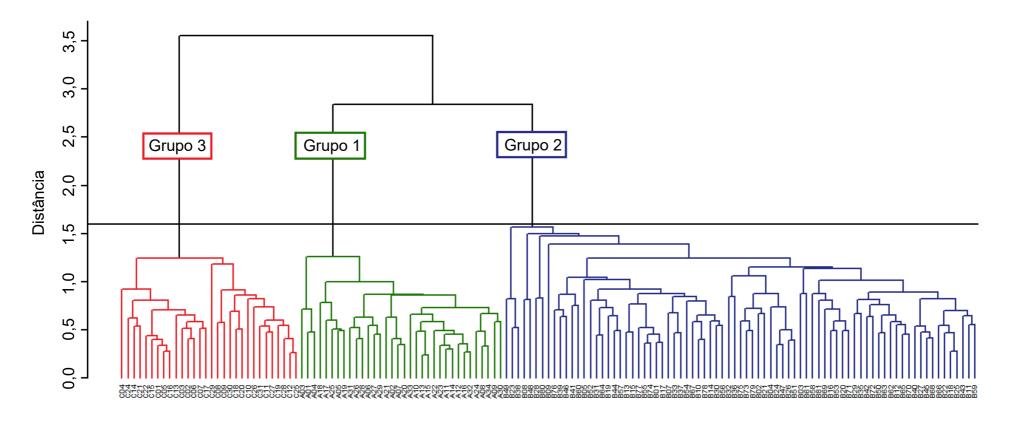
20

Na Figura 5.31, pode-se observar um maior salto entre as distâncias 1,60 e 2,84, logo, um corte adequado a ser realizado no dendrograma, para determinar o número de grupos, é na distância 1,60. Feito o corte nesta distância, formaram-se, dessa forma, três grupos, conforme se observa no dendrograma da Figura 5.32.

Tabela 5.14 - Agrupamentos gerados pelo método de ligação média com base na distância *Manhattan*.

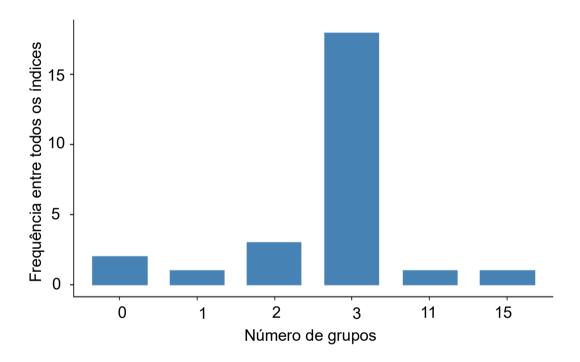
Grupo	Amostras
1	A03, A01, A04, A18, A17, A25, A05, A19, A31, A26, A28, A06, A27, A29,
	A21, A02, A07, A20, A33, A10, A13, A15, A22, A23, A11, A14, A12, A16,
	A32, A24, A08, A34, A09, A30.
	B49, B23, B38, B08, B48, B28, B80, B09, B76, B39, B46, B41, B60, B05,
2	B52, B31, B64, B19, B44, B57, B13, B15, B77, B55, B74, B01, B17, B07,
	B33, B37, B54, B67, B10, B78, B14, B30, B56, B32, B36, B75, B73, B79,
	B02, B21, B04, B24, B47, B26, B51, B03, B61, B58, B81, B69, B16, B53,
	B20, B71, B29, B35, B42, B72, B50, B63, B62, B12, B65, B70, B40, B27,
	B45, B68, B66, B22, B18, B25, B43, B11, B59.
3	C04, C24, C14, C21, C22, C15, C01, C05, C16, C13, C03, C02, C06, C07,
	C17, C29, C08, C09, C30, C18, C20, C10, C26, C31, C11, C27, C19, C28,
	C12, C25.

Figura 5.32 - Dendrograma do agrupamento das amostras de cerâmica, utilizando o método de ligação média e com base na distância *Manhattan*.



Os grupos formados foram denominados de Grupo 1 formado por 34 amostras, Grupo 2 formado por 79 amostras e Grupo 3 formado por 30 amostras. Os grupos são compostos por amostras de mesmo sítio arqueológico (Tabela 5.14).

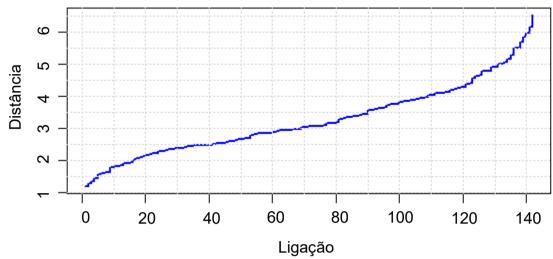
Figura 5.33 - Resultado do pacote *NbClust* para o método de ligação média com base na distância *Manhattan*.



Fonte: autora da dissertação.

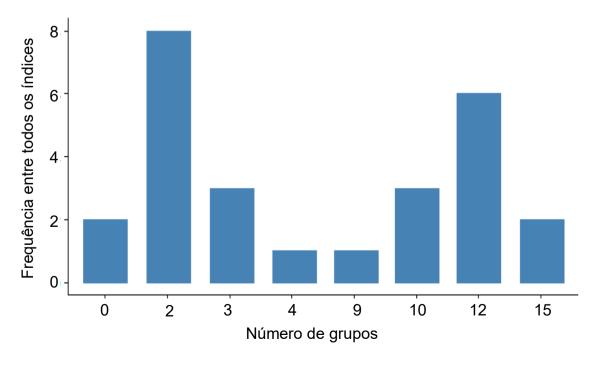
Ao observar a Figura 5.33, nota-se que a maior frequência entre os índices calculados está para a definição de três grupos, confirmando o resultado obtido pela análise do comportamento do nível de fusão (Figura 5.31).

Figura 5.34 - Gráfico da ligação em função da distância do agrupamento das amostras de cerâmica, utilizando o método de ligação média e com base na distância *Mahalanobis*.



Na Figura 5.34, fica difícil analisar entre quais distâncias houve um maior salto. Neste caso, utilizou-se o gráfico gerado pela análise do *NbClust* (Figura 5.35), a fim de determinar o número ideal de grupos gerados pelo agrupamento.

Figura 5.35 - Resultado do pacote *NbClust* para o método de ligação média com base na distância *Mahalanobis*.



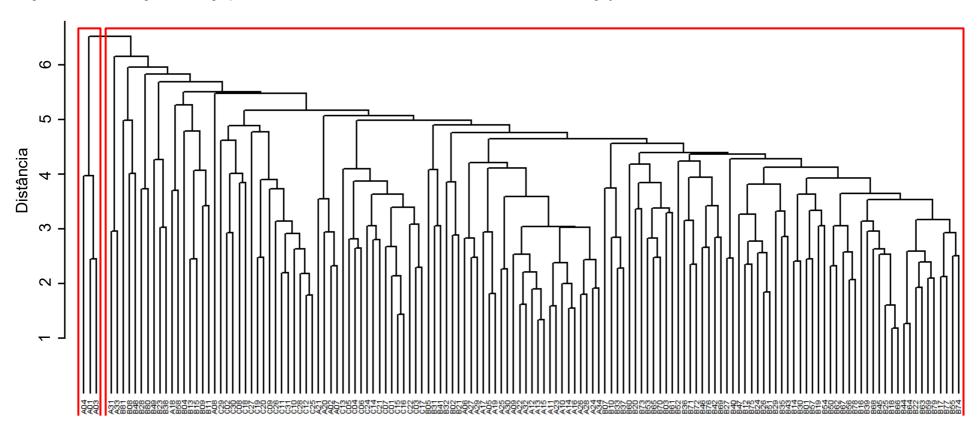
A partir da análise da Figura 5.35, observa-se que entre os índices comparados no pacote *NbClust*, a maioria dos índices propõe dois como o número ideal de grupos.

Sendo assim, foi feito um corte no dendrograma (Figura 5.36) de modo a gerar dois grupos. Os grupos formados foram denominados de Grupo 1 formado por 3 amostras e Grupo 2 formado por 140 amostras, descritos na Tabela 5.15.

Tabela 5.15 - Agrupamentos gerados pelo método de ligação média com base na distância *Mahalanobis*.

Grupo	Amostras
1	A04, A01, A03.
-	A31, A33, B81, B08, B48, B28, B80, B49, B23, B38, A18, B58, B04, B13,
	B15, B09, B11, A08, C29, C02, C30, C08, C18, C27, C19, C20, C09, C26,
2	C11, C31, C10, C28, C12, C25, A21, A20, A02, A07, C13, C05, C04, C06,
	C24, C14, C21, C07, C01, C15, C16, C22, C03, C17, B05, B31, B41, B32,
	B02, B21, A06, A27, A29, A17, A05, A19, A25, A30, A09, A22, A32, A12,
2	A13, A15, A11, A23, A10, A14, A16, A26, A28, A24, A34, B07, B10, B33,
	B37, B60, B20, B73, B53, B65, B70, B03, B61, B52, B36, B71, B72, B46,
	B76, B42, B69, B27, B40, B47, B12, B75, B24, B26, B51, B29, B35, B43,
	B14, B30, B01, B57, B19, B54, B50, B62, B67, B56, B78, B16, B39, B68,
	B45, B25, B18, B66, B44, B64, B22, B63, B59, B79, B17, B77, B55, B74.

Figura 5.36 - Dendrograma do agrupamento das amostras de cerâmica, utilizando o método de ligação média e com base na distância *Mahalanobis*.

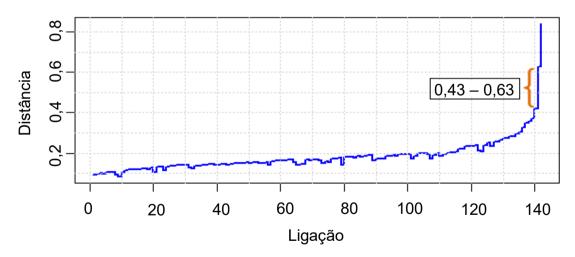


A partir da análise dos dendrogramas do agrupamento utilizando o método de ligação média com base nas distâncias Euclidiana, Euclidiana ao quadrado e *Manhattan* (Figuras 5.26, 5.29, 5.32) percebe-se a formação de três grupos bem definidos. Além disso, os grupos formados são os mesmos e compostos por amostras de mesmo sítio arqueológico.

Entretanto, o mesmo não acontece com o método associado à distância de *Mahalanobis* (Figura 5.36), no qual, formaram-se dois grupos: um formado por três amostras do mesmo sítio e o outro formado por 140 amostras de sítios arqueológicos diferentes.

Na Figura 5.37, pode-se observar um maior salto entre as distâncias 0,43 e 0,63, logo, um corte adequado a ser realizado no dendrograma, para determinar o número de grupos, é na distância 0,43. Feito o corte nesta distância, formaram-se, dessa forma, três grupos, conforme se observa no dendrograma da Figura 5.38.

Figura 5.37 - Gráfico da ligação em função da distância do agrupamento das amostras de cerâmica, utilizando o método do centróide e com base na distância Euclidiana.



Fonte: autora da dissertação.

Os grupos formados foram chamados de Grupo 1 representado pela cor verde, Grupo 2 representado pela cor azul e Grupo 3 representado pela cor vermelha, sendo o Grupo 1 formado por 34 amostras, o Grupo 2 formado por 79 amostras e Grupo 3 formado por 30 amostras, conforme a Tabela 5.16.

Figura 5.38 - Dendrograma do agrupamento das amostras de cerâmica, utilizando o método do centróide e com base na distância Euclidiana

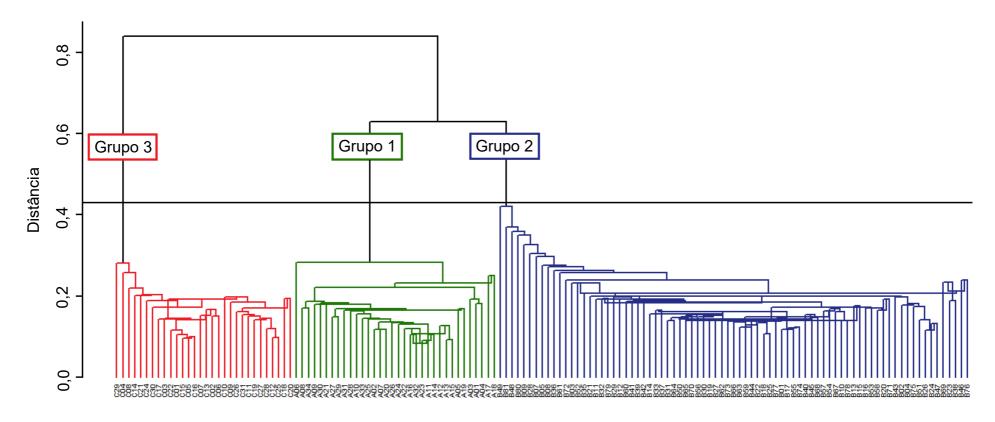
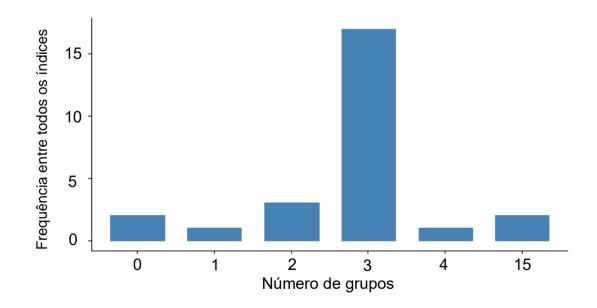


Tabela 5.16 - Agrupamentos gerados pelo método do centróide com base na distância Euclidiana.

Grupo	Amostras
1	A06, A08, A34, A09, A30, A21, A27, A29, A31, A28, A10, A33, A25, A02,
	A07, A20, A26, A24, A22, A16, A32, A23, A11, A14, A12, A13, A15, A05,
	A19, A03, A01, A04, A17, A18.
	B49, B81, B48, B80, B09, B28, B07, B05, B08, B36, B61, B72, B03, B52,
2	B35, B21, B11, B32, B79, B29, B12, B60, B41, B39, B42, B14, B33, B37,
	B31, B64, B50, B65, B70, B56, B30, B19, B27, B62, B73, B66, B63, B59,
	B44, B22, B18, B25, B77, B01, B17, B55, B74, B40, B45, B68, B57, B54,
	B67, B10, B78, B13, B15, B16, B53, B58, B20, B71, B43, B02, B04, B75,
	B51, B26, B24, B47, B69, B23, B38, B46, B76.
3	C29, C04, C08, C14, C21, C24, C30, C17, C03, C22, C01, C15, C05, C16,
	C07, C13, C02, C06, C10, C09, C26, C31, C11, C19, C27, C28, C12, C25,
	C18, C20.

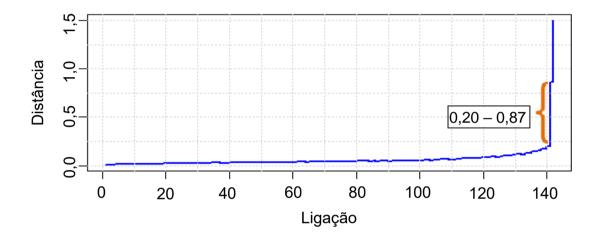
Com o dendrograma da Figura 5.38 e a Tabela 5.16, observa-se que os grupos formados são compostos por amostra de mesmo sítio arqueológico.

Figura 5.39 – Resultado do pacote *NbClust* para o método do centróide com base na distância Euclidiana.



Ao observar a Figura 5.39, nota-se que a maior frequência entre os índices calculados está para a definição de três grupos, indicando que este seria o número de grupos ideal para a partição do conjunto de dados analisado, corroborando, assim, com o resultado da análise do comportamento do nível de fusão (Figura 5.37).

Figura 5.40 - Gráfico da ligação em função da distância do agrupamento das amostras de cerâmica, utilizando o método do centróide e com base na distância Euclidiana ao quadrado.



Fonte: autora da dissertação.

Na Figura 5.40, pode-se observar um maior salto entre as distâncias 0,20 e 0,87, então, um corte adequado a ser realizado no dendrograma, para determinar o número de grupos, é na distância 0,20. Feito o corte nesta distância, formaram-se, dessa forma, três grupos, conforme se observa no dendrograma da Figura 5.41.

Os grupos formados foram chamados de Grupo 1 representado pela cor verde, Grupo 2 representado pela cor azul e Grupo 3 representado pela cor vermelha, sendo o Grupo 1 formado por 34 amostras, o Grupo 2 formado por 79 amostras e Grupo 3 formado por 30 amostras.

Figura 5.41 - Dendrograma do agrupamento das amostras de cerâmica, utilizando o método do centróide e com base na distância Euclidiana ao quadrado.

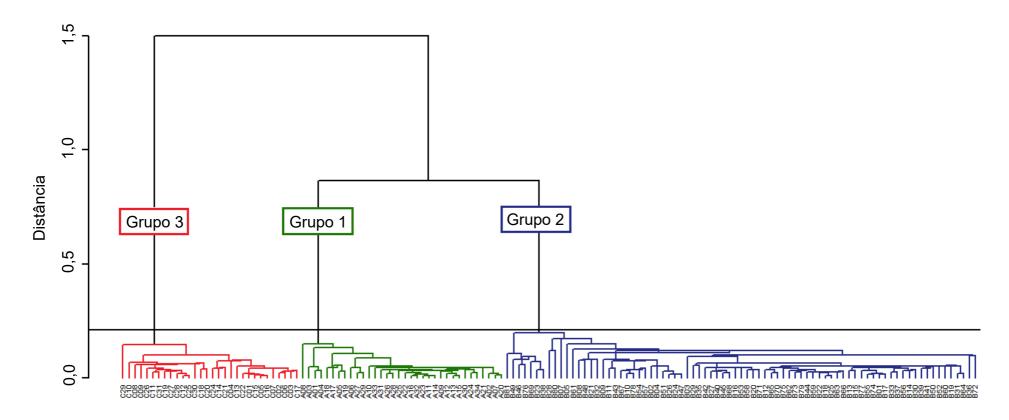
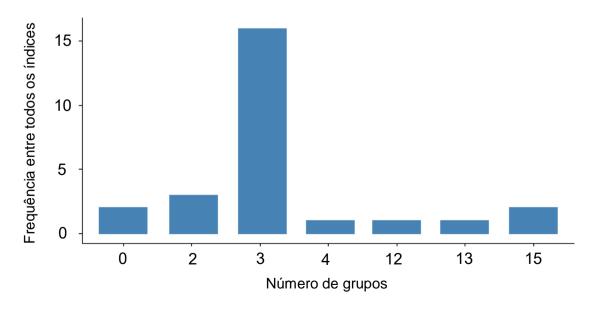


Tabela 5.17 - Agrupamentos gerados pelo método do centróide com base na distância Euclidiana ao quadrado.

Grupo	Amostras
1	A08, A03, A01, A04, A18, A17, A05, A19, A06, A27, A29, A10, A33, A31,
	A26, A28, A25, A22, A16, A32, A23, A11, A14, A09, A12, A13, A15, A30,
	A24, A34, A21, A02, A07, A20.
	B81, B49, B46, B76, B69, B23, B38, B28, B80, B07, B05, B61, B08, B48,
2	B21, B32, B09, B11, B43, B67, B10, B78, B54, B57, B02, B04, B51, B26,
	B24, B47, B03, B29, B35, B42, B27, B40, B45, B68, B16, B53, B58, B20,
	B71, B12, B65, B70, B75, B62, B73, B79, B44, B59, B22, B18, B25, B63,
	B66, B13, B15, B77, B55, B74, B01, B17, B33, B37, B56, B14, B30, B39,
	B41, B50, B52, B60, B19, B31, B64, B36, B72.
3	C29, C10, C08, C09, C26, C11, C31, C19, C27, C28, C12, C25, C30, C18,
	C20, C24, C14, C21, C04, C13, C22, C01, C15, C05, C16, C07, C02, C06,
	C03, C17.

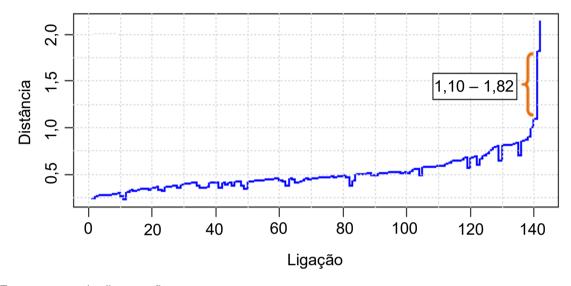
Com o dendrograma da Figura 5.41 e a Tabela 5.17, observa-se que os grupos formados são compostos por amostra de mesmo sítio arqueológico.

Figura 5.42 - Resultado do pacote *NbClust* para o método do centróide com base na distância Euclidiana ao quadrado.



Ao observar a Figura 5.42, nota-se que a maior frequência entre os índices calculados está na definição de três grupos, confirmando o resultado obtido com a análise do comportamento do nível de fusão (Figura 5.40).

Figura 5.43 - Gráfico da ligação em função da distância do agrupamento das amostras de cerâmica, utilizando o método do centróide e com base na distância *Manhattan*.



Fonte: autora da dissertação.

Na Figura 5.43, pode-se observar um maior salto entre as distâncias 1,10 e 1,82, portanto, um corte adequado a ser realizado no dendrograma, para determinar o número de grupos, é na distância 1,10. Feito o corte nesta distância, formaram-se, dessa forma, três grupos, conforme se observa no dendrograma da Figura 5.44.

Os grupos formados foram chamados de Grupo 1 representado pela cor verde, Grupo 2 representado pela cor azul e Grupo 3 representado pela cor vermelha, sendo o Grupo 1 formado por 34 amostras, o Grupo 2 formado por 79 amostras e Grupo 3 formado por 30 amostras, conforme a Tabela 5.18.

Figura 5.44 - Dendrograma do agrupamento das amostras de cerâmica, utilizando o método do centróide e com base na distância *Manhattan*.

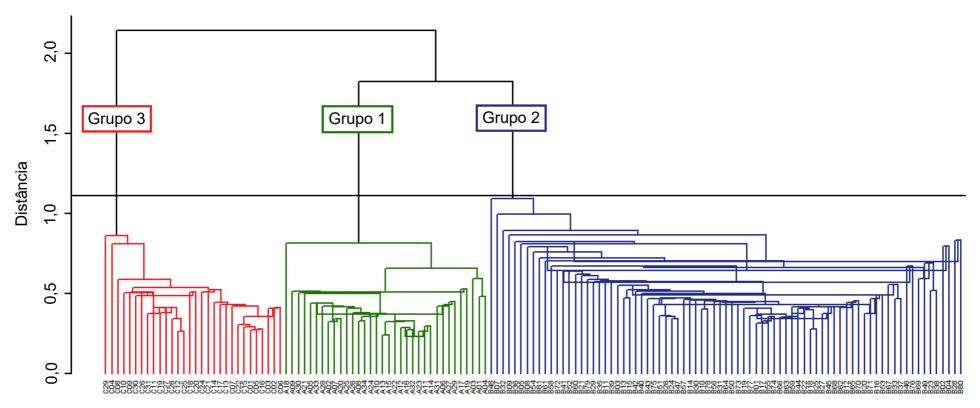
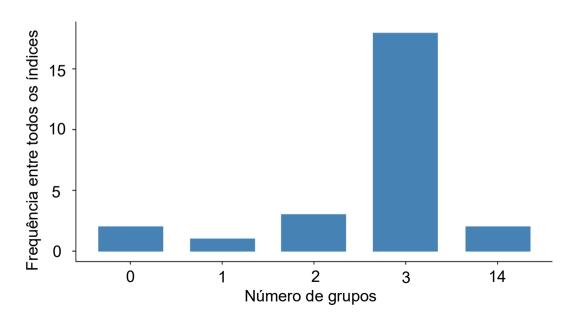


Tabela 5.18 - Agrupamentos gerados pelo método do centróide com base na distância Manhattan.

Grupo	Amostras
1	A18, A09, A30, A21, A05, A33, A28, A02, A07, A20, A25, A26, A08, A34,
	A24, A10, A13, A15, A22, A12, A16, A32, A23, A11, A14, A31, A06, A27,
	A29, A17, A19, A03, A01, A04.
	B48, B07, B32, B09, B36, B05, B08, B54, B81, B61, B58, B72, B41, B52,
2	B60, B21, B79, B29, B35, B11, B39, B03, B13, B15, B42, B40, B43, B75,
	B51, B26, B24, B47, B57, B14, B30, B10, B78, B56, B31, B64, B50, B73,
	B19, B77, B01, B17, B55, B74, B66, B63, B59, B44, B22, B18, B25, B27,
	B45, B68, B62, B12, B65, B70, B20, B71, B16, B53, B67, B33, B37, B46,
	B76, B69, B49, B23, B38, B02, B04, B28, B80.
3	C29, C04, C08, C10, C09, C30, C26, C31, C11, C19, C27, C28, C12, C25,
	C18, C20, C24, C21, C14, C17, C13, C07, C22, C15, C01, C05, C16, C03,
	C02, C06.

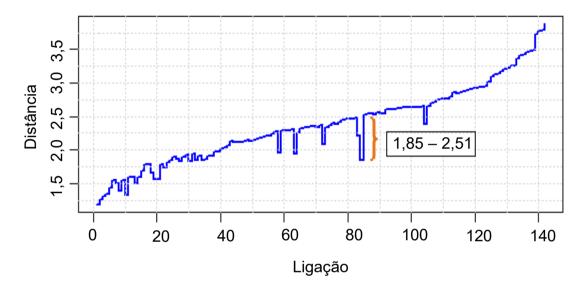
Com o dendrograma da Figura 5.44 e a Tabela 5.18, observa-se que os grupos formados são compostos por amostras de mesmo sítio arqueológico.

Figura 5.45 – Resultado do pacote *NbClust* para o método do centróide com base na distância *Manhattan*.



Ao observar a Figura 5.45, nota-se que a maioria dos índices comparados pelo pacote *NbClust* seleciona três como o número relevante de grupos, indicando que este seria o número de grupos ideal para a partição do conjunto de dados analisados, confirmando o resultado obtido pela a análise do comportamento do nível de fusão (Figura 5.43).

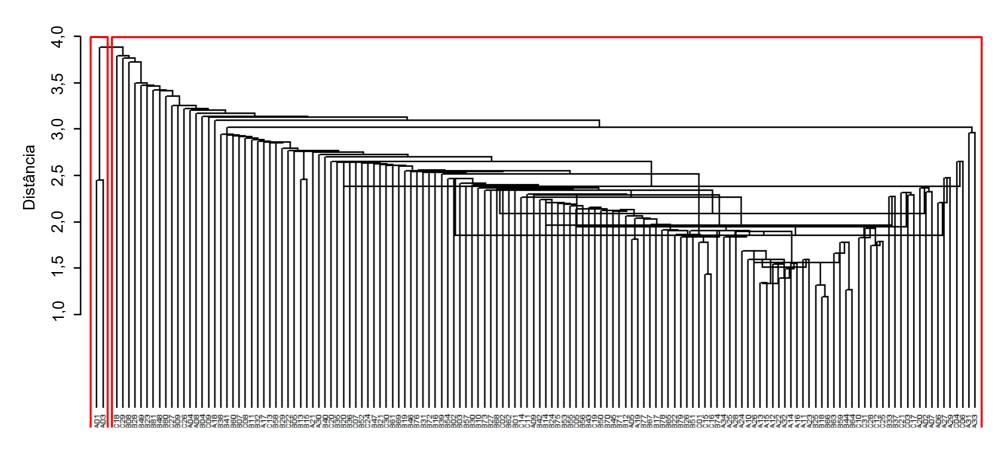
Figura 5.46 - Gráfico da ligação em função da distância do agrupamento das amostras de cerâmica, utilizando o método do centróide e com base na distância *Mahalanobis*.



Fonte: autora da dissertação.

Na Figura 5.46, pode-se observar um maior salto entre as distâncias 1,85 e 2,51, então, um corte adequado a ser realizado no dendrograma, para determinar o número de grupos, é na distância 1,85. Entretanto, o dendrograma gerado para o método do centróide, com base na distância de *Mahalanobis* não gerou grupos definidos (Figura 5.47). Dessa forma, optou-se por utilizar o resultado do pacote *NbClust*, para identificar o número ideal de grupos.

Figura 5.47 - Dendrograma do agrupamento das amostras de cerâmica, utilizando o método do centróide e com base na distância *Mahalanobis*.



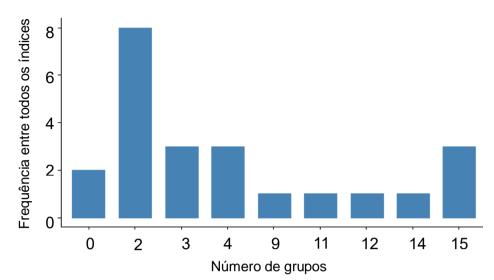


Figura 5.48 - Resultado do pacote *NbClust* para o método do centróide com base na distância *Mahalanobis*.

A partir da análise da Figura 5.48, observa-se que a maioria dos índices comparados no pacote *NbClust* propõe dois como o melhor número de grupos. Sendo assim, foi feito um corte no dendrograma (Figura 5.47), de modo a gerar dois grupos. Os grupos formados foram denominados de Grupo 1 formado por 2 amostras e Grupo 2 formado por 141 amostras, descritos na Tabela 5.19.

Tabela 5.19 - Agrupamentos gerados pelo método do centróide com base na distância Mahalanobis.

Grupo	Amostras
1	A01, A03.
_	C18, C29, B08, B28, B49, B23, B81, B48, B80, C27, B09, C26, A04, A08,
	B04, C09, A18, B38, B41, B60, B07, C08, B11, B32, A17, C13, B58, B29,
	C22, B05, B13, B15, A21, A30, B40, C20, B35, B20, B36, C07, B52, C24,
	B47, B21, C30, B61, B69, B19, B46, B76, B31, B72, B16, B39, B54, B02,
	B03, B57, B30, B10, B73, B27, B68, C02, B62, B01, C14, C11, A09, B42,
2	B14, B24, B75, B53, B55, C05, B56, B43, C19, B50, B70, B45, B71, B12,
	A05, A19, B77, B67, B17, B78, B65, B22, B79, B26, B51, C01, C15, C16,
	B74, A34, A25, A28, A24, A10, A26, A13, A15, A12, A22, A32, A14, A16,
	A11, A23, B25, B18, B66, B63, B59, B44, B64, C10, C31, C28, C12, C25,
	B33, B37, C21, C03, C17, A20, A02, A07, A06, A27, A29, C04, C06, A31,
	A33.

A partir da análise dos dendrogramas do agrupamento utilizando o método do centróide com base nas distâncias Euclidiana, Euclidiana ao quadrado e *Manhattan* (Figuras 5.38, 5.41, 5.44) percebe-se a formação de três grupos bem definidos. Além disso, os grupos formados são os mesmos e compostos por amostras de mesmo sítio arqueológico.

Entretanto, o mesmo não acontece com o método associado à distância de *Mahalanobis* (Figura 5.47), no qual, formaram-se dois grupos: um formado por duas amostras do mesmo sítio e o outro formado por amostras de sítios arqueológicos diferentes.

Além disso, os dendrogramas gerados utilizando como método de agrupamento o método do centróide ficaram mais confusos do que os dendrogramas gerados pelos métodos anteriores. HAIR JR. et al. (2005) ressaltam que este método pode produzir resultados confusos, devido a inversões, ou seja, casos em que a distância entre centróides de um par pode ser menor do que a distância entre os centróides de outro par fundindo em uma combinação anterior.

Este fato fica evidente nas Figuras 5.37, 5.43 e 5.46, onde se observa que apesar de aumentar o número de ligações (fusão), em alguns casos, houve uma diminuição de distância (nível de fusão). Isso poderá ocorrer no método centróide, quando em alguma etapa do algoritmo de agrupamento houver empate entre os valores da matriz de distâncias (MINGOTI, 2017).

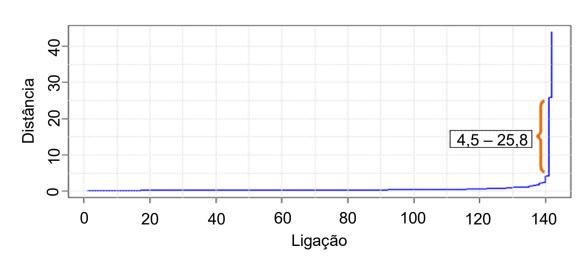


Figura 5.49 - Gráfico da ligação em função da distância do agrupamento das amostras de cerâmica, utilizando o método de Ward e com base na distância Euclidiana.

Na Figura 5.49, pode-se observar um maior salto entre as distâncias 4,5 e 25,8, então, um corte adequado a ser realizado no dendrograma, para determinar o número de grupos, é na distância 4,5. Feito o corte nesta distância, formaram-se, dessa forma, três grupos, conforme se observa no dendrograma da Figura 5.50.

Os grupos formados foram chamados de Grupo 1 representado pela cor verde, Grupo 2 representado pela cor azul e Grupo 3 representado pela cor vermelha, sendo o Grupo 1 formado por 34 amostras, o Grupo 2 formado por 79 amostras e Grupo 3 formado por 30 amostras.

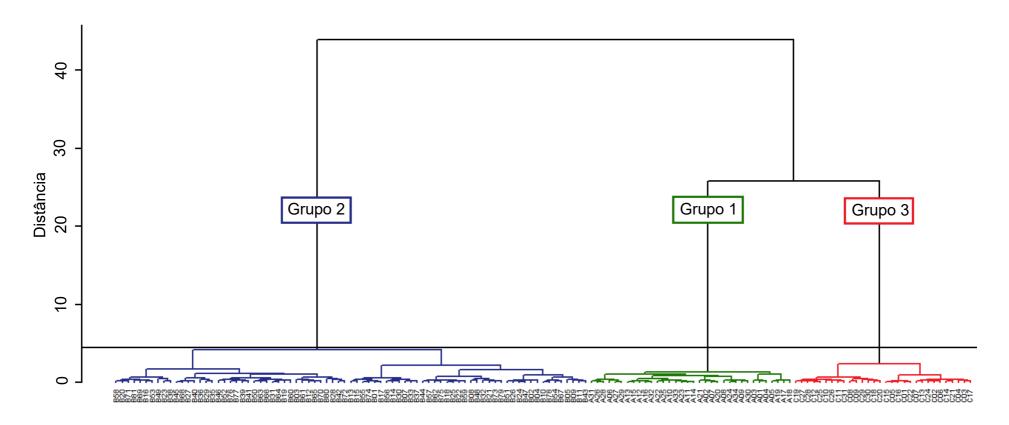
Tabela 5.20 - Agrupamentos gerados pelo método de Ward com base na distância Euclidiana.

Grupo	Amostras					
1	A31, A26, A28, A06, A27, A29, A13, A15, A12, A16, A32, A22, A25, A10,					
	A33, A23, A11, A14, A21, A02, A07, A20, A08, A24, A34, A09, A30, A03,					
	A01, A04, A05, A19, A17, A18.					
2	B58, B20, B71, B81, B69, B16, B53, B49, B23, B38, B45, B68, B27, B40,					
	B36, B29, B35, B46, B52, B76, B77, B39, B41, B50, B63, B66, B31, B64,					
	B19, B60, B03, B61, B12, B65, B70, B80, B28, B42, B72, B13, B15, B55,					
	B74, B01, B17, B56, B14, B30, B07, B33, B37, B44, B57, B62, B75, B18,					
	B25, B22, B59, B08, B48, B32, B21, B73, B79, B51, B26, B24, B47, B02,					
	B04, B10, B78, B54, B67, B05, B09, B11, B43.					
3	C19, C27, C28, C12, C25, C10, C26, C11, C31, C08, C09, C29, C30, C18,					
	C20, C15, C05, C16, C01, C22, C07, C13, C24, C02, C06, C14, C21, C04,					
	C03, C17.					

Fonte: autora da dissertação.

Com o dendrograma da Figura 5.50 e a Tabela 5.20, observa-se que os grupos formados são compostos por amostra de mesmo sítio arqueológico.

Figura 5.50 - Dendrograma do agrupamento das amostras de cerâmica, utilizando o método de Ward com base na distância Euclidiana.



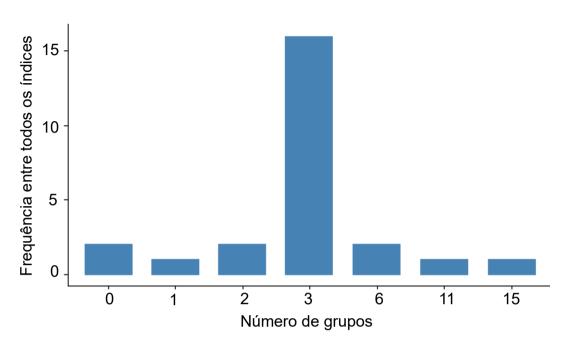
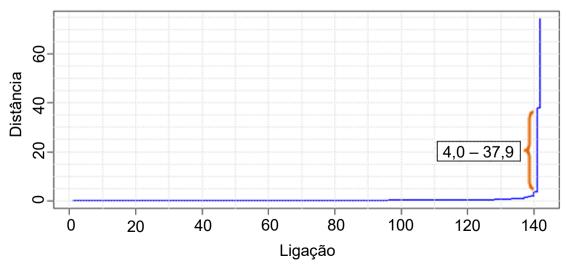


Figura 5.51 - Resultado do pacote *NbClust* para o método de Ward com base na distância Euclidiana.

Ao observar a Figura 5.51, nota-se que a maior frequência entre os índices calculados está para a definição de três grupos, indicando que este seria o número de grupos ideal para a partição do conjunto de dados analisado, confirmando o resultado obtido com a análise do comportamento do nível de fusão (Figura 5.49).

 Figura 5.52 - Gráfico da ligação em função da distância do agrupamento das amostras de cerâmica, utilizando o método de Ward e com base na distância Euclidiana ao quadrado.



Na Figura 5.52, pode-se observar um maior salto entre as distâncias 4,0 e 37,9, então, um corte adequado a ser realizado no dendrograma, para determinar o número de grupos, é na distância 4,0. Feito o corte nesta distância, formaram-se, dessa forma, três grupos, conforme se observa no dendrograma da Figura 5.53.

Os grupos formados foram chamados de Grupo 1 representado pela cor verde, Grupo 2 representado pela cor azul e Grupo 3 representado pela cor vermelha, sendo o Grupo 1 formado por 34 amostras, o Grupo 2 formado por 79 amostras e Grupo 3 formado por 30 amostras.

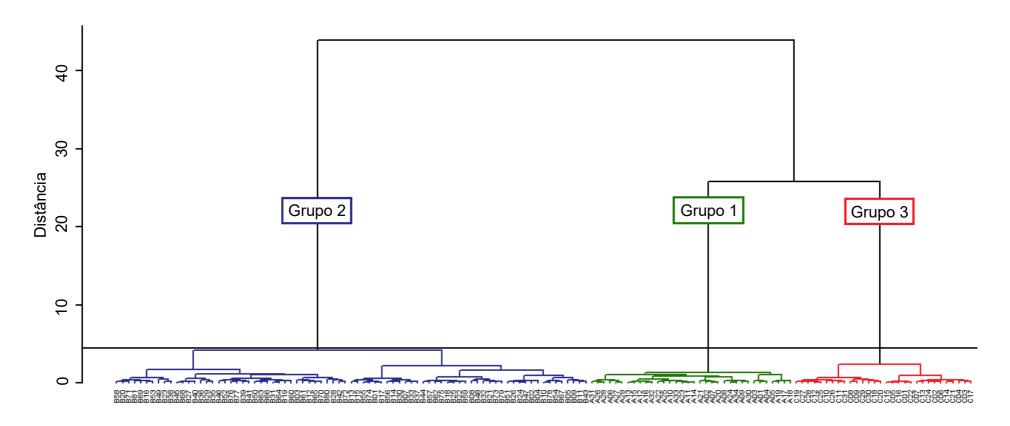
Tabela 5.21 - Agrupamentos gerados pelo método de Ward com base na distância Euclidiana ao quadrado.

Grupo	Amostras
1	A31, A26, A28, A06, A27, A29, A08, A24, A34, A09, A30, A21, A02, A07,
	A20, A12, A13, A15, A23, A11, A14, A22, A16, A32, A10, A33, A03, A01,
	A04, A25, A05, A19, A17, A18.
2	B58, B20, B71, B81, B69, B16, B53, B39, B41, B46, B76, B77, B49, B23,
	B38, B51, B26, B24, B47, B02, B04, B10, B78, B54, B67, B05, B09, B11,
	B43, B13, B15, B55, B74, B01, B17, B56, B14, B30, B07, B33, B37, B28,
	B80, B03, B61, B12, B65, B70, B52, B19, B60, B31, B64, B44, B57, B45,
	B68, B27, B40, B50, B63, B66, B42, B72, B36, B29, B35, B08, B48, B62,
	B75, B18, B25, B22, B59, B32, B21, B73, B79.
3	C29, C30, C18, C20, C08, C09, C19, C27, C28, C12, C25, C10, C26, C11,
	C31, C15, C05, C16, C01, C22, C07, C13, C24, C02, C06, C14, C21, C04,
	C03, C17.

Fonte: autora da dissertação.

Com o dendrograma da Figura 5.53 e a Tabela 5.21, observa-se que os grupos formados são compostos por amostras de mesmo sítio arqueológico.

Figura 5.53 - Dendrograma do agrupamento das amostras de cerâmica, utilizando o método de Ward e com base na distância Euclidiana ao quadrado.

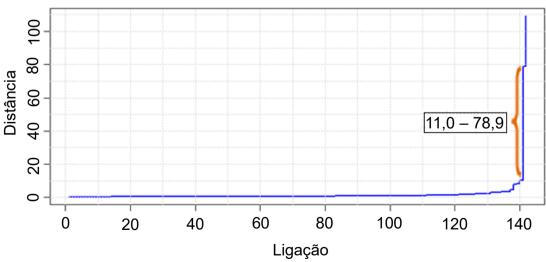


Seguipus 15-0 1 2 3 4 13 15 Número de grupos

Figura 5.54 - Resultado do pacote *NbClust* para o método de Ward com base na distância Euclidiana ao quadrado.

Ao observar a Figura 5.54, nota-se que a maior frequência entre os índices calculados está na definição de três grupos, corroborando, assim, com o resultado da Figura 5.52.

Figura 5.55 - Gráfico da ligação em função da distância do agrupamento das amostras de cerâmica, utilizando o método de Ward e com base na distância *Manhattan*.



Na Figura 5.55, pode-se observar um maior salto entre as distâncias 11,0 e 78,9 e, portanto, um corte adequado a ser realizado no dendrograma, para determinar o número de grupos, é na distância 11,0. Feito o corte nesta distância, formaram-se, então, três grupos, conforme se observa no dendrograma da Figura 5.56.

Os grupos formados foram chamados de Grupo 1 representado pela cor verde, Grupo 2 representado pela cor azul e Grupo 3 representado pela cor vermelha, sendo o Grupo 1 formado por 34 amostras, o Grupo 2 formado por 79 amostras e Grupo 3 formado por 30 amostras.

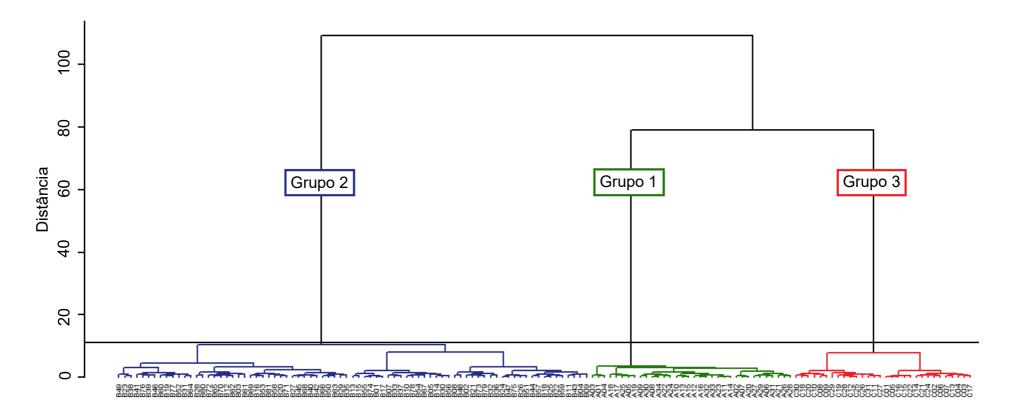
Tabela 5.22 - Agrupamentos gerados pelo método de Ward com base na distância Manhattan.

Grupo	Amostras				
1	A03, A01, A04, A18, A17, A25, A05, A19, A09, A30, A08, A34, A22, A24,				
	A10, A13, A15, A12, A16, A32, A33, A23, A11, A14, A02, A07, A20, A27,				
	A29, A06, A31, A21, A26, A28.				
2	B49, B23, B38, B41, B76, B39, B46, B60, B19, B77, B52, B31, B64, B28,				
	B80, B72, B65, B70, B12, B62, B03, B61, B69, B16, B53, B81, B58, B20,				
	B71, B27, B45, B68, B40, B42, B66, B50, B63, B29, B35, B13, B15, B55,				
	B74, B01, B17, B07, B33, B37, B10, B78, B54, B67, B05, B14, B30, B56,				
	B08, B48, B02, B21, B73, B79, B32, B36, B24, B47, B75, B26, B51, B44,				
	B57, B18, B25, B22, B59, B11, B43, B04, B09.				
3	C30, C18, C20, C10, C08, C09, C29, C19, C28, C12, C25, C26, C31, C11,				
	C27, C01, C05, C16, C15, C22, C14, C21, C24, C02, C06, C07, C13, C04,				
	C03, C17.				

Fonte: autora da dissertação.

Com o dendrograma da Figura 5.56 e a Tabela 5.22, observa-se que os grupos formados são compostos por amostra de mesmo sítio arqueológico.

Figura 5.56 - Dendrograma do agrupamento das amostras de cerâmica, utilizando o método de Ward e com base na distância *Manhattan*.



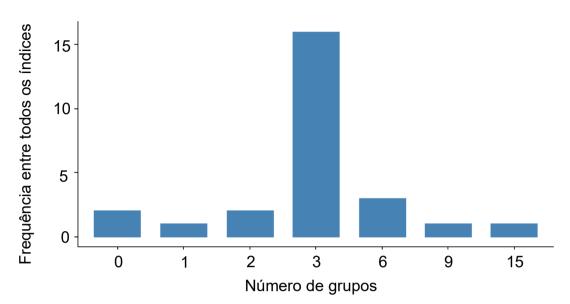


Figura 5.57 - Resultado do pacote *NbClust* para o método de Ward com base na distância *Manhattan*.

Ao observar a Figura 5.57, nota-se que a maior frequência entre os índices calculados está na definição de três grupos, indicando que este seria o número de grupos ideal para a partição do conjunto de dados analisado. Confirmando, então, o resultado obtido com a análise do comportamento do nível de fusão (Figura 5.55).

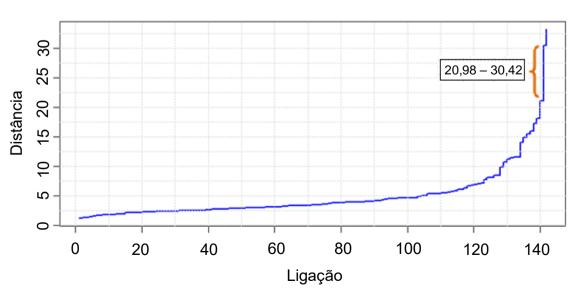


Figura 5.58 - Gráfico da ligação em função da distância do agrupamento das amostras de cerâmica, utilizando o método de Ward e com base na distância *Mahalanobis*.

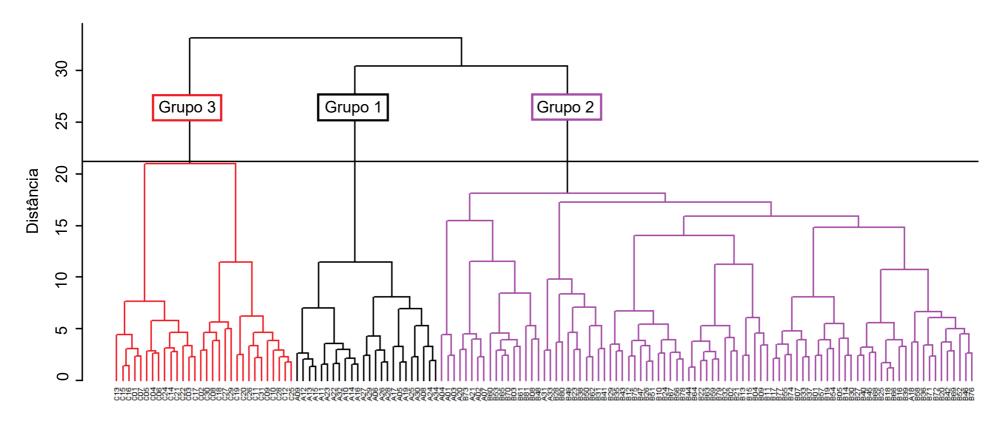
Na Figura 5.58, pode-se observar um maior salto entre as distâncias 20,98 e 30,42, logo, um corte adequado a ser realizado no dendrograma, para determinar o número de grupos, é na distância 20,98. Feito o corte nesta distância, formaram-se três grupos, conforme se observa no dendrograma da Figura 5.59.

Os grupos formados foram chamados de Grupo 1 representado pela cor preta, Grupo 2 representado pela cor roxa e Grupo 3 representado pela cor vermelha, sendo o Grupo 1 formado por 24 amostras, o Grupo 2 formado por 89 amostras e Grupo 3 formado por 30 amostras (Tabela 5.23).

Tabela 5.23 - Agrupamentos gerados pelo método de Ward com base na distância Mahalanobis.

Grupo	Amostras					
1	A09, A12, A13, A15, A11, A23, A22, A32, A10, A14, A16, A27, A29, A06,					
	A26, A28, A17, A05, A19, A25, A30, A08, A24, A34.					
2	A04, A01, A03, A20, B73, A21, A02, A07, B60, B53, B65, B70, B03, B61,					
	B81, B08, B48, A31, A33, B28, B80, B49, B23, B38, B50, B62, B31, B41,					
	B29, B35, B43, B12, B75, B47, B26, B51, B10, B24, B67, B56, B78, B44,					
	B64, B22, B63, B59, B79, B32, B02, B21, B13, B15, B04, B09, B11, B17,					
	B77, B55, B74, B07, B33, B37, B01, B57, B19, B54, B05, B14, B30, B27,					
	B40, B45, B68, B25, B18, B66, B16, B39, A18, B58, B36, B71, B72, B20,					
	B42, B69, B52, B46, B76.					
3	C13, C15, C16, C01, C07, C05, C04, C06, C24, C14, C21, C22, C03, C17,					
	C02, C30, C08, C18, C27, C29, C19, C20, C26, C11, C31, C09, C10, C28,					
	C12, C25.					

Figura 5.59 - Dendrograma do agrupamento das amostras de cerâmica, utilizando o método de Ward e com base na distância *Mahalanobis*.



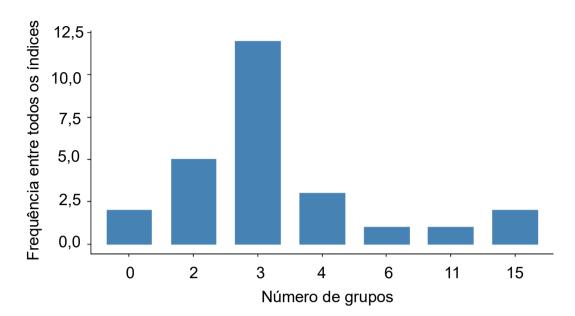


Figura 5.60 - Resultado do pacote *NbClust* para o método de Ward com base na distância *Mahalanobis*.

Ao observar a Figura 5.60, nota-se que a maior frequência entre os índices calculados está na definição de três grupos, corroborando, assim, com o resultado da Figura 5.58.

A partir da análise dos dendrogramas do agrupamento, utilizando o método de Ward com base nas distâncias Euclidiana, Euclidiana ao quadrado e *Manhattan* (Figuras 5.50, 5.53, 5.56), percebe-se a formação de três grupos bem definidos. Além disso, os grupos formados são os mesmos e compostos por amostras de mesmo sítio arqueológico.

Ainda que, o resultado do método de Ward com base na distância Mahalanobis tenha gerado três grupos:

- Grupo 1 composto por 24 amostras do sítio arqueológico Prado;
- Grupo 2 composto por todas as amostras do sítio arqueológico
 Água Limpa mais 10 amostras do sítio arqueológico Prado;
- Grupo 3 composto por todas as amostras do sítio arqueológico Rezende.

Os Grupos 1 e 2 diferem dos grupos gerados pelos outros agrupamentos. Ademais, o Grupo 2 é composto de amostras de sítios diferentes.

5.4.2 Comparação dos agrupamentos gerados pelos métodos utilizados

A partir da análise dos dendrogramas das Figuras 5.2, 5.5, 5.8, 5.14, 5.17, 5.20, 5.26, 5.29, 5.32, 5.38, 5.41, 5.44, 5.50, 5.53, 5.56, percebe-se a formação de três grupos homogêneos.

Embora a estrutura geral dos agrupamentos seja bastante similar, pode-se observar que há pequenas alterações nos níveis em que as amostras são agrupadas, ou seja, as amostras que estão dentro de um mesmo grupo podem ser agrupadas em outra ordem, quando se mudam os métodos.

Para as distâncias Euclidiana, Euclidiana ao quadrado e *Manhattan* associadas aos cinco métodos de agrupamentos, foram formados os mesmos grupos, compostos por amostras do mesmo sítio arqueológico:

- Grupo 1 composto por todas as amostras do sítio Prado;
- Grupo 2 composto por todas as amostras do sítio Água Limpa;
- Grupo 3 composto por todas as amostras do sítio Rezende.

Em relação aos dendrogramas das Figuras 5.12, 5.24, 5.36, 5.47, obtidos pelos métodos: ligação simples, ligação completa, ligação média e centróide com base na distância de *Mahalanobis*, não se tem a formação de grupos homogêneos e os grupos formados misturam amostras de sítios arqueológicos diferentes.

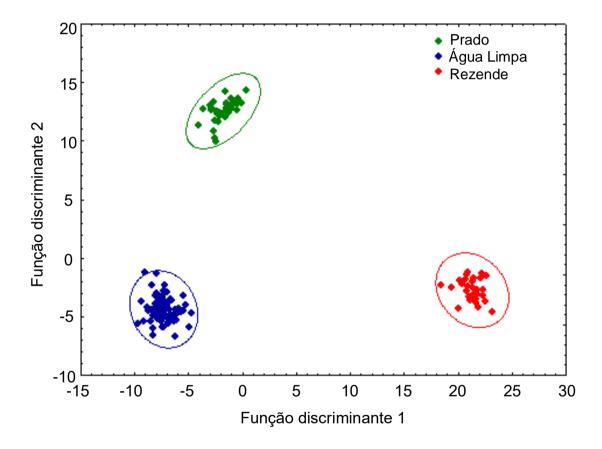
Por último, a partir da análise do dendrograma da Figura 5.59, observase a formação de três grupos. Estes grupos, exceto o Grupo 3, diferem dos grupos gerados pelos outros agrupamentos. Além disso, tem-se grupo composto de amostras de sítios diferentes.

Portanto, pode-se observar que há divergências entre os métodos, corroborando com a afirmativa de JOHNSON e WICHERN (1992), de que dificilmente os dendrogramas obtidos por métodos de agrupamentos diferentes sejam semelhantes.

5.5 Análise discriminante

Foi realizada a análise discriminante a partir dos três grupos gerados pelos cinco métodos associados às distância: Euclidiana, Euclidiana ao quadrado e *Manhattan*, sendo possível gerar o gráfico da Figura 5.61. A partir da Figura 5.61, é possível confirmar os três grupos de composição química similar.

Figura 5.61 - Gráfico da função discriminante 1 vs função discriminante 2. A elipse representa um nível de confiança de 95%.



Fonte: autora da dissertação.

5.6 Coeficiente de Correlação Cofenética

Para os cinco métodos de agrupamento utilizados foram obtidas as respectivas matrizes cofenéticas resultantes da simplificação proporcionada pelo método. A matriz cofenética foi obtida após a construção do dendrograma. Com base nas matrizes de dissimilaridade original e cofenética, foi obtida o coeficiente

de correlação cofenética conforme a Equação 4.5. Os resultados obtidos estão representados na Tabela 5.24.

Tabela 5.24 - Correlações cofenéticas entre as matrizes cofenéticas e as matrizes de dissimilaridades obtidas conforme método de agrupamento utilizado.

Distância	Euclidiana	Euclidiana	Manhattan	Mahalanobis
Método		ao quadrado		
Ligação simples	0,9345	0,8961	0,9257	0,6406
Ligação completa	0,8614	0,7822	0,8886	0,4338
Ligação média	0,9412	0,9024	0,9556	0,6833
Centróide	0,9383	0,9020	0,9496	0,6351
Ward	0,8648	0,7923	0,8937	0,2974

Fonte: autora da dissertação.

Os valores das correlações cofenéticas (Tabela 5.24), para os métodos associados às distâncias Euclidiana, Euclidiana ao quadrado e *Manhattan*, de forma geral, foram todos de magnitude elevada. Isso mostra que há boa representação das matrizes de dissimilaridade na forma de dendrogramas e que isso independe do método usado.

Entretanto, os valores obtidos para os cinco métodos associados à distância de *Mahalanobis* foram de magnitude inferior a 0,70. Segundo ROHLF (1970), na prática, dendrogramas com coeficiente de correlação cofenética menor que 0,70 indicariam a inadequação do método de agrupamento para resumir a informação do conjunto de dados, o que explica o fato dos dendrogramas gerados pelos cinco métodos associado à distância *Mahalanobis* gerar grupos não definidos.

Sendo assim, os resultados da Tabela 5.24 revelam a maior consistência dos agrupamentos formados com base nas distâncias Euclidiana, Euclidiana ao quadrado e *Manhanttan* em relação à *Mahalanobis*. Levando em consideração que os resultados correspondem a três sítios arqueológicos, o uso da distância de *Mahalanobis* levaria a falsas interpretações, uma vez que estaria indicando que foi usada a mesma matéria-prima na fabricação dessas peças.

Além disso, comparando os valores do coeficiente de correlação cofenética, observa-se que independentemente da métrica de distância utilizada o

método de ligação média foi o que forneceu os melhores resultados, corroborando com a literatura (KUIPER e FISHER, 1975; MILLIGAN e COOPER, 1988; SARAÇLI et al., 2013; SOKAL e ROHLF, 1962).

5.7 Interpretação dos grupos gerados pela análise de agrupamentos

Levando em consideração o que foi discutido no item anterior, não será feita a interpretação dos grupos gerados pelos os cinco métodos associados à distância de *Mahalanobis*.

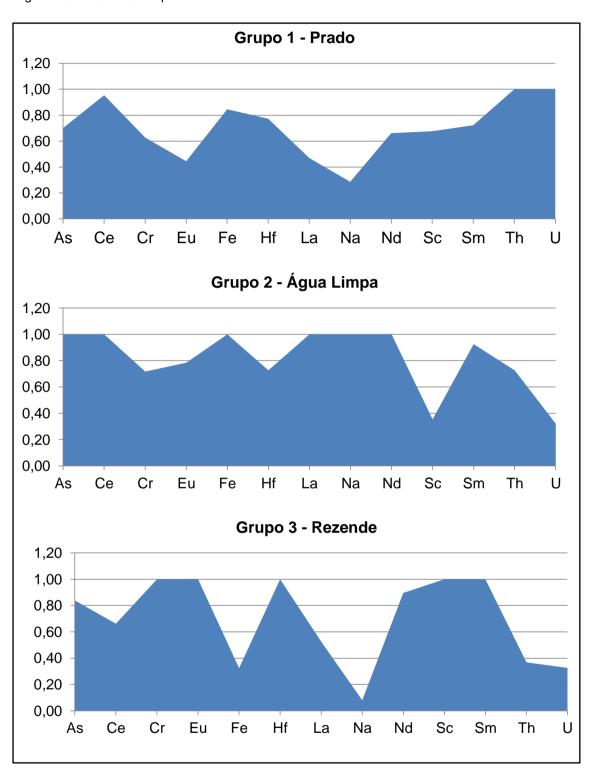
Segundo BARROSO e ARTES (2003), quando se tem muitas variáveis, é difícil interpretar uma tabela com as médias observadas para as variáveis em cada grupo. Neste contexto, a utilização de gráficos de representação de casos facilita a observação de semelhanças e diferenças entre os grupos.

Sendo assim, foi utilizado o gráfico de perfil para fazer a interpretação dos grupos. Neste gráfico, é conveniente que os valores máximos de cada variável não difiram muito. Para garantir a igualdade dos valores máximos, o gráfico foi construído com os valores obtidos dividindo-se o valor de cada média, pela maior média observada para a variável (BARROSO e ARTES, 2003).

No gráfico de perfil as observações de cada grupo são representadas separadamente. No eixo das abscissas, é indicado as variáveis e no eixo das ordenadas, as escalas de medida. Cada média é representada por um ponto nos eixos cartesianos. Unindo-se os pontos obtêm-se os perfis de cada grupo.

A partir da análise do gráfico de perfil (Figura 5.62), nota-se que as frações de massa para os elementos analisados diferem de acordo com os grupos. Justificando, assim, a separação observada na análise de agrupamentos.

Figura 5.62 – Gráfico de perfis.



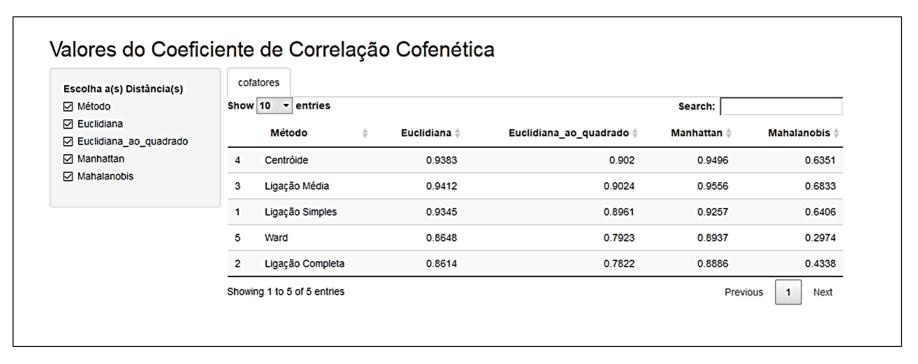
5.8 Script do R

Por último, um *script* com funções do programa estatístico R foi desenvolvido para calcular o coeficiente de correlação cofenética, com o intuito de identificar o método hierárquico de análise de agrupamento mais adequado para uma base de dados. O objetivo deste *script* é facilitar o estudo de pesquisadores que não são da área estatística ou que não estão familiarizados com o programa.

Sendo assim, o *script* desenvolvido torna-se útil, já que basta apenas inserir a base de dados no programa estatístico R e executá-lo, obtendo uma tabela com todos os valores de correlação cofenética. Desta forma, o pesquisador pode facilmente verificar qual método e distância são mais apropriados para sua base de dados. A tabela gerada pelo *script*, desenvolvido neste trabalho, está representada na Figura 5.63.

O script, desenvolvido neste trabalho, encontra-se no Apêndice.

Figura 5.63 - Tabela gerada pelo script desenvolvido.



CONCLUSÃO

Neste trabalho, realizou-se um estudo comparativo entre os métodos hierárquicos de análise de agrupamentos utilizando uma base de dados de 146 amostras de fragmentos cerâmicos de três sítios arqueológicos, Prado, Água Limpa e Rezende. O principal objetivo deste trabalho era apontar o método mais adequado a ser utilizado para a base de dados dos sítios arqueológicos.

Com base nas interpretações dos dedrogramas gerados, pelos métodos ligação simples, ligação completa, ligação média, centróide e de Ward associados às distâncias Euclidiana, Euclidiana ao quadrado e *Manhattan*, foi observado à formação de três grupos homogêneos. Os grupos formados são os mesmos e compostos por amostras do mesmo sítio arqueológico.

Entretanto, o mesmo não aconteceu para os cincos métodos associados à distância de *Mahalanobis*. Para essa distância, não houve a formação de grupos homogêneos e os grupos formados misturaram amostras de sítios arqueológicos diferentes.

Para a determinação do número de grupos foram utilizados dois critérios: a análise de comportamento do nível de fusão e o pacote *NbClust* do programa R. Contudo, para a distância de *Mahalanobis*, não foi possível determinar o número de grupos utilizando a análise de comportamento do nível de fusão, sendo necessário à utilização do outro critério. Pelos resultados obtidos verificou-se que o pacote *NbClust* é mais eficiente para determinar o número ideal de grupos.

A validação dos grupos formados pela análise de agrupamentos foi realizada por meio da análise discriminante e do coeficiente de correlação cofenética. Por meio da análise discriminante confirmou-se a formação de três grupos de composição química similar, constituídos por amostras de mesmo sítio arqueológico. Certificando, dessa forma, que os grupos obtidos pelos cinco métodos de agrupamentos, associados às distâncias Euclidiana, Euclidiana ao quadrado e *Manhattan*, realmente, diferem e que os agrupamentos gerados pelos

cinco métodos associados à distância de *Mahalanobis* levam a falsas interpretações. Este fato também pôde ser observado, comparando os valores da correlação cofenética, uma vez que, somente para a distância de *Mahalanobis* obteve-se valores abaixo de 0,70, indicando, assim, a inadequação do método de agrupamento para se resumir a informação do conjunto de dados.

Em todas as análises efetuadas nos dados de fragmentos cerâmicos dos sítios arqueológicos Prado, Água Limpa e Rezende, o uso da distância de *Mahalanobis*, independentemente do método de agrupamento utilizado, gerou grupos em que misturam amostras de sítios diferentes, menores valores de coeficiente de correlação cofenética e dendrogramas menos úteis aos propósitos desejados.

Para a determinação do melhor método de agrupamento para a base de dados estudada, foram avaliados os dendrogramas conjuntamente com o valor dos coeficientes de correlação cofenética. Apesar dos cinco métodos avaliados neste trabalho associados às quatro distâncias, exceto a *Mahalanobis*, apresentarem resultados concordantes entre si quanto à formação dos grupos, o método de ligação média apresentou os maiores valores do coeficiente de correlação cofenética. Sendo assim, para essa base de dados, o método de ligação média, é o mais adequado.

Por último, o programa R mostrou-se uma ferramenta poderosa para aplicação das técnicas de análises multivariadas, principalmente pelos seus recursos gráficos. No entanto, o seu uso requer conhecimento teórico dos conceitos envolvidos e uma noção de programação adequada por parte do pesquisador, para que tenha uma boa utilização da ferramenta. Dessa forma, o *script* desenvolvido é de grande importância, uma vez que ajudará os pesquisadores a encontrar o método de agrupamento mais apropriado para sua base de dados.

REFERÊNCIAS BIBLIOGRÁFICAS

AINSWORTH, M. W. From connoisseurship to Technical Art History: The Evolution of the Interdisciplinary Study of Art. *The Getty Conservation Institute Newsletter*, v. 20, n. 1, p. 4-10, 2005.

ALBUQUERQUE, M. A. *Estabilidade em análise de agrupamento (Cluster Analysis)*. 2005. 62 p. Dissertação (Mestrado em Biometria e Estatística Aplicada) - Universidade Federal Rural de Pernambuco, Recife. Disponível em: < http://www.tede2.ufrpe.br:8080/tede/handle/tede2/5178>. Acesso em: 10 jul. 2018.

ALDENDERFER, M. S.; BLASHFIELD, R. K. *Cluster Analysis*. Beverly Hills, CA: SAGE, 1984.

ANDERBERG, M. R. *Cluster analysis for applications*. New York, NY: Acafenic, 1973.

BAKER, F. B. Stability of two hierarchical grouping techniques case I: sensitivity to data errors. *Journal of the American Statistical Association*, v. 69, n. 346, p. 440-445, 1974. DOI: 10.1080/01621459.1974.10482971

BARIA, R.; CANO, N. F.; SILVA-CARRERA, B. N.; WATANABE, S.; NEVES, E. G.; TATUMI, S. H.; MUNITA, C. S. Archaeometric studies of ceramics from the São Paulo II archaeological site. *Journal of Radioanalytical and Nuclear Chemistry*, v. 306, n.3, p. 721-727, 2015. DOI: 10.1007/s10967-015-4183-5

BARROSO, L. P.; ARTES, R. *Análise multivariada*: minicurso do 10º Simpósio de Estatística Aplicada à Experimentação Agronômica – SEAGRO, 48ª Região Brasileira da Sociedade Internacional de Biometria – RBRAS. Lavras: UFLA, 2003.

BAXTER, M. J.; BEARDAH, C. C.; PAPAGEORGIOU, I.; CAU, M. A.; DAY, P. M.; KILIKOGLOU, V. On statistical approaches to the study of ceramics artefacts using geochemical and petrographic data. *Archaeometry*, v. 50, n.1, p.142-157, 2008. DOI: 10.1111/j.1475-4754.2007.00359-x

BEIER, T.; MOMMSEN, H. Modified Mahalanobis filters for grouping pottery by chemical composition. *Archaeometry*, v. 36, n. 2, p. 287-306, 1994. DOI: 10.1111/j.1475-4754.1994.tb00971.x

BLASHFIELD, R. K. Mixture model tests of cluster analysis: accuracy of four agglomerative hierarchical methods. *Psychological Bulletin*, v. 83, n. 3, p. 377-388, 1976. DOI: 10.1037/0033-2909.83.3.377

BLASIUS, E.; WAGNER, H.; BRAUN, H.; KRUMBHOLZ, R.; SCHWARTZ, B. Tile fragments as characteristic evidence for ancient Roman settlements: scientific investigations. *Archaeometry*, v. 25, n. 2, p. 165-178, 1983. DOI: 10.1111/j.1475-4754.1983.tb00673.x

BUSSAB, W. O.; MIAZAKI, E. S.; ANDRADE, D. F. *Introdução à Análise de Agrupamentos*. São Paulo: Associação Brasileira de Estatística, 1990.

CARANDINI, A. *Archeologia e cultura materiale*. Il lavoro senza Gloria nella antichità clásica. Bari: de Donato, 1975. p. 53-93.

CHARRAD, M.; GHAZZALI, N.; BOITEAU, V.; NIKNAFS, A. NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. *Journal of Statistical Software*, v. 61, n. 6, p. 1-36, 2014. DOI: 10.18637/jss.v061.i06

COSTIN, L. C. The use of ethnoarchaelogy for the archaeological study of ceramic production. *Journal of Archaeological Method and Theory*, v. 7, n. 4, p. 377-403, 2000. DOI: 10.1023/A:1026523023550

DALGAARD, P. *Introductory Statistics with R:* Statistics and Computing. New York, NY: Springer, 2002.

D'ERCOLE, G.; ERAMO, G.; GARCEA, E. A. A.; MUNTONI, I. M.; SMITH, J. R. Raw material and technological changes in ceramic productions at Sai Island, Northern Sudan, from the seventh to the third millennium B.C. *Archaeometry*, v. 57, n. 4, p. 597-616, 2015. DOI: 10.1111/arcm.12113

DJINGOVA, R.; KULEFF, I. An archaeometric study of Medieval glass from the first Bulgarian capital, Pliska (ninth to tenth century AD). *Archaeometry*, v. 34, n. 1, p. 53-61, 1992. DOI: 10.1111/j.1475-4754.1992.tb00477.x

DUNN, J. Well Separated Clusters and Optimal Fuzzy Partitions. *Journal Cybernetics*, v. 4, n. 1, p.95-104, 1974. DOI: 10.1080/01969727408546059

EVERITT, B. S.; LANDAU, S.; LEESE, M.; STAHL, D. *Cluster analysis*. 5ht ed. London: Edward, 2011.

FÁVERO, L. P.; BELFIORE, P.; SILVA, F. L.; CHAN, B. L. *Análise de dados:* modelagem multivariada para tomada de decisões. 2. ed. Rio de Janeiro: Elsevier, 2009.

FELICÍSSIMO, M. P.; PEIXOTO, J. L.; PIREAUX, J. J.; DEMORTIER, G.; RODRIGUES, U. P. F. Estudos arqueométricos de cerâmicas indígenas précoloniais das lagoas Castelo e Vermelha, localizadas no Pantanal Sul-Mato-Grossense. *Canindé*, v. 4, p. 325-368, 2004.

FERREIRA, D. F. *Análise multivariada*, Lavras: UFLA, 2008.

FLOREK, K.; LUKASZEWICZ, L.; PERKAL, J.; STEINHAUS, H.; ZUBRZYCKI, S. Sur la liaison et la division des points d'un ensemble fini. *Colloquium Mathematicum*, v. 2, n. 3-4, p. 282-285, 1951.

FUNTUA, I. I.; OLADIPO, M. O. A; NJINGA, R. L.; JONAH, S. A; YUSUF, I.; AHMED, Y. A. Evaluation for accuracy and applicability of instrumental neutron activation analysis of geological materials on Nigeria Nuclear Research Reactor – 1(NIRR - 1). *International Journal of Applied Science and Technology*, v. 2, n. 1, p. 286 – 292, 2012.

GLASCOCK, M. D. Characterization of archaeological ceramics at MURR by neutron activation analysis and multivariate statistics. In: NEFF, H.(Ed.) *Chemical characterization of ceramic pastes in archaeology*, p. 11-26, Madison: Prehistory, 1992. p. 11-26.

GLASCOCK, M. D.; NEFF, H. Neutron activation analysis and provenance research in archaeology. *Measurement Science and Tecnology*, v. 14, n. 9, p. 1516-1526, 2003. DOI: 10.1088/0957-0233/14/9/304

GLIOZZO, E.; VIVACQUA, P.; MEMMI, I. T. Integrating archaeology, archaeometry and geology: local production technology and imports at Paola (Cosenza, Southern Italy). *Journal of Archaeological Science*, v. 35, n. 4, p. 1074-1089, 2008. DOI: 10.1016/j.jas.2007.07.008

GUIRAO, D.; PLA, F.; ACOSTA, A. The archaeometric characterization of Majolica ceramics from Talavera de la Reina and El Puente del Arazobispo (Toledo, Spain). *Archaeometry*, v. 56, n. 5, p. 746-763, 2014. DOI: 10.1111/arcm.12048

HAIR JR, J. F.; ANDERSON, R. E.; TATHAM, R. L.; BLACK, W. C. *Análise multivariada de dados*. Tradução: Adonai Schlup Sant'Anna e Anselmo Chaves Neto. 5. ed. Porto Alegre: Bookman, 2005.

HAIR JR., F. J; BLACK, W. C.; BABIN, B. J.; ANDERSON, R. E. Cluster Analysis. In: HAIR JR., F. J; BLACK, W. C.; BABIN, B. J.; ANDERSON, R. E. *Multivariate Data Analysis:* Pearson New International Edition. 7th ed. Pearson, 2014. p. 415-474.

HALKIDI, M.; VAZIRGIANNIS, M. Clustering validity assessment: finding the optimal partitioning of a data set. In: *Proceedings 2001 IEEE International Conference on Data Mining*, 29 Nov. - 2 Dec., 2001, San Jose, CA, USA. IEEE, 2001. p.187-194.

HALKIDI, M.; VAZIRGIANNIS, M.; BATISTAKIS, I. Quality scheme assessment in the clustering process. In: ZIGHED, D. A.; KOMOROWSKI, J.; ZYTKOW, J. (Ed.) *Principles of data mining and knowledge discovery:* 4th European Conference, PKDD 2000 Lyon, France, September 13-16, 2000, Proceedings, v. 1910 of Lecture Notes in Computer Science. Springer, 2000. p. 265-276.

HARBOTTLE, G. Activation Analysis in Archaeology. In: NEWTON, G. W. A. (Ed.) *Radiochemistry*, v. 3. London: Royal Society of Chemistry, 1976. p. 33-72. DOI: 10.1039/9781847556882-00033

HAYASHIDA, F. Bridging the gap between archaeology and the physical sciences. *Hyperfine Interactions*, v. 150, n. 1, p. 7-11, 2003. DOI: 10.1023/B:HYPE.0000007174.27143.85

HAZENFRATZ-MARKS, R. *Nêutrons, radiação e arqueologia:* estudo de caso multianalítico de cerâmicas da tradição borda incisa na Amazônia Central. 2014. 453 p. Tese (Doutorado em Tecnologia Nuclear) – Instituto de Pesquisas Energéticas e Nucleares, IPEN-CNEN/SP, São Paulo. Disponível em: < http://www.teses.usp.br/teses/disponiveis/85/85131/tde-02072014-140014/en.php > Acesso em: 10 jul. 2018.

HAZENFRATZ, R.; MUNITA, C. S.; GLASCOCK, M. D.; NEVES, E. G. Study of exchange networks between two Amazon archaeological sites by INAA. *Journal of Radioanalytical and Nuclear Chemistry*, v. 309, n. 1, p.195-205, 2016. DOI: 10.1007/s10967-016-4758-9

HEGMON, M. Advances in ceramic ethnoarchaeology. *Journal of Archaeological Method and Theory*, v. 7, n. 3, p. 129-137, 2000. DOI: 10.1023/A:1026502419007

IAEA – TRS416. *Nuclear Analytical Techniques in Archaeological Investigations*. Viena, 2003.

JOHNSON, R. A.; WICHERN, D. W. *Applied multivariate statistical analysis*. 3rd ed. Englewood Cliffs, NJ: Prentice Hall, 1992.

JOHNSON, S. C. Hierarchical clustering schemes. *Psychometrika*, v. 32, n. 3, p. 241–254, 1967.

JONES, A. Archaeometry and materiality: materials-based analysis in theory and practice. *Archaeometry*, v. 46, n. 3, p. 327-338, 2004. DOI: 10.1111/j.1475-4754.2004.00161.x

KALENTZIDOU, O. Discontinuing traditions: using historically informed ethnoarchaeology in the study of Evros ceramics. *Journal of Archaeological Method and Theory*, v. 7, n. 3, p. 165-186, 2000. DOI: 10.1023/A:1026506519916

KAUFMAN, L.; ROUSSEEUW, P. J., *Finding groups in data:* an introduction to cluster analysis. New York, NY: Wiley, 1990.

KOCH, G. S.; LINK, R. F. *Statistical analysis of geological data*. New York, NY: Courier Dove Publications, 2002.

KUIPER, F. K; FISHER, L. A Monte Carlo comparison of six clustering procedures. *Biometrics*, v. 31, n. 3, p. 777-783, 1975. DOI: 10.2307/2529565

KULEFF, I.; DJINGOVA, R.; DJINGOV, G. Provenience study of medieval Bulgarian glasses by NAA and cluster analysis. *Archaeometry*, v. 27, n. 2, p. 185-193, 1985. DOI: 10.1111/j.1475-4754.1985.tb00361.x

- LANCE, G. N.; WILLIAMS, W. T. A general theory of classificatory sorting strategies: 1. Hierarchical systems. *Computer Journal*, v. 9, n. 4, p. 373-380, 1967. DOI: 10.1093/comjnl/9.4.373
- LEBART, L.; MORINEAU, A.; PIRON, M. *Statistique Exploratoire Multidimensionnelle*. Paris: Dunod, 2000.
- LEUTE, U. *Archaeometry:* an introduction to physical methods in archaeology and the history of art. New York, NY: VCH, 1987.
- MARTINS, C. M. Cultura material e iconografia: um estudo das ânforas gregas do festival das Panateneias. *Cadernos do LEPAARQ*, v. 8, n. 15/16, p. 129-151, 2011. DOI: 10.15210/lepaarq.v8i15/16.1678
- MILLIGAN, G. W.; COOPER, M. C. A study of standardization of variables in cluster analysis. *Journal of Classification*, v. 5, n. 2, p. 181-204, 1988. DOI: 10.1007/BF01897163
- MILLIGAN, G. W.; COOPER, M. C. An examination of procedures for determining the number of clusters in a data set. *Psychometrika*, v. 50, n. 2, p.159-179, 1985. DOI: 10.1007/BF02294245
- MINGOTI, S. A. *Análise de dados através de métodos estatísticos multivariada:* uma abordagem aplicada. Belo Horizonte: UFMG, 2017.
- MIRTI P.; ZELANO V.; ARUGA, R.; FERRARA E.; APPOLONIA L.; Roman pottery from Augusta Praetoria (Aosta, Italy): a provenance study. *Archaeometry*, v. 32, n. 2, p. 163-175, 1990. DOI: 10.1111/j.1475-4754.1990.tb00463.x
- MUCHA, H. J.; BARTEL, H. G. Resampling techniques in cluster analysis: is subsampling better than bootstrapping? In: LAUSEN, B; KROLOK-SCHWERDT, S; BOHMER, M. (Ed.) *Data science, learning by latent structures, and knowledge discovery*. Springer, 2015. p. 113-122.
- MUNITA, C. J. A. S. Arqueometria: estudo de interfaces. *CLIO Arqueológica*, v. 2, n.19, p. 30-42, 2005.
- MUNITA, C. S. Contribución de la química a la arqueología. In: JIMÉNEZ, E. N. (Org.). *Las Técnicas Analíticas Nucleares y el Patrimonio Cultural*. Quito: INPC, 2010. p. 23-33.

- MUNITA, C. S.; CARVALHO, P. R. Grupo de Estudos Arqueométricos do IPEN-CNEN/SP. *Cadernos do CEOM*, v. 28, n. 43, p. 53-59, 2015.
- MUNITA, C. S.; NASCIMENTO, A.; SCHREIBER, S. B.; LUNA, S.; OLIVEIRA, P. M. S. Chemical study of some ceramics from Brazilian Northeast. *Journal of Radioanalytical and Nuclear Chemistry*, v. 259, n. 2, p. 305-309, 2004. DOI: 10.1023/B:JRNC.0000017308.65405.b5
- MUNITA, C. S.; PAIVA, R. P.; ALVES, M. A.; OVLIEIRA, P. M. S.; MOMOSE, E. F. Provenance study of archaeological ceramic. *Journal of Trace and Microprobe Techniques*, v. 21, n. 4, p. 697-706, 2003. DOI: 10.1081/TMA-120025819
- MURRELL, P. *R Graphics*. Boca Raton: Chapman & Hall/CRC, 2006.
- MURTAGH, F.; CONTRERAS, P. Algorithms for hierarchical clustering: na overview. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, v. 2, n. 1, p. 86-97, 2012. DOI: 10.1002/widm.53
- NUNES, K. P.; TOYOTA, R. G.; OLIVEIRA, P. M. S.; NEVES, E. G.; SOARES, E. A. A.; MUNITA, C. J. A. S. Preliminary compositional evidence of provenance of ceramics from Hatahara Archaeological site, Central Amazonia. *Journal of Chemistry*, v. 2013, p. 1-6, 2013. DOI: 10.1155/2013/701748
- OLIVEIRA P. M. S.; MUNITA C. S. Influência do Valor Crítico na Detecção de Valores Discrepantes em Arqueometria, In: *48ª Reunião Anual Região Brasileira da Sociedade Internacional de Biometria, 07-11 de Julho*, Lavras, MG Brasil, p. 545-550, 2003.
- OLIVEIRA, M. G.; ZOTES, L. P.; QUELHAS, O. L. G.; PEREIRA, V.; CARVALHO, C. A. S. Comparação dos métodos de análise agrupamentos k-means e Partitioning Around Medoids (PAM). *Revista Brasileira de Estatística*, v. 76, n. 241, p. 51-78, 2015.
- PAIVA, S. The determination of brick provenance and technology using analytical techniques from the physical sciences. *Archaeometry*, v. 48, n.2, p.201-218, 2006. DOI: 10.1111/j.1475-4754.2006.00251.x
- PAPAGEORGIOU, J.; BAXTER, M. J. Model-based cluster analysis of artefact compositional data. *Archaeometry*, v. 43, n. 4, p. 571-588, 2001. DOI: 10.1111/1475-4754.00037

R Development Core Team. R Foundation for Statistical Computing. R: A language and environment for statistical computing, Vienna, Austria. ISBN 3-900051-07-0, URL. Disponível em: http://www.R-project.org. Acesso em: 13 jul. 2018.

RAURET, G., CASASSAS E.; RIUS F. X.; MUNOZ, M. Cluster analysis applied to spectrochemical data of European Medieval stained glass. *Archaeometry*, v. 29, n. 2, p. 240-249, 1987. DOI: 10.1111/j.1475-4754.1987.tb00417.x

RIZZUTTO, M. A. Métodos físicos e químicos para estudo de bens culturais. *Cadernos do CEOM*, v. 28, n. 43, p. 67-76, 2015.

ROHLF, F. J. Adaptive hierarchical clustering schemes. *Systematic Zoology*, v. 18, n. 1, p. 58-82, 1970. DOI: 10.1093/sysbio/19.1.58

ROMESBURG, C. H. *Cluster analysis for researchers*. Belmont: Lifetime Learning Publications, 1984.

ROUSSEEUW, P. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, v. 20, p. 53-65, 1987. DOI: 10.1016/0377-0427(87)90125-7

SARAÇLI, S.; DOGAN, N.; DOGAN, I. Comparison of hierarchical cluster analysis methods by cophenetic correlation. *Journal of Inequalities and Applications*, v. 203, p.1-8, 2013. DOI: 10.1186/1029-242X-2013-203

SAXENA, A.; PRASAD, M.; GUPTA, A.; BHARILL, N.; PATEL, O. P.; TIWARI, A.; ER, M. J.; DING, W.; LIN, C. A review of clustering techniques and developments. *Neurocomputing*, v. 267, p. 664-681, 2017. DOI: 10.1016/j.neucom.2017.06.053

SAYRE, E. V. *Brookhaven procedures for statistical analyses of multivariate archaeometric data*. New York: Brookhaven National Laboratory, 1975.

SCARLETT, T. J.; SPEAKMAN, R. J.; GLASCOCK, M. D. Pottery in the mormon economy: an historical, archaeological, and archaeometric study. *Historical Archaeology*, v. 41, n. 4, p. 72-97, 2007. DOI: 10.1007/BF03377296

SCHUBERT, P. Petrographic modal analysis – a necessary complement to chemical analysis of ceramic coarse wear. *Archaeometry*, v. 28, n. 2, p. 163-178, 1986. DOI: 10.1111/j.1475-4754.1986.tb00384.x

- SEMAAN, G. S.; TORRES, C. C.; BRITO, J. A. M.; OCHI, L. S. Um método baseado em combinação de soluções com coassociação para o problema de agrupamento automático. *Revista Brasileira de Estatística*, v. 74, n. 239, p. 43-68, 2013.
- SHACKLEY, M. S. Archaeological petrology and the archaeometry of lithic materials. *Archaeometry*, v. 50, n. 2, p. 194-215, 2008. DOI: 10.1111/j.1475-4754.2008.00390.x
- SNEATH, P. H. A. The application of computers to taxonomy. *Journal of General Microbiology*, v. 17, p. 201-226, 1957. DOI: 10.1099/00221287-17-1-201
- SOKAL, R. R.; MICHENER, C.D. A statistical method for evaluating systematic relationships. *Bulletin of the Society University of Kansas*, n. 38, p. 1409-1438, 1958. DOI: 13960/t84j1fk1m
- SOKAL, R. R.; ROHLF, F. J. The comparison of dendrograms by objective methods. *Taxon*, v. 11, n. 2, p. 33-40, 1962. DOI: 10.2307/1217208
- SOKAL, R. R.; SNEATH, P. H. A. *Principles of Numerical Taxonomy*. San Francisco, CA: W. H. Freeman, 1963.
- STEWART, J. D.; FRALICK, P.; HANCOCK, R. G. V.; KELLEY J. H.; GARRETT E. M. Petrographic analysis and INAA geochemistry of prehistoric ceramics from Robinson Pueblo, New Mexico. *Journal of Archaeological Science*, v. 17, N. 6, p. 601-625, 1990. DOI: 10.1016/0305-4403(90)90043-5
- TIBSHIRANI, R.; WALTHER, G.; HASTIE, T. Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society, Series* **B**, v. 63, n. 2, p. 411-423, 2001. DOI: 10.1111/1467-9868.00293
- TREBUNA, P.; HALCINOVÁ, J. Mathematical tools of cluster analysis. *Applied Mathematics*, v. 4, p. 814-816, 2013. DOI: 10.4236/am.2013.45111
- TUDELA, D. R. G.; TATUMI, S. H.; YEE, M.; BRITO, S. L. M.; MORAIS, J. L.; MORAIS, D.; PIEDADE, S. C.; MUNITA, C. S.; HAZENFRATZ, R. TL, OSL and C-14 datind results of the sediments and bricks from mummified nun's grave. *Anais da Academia Brasileira de Ciências*, v. 84, n. 2, p. 237-244, 2012. DOI: 10.1590/S0001-37652012005000031

VALENTIN, J. L. *Ecologia numérica: uma introdução à análise multivariada de dados ecológicos*. 2. ed. Rio de Janeiro: Interciência, 2012.

VENABLES, W. N.; SMITH, D. M.; THE R CORE TEAM. *An introduction to R*, 2018. Disponível em: < https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf > Acesso em: 10 jul. 2018.

WARD, J. H. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, v. 58, n. 301, p. 236-244, 1963. DOI: 10.1080/01621459.1963.10500845

WARNER, R. M. *Applied statistics:* from bivariate through multivariate techniques. 2nd ed. Los Angeles, CA: SAGE, 2013.

WATANABE, S. *Propriedades de termoluminescência do LiF:Mg*. 1969. 85 p. Tese (Livre Docência em Física Nuclear) - Instituto de Física, IF-USP, São Paulo. Disponível em:< http://www.teses.usp.br/teses/disponiveis/livredocencia/43/tde-11082015-091646/>. Acesso em: 10 jul. 2018.

WEIGAND, P. C.; HARBOTTLE, G.; SAYRE, E. V. Turquoise sources and source analysis: Mesoamerica and the Southwestern U.S.A.. In: EARLE, T. K.; ERICSON, J. E. (Ed.) *Exchange systems in prehistory*. New York, NY: Academic Press, 1977. p.15-34. DOI: 10.1016/B978-0-12-227650-7.50008-0

YAN, L. T.; FENG, S. L.; LI, L.; MA, B.; LIU, L.; FENG, X. Q. The compositional study of ceramic samples from tombs dating to the Zhou Dynasty at Xiaoshan, Zhejiang Province (China). *Archaeometry*, v. 57, n. 5, p. 822-836, 2015. DOI: 10.1111/arcm.12136

APÊNDICE

Script do programa desenvolvido

```
LEIA ATENTAMENTE ANTES DE USAR
#
                                                                    #
  Procedimento para o cálculo dos CCC pelos diferentes Métodos
                                                                    #
                                                                    #
# 1. Importe a matriz de interesse sem os títulos de linha e coluna:
                                                                    #
    Comando Import Dataset (menu superior do quadro direito)
                                                                    #
                                                                    #
# obs: Para que tudo ocorra bem é necessário:
                                                                    #
                                                                    #
    a. Baixar os pacotes vegan e DT (menu inferior do quadro direito)
    b. Setar ambos os pacotes na 4ª janela na aba Packages
                                                                    #
      obs: se a library nas linhas 27 deste arquivo, linhas 3
# e 4 do ui.R e 1 server.R estiverem direcionadas corretamente,
# não será necessário o set.
 2. Gravar os arquivos ui.R e server.R em um diretorio a partir
    do menu file>new file>shiny web app...
#
    obs: redirecione o arquivo da linha 65 em runApp
                                                                    #
# 3. Digite CTRL+R na linha 24 (ou no botão [>Run App] no canto
                                                                    #
    superior direito desta janela) para realização do cálculo
#
#======>Criando as Funções cofnt e calcof<======
cofht < -function(x,y,z){
library("vegan", lib.loc="C:/Program Files/R/R-3.3.2/library")
ifelse (y == "euclidian2",signif(cor(vegdist(x,"euclidian")^2,
                      cophenetic(hclust(vegdist(x,"euclidian")^2,z))),digits=4),
     signif(cor(vegdist(x,y),cophenetic(hclust(vegdist(x,y),z))),digits=4))
};calcof<-function(matriz){</pre>
 #### Métodos hierárquicos #########
 sl<-"single"
 cl<-"complete"
 al<-"average"
 cm<-"centroid"
 w<-"ward.D"
 ###### Distâncias #######
 eu<-"euclidean"
 eu2<-"euclidian2"
 man<-"manhattan"
 mah<-"mahalanobis"
 ###### Matriz de Dados ########
 dados<-matriz
```

Montagem do Data Frame

Método<-c("Ligação Simples","Ligação Completa","Ligação Média","Centróide","Ward") Euclidiana<-c(cofht(dados,eu,sl),cofht(dados,eu,cl),cofht(dados,eu,al),cofht(dados,eu,cm),cofht(dados,eu,w))

Euclidiana_ao_quadrado<-c(cofht(dados,eu2,sl),cofht(dados,eu2,cl),cofht(dados,eu2,al),cofht(dados,eu2,cm),cofht(dados,eu2,w))

Manhattan<-c(cofht(dados,man,sl),cofht(dados,man,cl),cofht(dados,man,al),cofht(dados,man,cm),cofht(dados,man,w))

Mahalanobis<-c(cofht(dados,mah,sl),cofht(dados,mah,cl),cofht(dados,mah,al),cofht(dados,mah,cm),cofht(dados,mah,w))

data.frame(Método,Euclidiana,Euclidiana_ao_quadrado,Manhattan,Mahalanobis)

}; cofatores<-calcof(TabelaFinal1); shiny::runApp('E:/andre/R/novo') #troque o nome da matriz

Arquivo ui.R

```
library(shiny)
library("DT", lib.loc="C:/Program Files/R/R-3.3.2/library")
#library(ggplot2) # for the diamonds dataset
fluidPage(titlePanel(HTML("Valores do Coeficiente de Correla&ccedil:&atildeo
Cofenética")),
 title = 'Examples of DataTables',
# cofatores<-calcof(),
 sidebarLayout(
  sidebarPanel(
    conditionalPanel(
     'input.dataset === "cofatores"',
     checkboxGroupInput('show_vars',HTML('Escolha a(s) Distância(s)'),
                 names(cofatores), selected = names(cofatores))
   )),
  mainPanel(
    tabsetPanel(
     id = 'dataset',
     tabPanel('cofatores', DT::dataTableOutput('mytable1'))
```

Arquivo serve.R

```
library(shiny)
#library(ggplot2)

function(input, output) {

    # choose columns to display
    diamonds2 = cofatores[sample(nrow(cofatores), 5), ]
    output$mytable1 <- DT::renderDataTable({
        DT::datatable(diamonds2[, input$show_vars, drop = FALSE])
    })}</pre>
```



INSTITUTO DE PESQUISAS ENERGÉTICAS E NUCLEARES Diretoria de Pesquisa, Desenvolvimento e Ensino Av. Prof. Lineu Prestes, 2242 – Cidade Universitária CEP: 05508-000 Fone/Fax(0XX11) 3133-8908 SÃO PAULO – São Paulo – Brasil http://www.ipen.br

O IPEN é uma Autaquia vinculada à Secretaria de Desenvolvimento, associada à Universiade de São Paulo e gerida técnica e administrativamente pela Comissão Nacional de Energia Nuclear, órgão do Ministério da Ciência, Tecnologia e Inovação.