

UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA E ESTATÍSTICA APLICADA
DIEGO ALVES GOMES

**UTILIZAÇÃO DE ANÁLISE MULTIVARIADA PARA SELEÇÃO DE GRUPOS
DIVERGENTES EM GENÓTIPOS DE MANDIOCA**

RECIFE

2018



DIEGO ALVES GOMES

**UTILIZAÇÃO DE ANÁLISE MULTIVARIADA PARA SELEÇÃO DE GRUPOS
DIVERGENTES EM GENÓTIPOS DE MANDIOCA**

Dissertação apresentada ao programa de Pós-Graduação em Biometria e Estatística Aplicada da Universidade Federal Rural de Pernambuco como pré-requisito para obtenção do grau de Mestre, sob orientação do prof. Dr. Guilherme Rocha Moreira.

**Área de concentração: Biometria e
Estatística Aplicada.**

Recife

2018

Dados Internacionais de Catalogação na Publicação (CIP)
Sistema Integrado de Bibliotecas da UFRPE
Biblioteca Central, Recife-PE, Brasil

G633u Gomes, Diego Alves
Utilização de análise multivariada para seleção de grupos divergentes em
genótipos de mandioca / Diego Alves Gomes. – 2018.
35 f. : il.

Orientador: Guilherme Rocha Moreira.
Dissertação (Mestrado) – Universidade Federal Rural de Pernambuco,
Programa de Pós-Graduação em Biometria e Estatística Aplicada, Recife, BR-PE,
2018.

Inclui referências.

1. Estatística 2. Mandioca 3. Biometria 4. Análise multivariada I. Moreira,
Guilherme Rocha, orient. II. Título

CDD 574.018

DIEGO ALVES GOMES

**UTILIZAÇÃO DE ANÁLISE MULTIVARIADA PARA SELEÇÃO DE GRUPOS
DIVERGENTES EM GENÓTIPOS DE MANDIOCA**

Dissertação apresentada ao programa de Pós-Graduação em Biometria e Estatística Aplicada da Universidade Federal Rural de Pernambuco como pré-requisito para obtenção do grau de Mestre.

Orientador:

Prof. Dr. Guilherme Rocha Moreira
Universidade Federal Rural de Pernambuco - UFRPE

Banca examinadora:

Prof. Dr. Moacyr Cunha Filho
Universidade Federal Rural de Pernambuco - UFRPE

Prof. Dr. Alexandre Lima Ferreira
Universidade Federal de São João Del Rey – UFSJ

Aprovado em ____/____/____

Dedico a DEUS e a minha família.

Agradecimentos

Agradeço a Deus por ter saúde e discernimento para poder seguir com meus objetivos e sonhos, a Nossa Senhora Aparecida por sempre interceder para que eu alcance meus sonhos.

Ao professor e orientador Guilherme Rocha Moreira pela paciência, estímulo, confiança e dedicação durante a orientação deste trabalho. A minha amada e querida família meus pais Vânia e Antônio ao meu grande amigo e irmão Diogo, aos meus tios Carlos e Zefinha pelo apoio nos momentos de dificuldades.

Ao auxílio da CAPES com uma bolsa de estudos a Universidade Federal Rural de Pernambuco (UFRPE) ao programa de pós-graduação em Biometria e Estatística Aplicada e a todos os professores que ajudaram ao longo dos meus estudos.

Aos meus amigos em especial Fábio e Rosendo, a minha noiva Juliana Maria por se minha amiga e companheira e por me apoiar nos meus estudos e sempre me dar seu carinho e amor, e aos meus queridos amigos André, Kerolly, Fernando Henrique, agradeço a DEUS por ter conhecido vocês e a todos que de uma forma ou de outra me ajudaram durante o mestrado.

RESUMO

Objetivou-se a seleção de genótipos de mandioca (*Manihot esculenta crantz*) para formação de grupos divergentes de genótipos. Foram utilizadas 23 variáveis quantitativas de 28 genótipos, Amansa Burro, Aramaris, Bom Jardim, Bromadeira, Caipira, Caitité, Caravela, Kiriris, Lagoão, Lavra Velha, Malacacheta MR, Mulatinha, Parazinha, Peru, Poti Branca, Salangor, Sergipana, Sergipe, Sergipe MR, Simbé, Tapioqueira, Tussuma, Verdinha, 2006-4, 2006-5, 2006-8, 2006-10 e 2006-12. Por meio de análise fatorial reduziu-se o número de variáveis a treze, alturas das plantas, diâmetro de caule, comprimento das raízes tuberosas, diâmetro das raízes tuberosas, peso das raízes tuberosas, índice de colheita, produtividade da parte aérea, produtividade de raízes tuberosas, porcentagem de massa seca de raízes tuberosas, teor de amido em raízes tuberosas, produtividade de amido em raízes tuberosas, teor de farinha em raízes tuberosas e produtividade de farinha em raízes tuberosas. Utilizou-se a análise de agrupamento hierárquico, em que a melhor combinação dos métodos hierárquicos para essa situação experimental foi a Distância de Chebyshev com o método de ligação médio. Os grupos foram formados com o auxílio do índice de Ratkowsky. O grupo G1 formado com o genótipo Mulatinha possui características balanceadas entre parte superior e inferior da planta, na produção da parte aérea da mandioca e raiz com grande diâmetro e no teor de amido na raiz, apresentando baixo diâmetro de caule e teor de matéria seca. O grupo G2 formado com os genótipos Salangor, Tussuma, Bom jardim, Lavra velha e Sergipana apresentaram em comum, bom teor de farinha nas raízes tuberosas apesar da baixa produtividade de raízes tuberosas, pouco teor de amido e baixa produtividade da parte aérea. No grupo G3 formado pelos genótipos Poti Branco, Caitite, Sergipe, Simbé, Amansa burro e Peru encerraram como principais características, diâmetro de raízes tuberosas, a porcentagem de massa seca nas raízes tuberosas, o teor de amido nas raízes tuberosas, o teor de farinha em raízes tuberosas, os baixos valores de diâmetro do caule e o comprimento das raízes tuberosas. O grupo G4 formado por 2006-5, possui características semelhantes ao grupo G3, visto esse genótipo ter sido modificado geneticamente na Embrapa. No grupo G5, formado por Parazinha, Sergipe MR, Lagoão, Bromadeira, Caravela, Aramaris e Tapioqueira, obteve-se valores altos para o diâmetro das raízes tuberosas, o peso de raízes tuberosas, o índice de colheita, a porcentagem de massa seca nas raízes tuberosas, o teor de amido nas raízes

tuberosas e o teor de farinha nas raízes tuberosas, e, baixos valores de diâmetro do caule, de produtividade de raízes tuberosas e de produtividade de parte aérea. Para o grupo G6, formado com os genótipos 2006-8 e 2006-12 destaca-se as características índice de colheita e de produtividade de raízes tuberosas, com baixos valores para o diâmetro do caule e o comprimento das raízes tuberosas. No grupo G7, composto pelos genótipos 2006-4 e 2006-12, apresentam características satisfatória para índice de colheita e produtividade de raízes tuberosas, e baixo valores para altura de planta e produtividade de parte aérea. O último grupo, o G8 formado por Kiriris, Malacacheta MR, Caipira e Verdinha a principal característica desse grupo é a presença de todas as variáveis.

Palavras-chave: Métodos preditivos, Dendrograma, Raízes Tuberosas.

ABSTRACT

The aim of the selection of genotypes of cassava (*Manihot esculenta* crantz) to divergent groups of genotypes. 23 quantitative variables were used 28 genotypes, Amansa Burro, Aramaris, Bom Jardim, Bromadeira, Caipira, Caitité, Caravela, Kiriris, Lagoão, Lavra Velha, Malacacheta MR, Mulatinha, Parazinha, Peru, Poti Branca, Salangor, Sergipana, Sergipe, Sergipe MR, Simbé, Tapioqueira, Tussuma, Verdinha, 2006-4, 2006-5, 2006-8, 2006-10 e 2006-12. Through factor analysis reduced the number of variables to thirteen, heights of the plants, stem diameter, length of the tuberous roots, diameter of Tuberous roots, weight of the tuberous roots, harvest index, aboveground productivity, productivity of Tuberous roots, percentage of dry mass of Tuberous roots, , levels of starch in tuberous roots, productivity of starch in tuberous roots, content of flour in tuberous roots and productivity of flour in tuberous roots. Using the hierarchical cluster analysis, in which the best combination of hierarchical methods for this situation was the distance of Chebyshev with the connection method. The Group G2 formed with the genotypes Salangor, Tussuma, Bom jardim, Lavra velha and Sergipana presented in common, good flour content in tuberous roots despite the low productivity of Tuberous roots, little starch content and low productivity of the shoot. The G3 Group formed by genotypes Poti Branco, Caitite, Sergipe, Simbé, Amansa burro and Peru , closed as key features, diameter of Tuberous roots, the percentage of dry pasta in tuberous roots, the starch content in tuberous roots, the flour content in tuberous roots, the low values of stem diameter and length of the tuberous roots. The G4 group formed by 2006-5, has similar characteristics to the G3 Group, seen this genotype have been genetically modified at Embrapa. The G5 group, formed by Parazinha, Sergipe MR, Lagoão, Bromadeira, Caravela, Aramaris e Tapioqueira, high values for the diameter of the tuberous roots, tuberous root weight, harvest index, the percentage of dry pasta in tuberous roots, the content of starch in tuberous roots, and the flour content in tuberous roots and low values of diameter of stem, productivity of Tuberous roots and aboveground productivity. For the G6 group, formed with the genotypes 2006-8 and 2006-12 stands out the characteristics and crop productivity index of Tuberous roots, with low values for the diameter of the stem and the length of the tuberous roots. In the G7 group, composed of the 2006-4 and 2006-12 genotypes, satisfactory characteristics for harvest and productivity index of Tuberous roots, and productivity of Tuberous roots,

and low values for plant height and productivity of the shoot. The latter group, the G8 formed by Kiriris, Malacacheta MR, Caipira and Verdinha the main feature of this group is the presence of all variables

Key-words: Predictive methods, Dendrogram, Tuberous Roots.

LISTA DE FIGURAS

Figura 1 - Gráfico de correlações das variáveis observados nos genótipos de mandioca da cidade de Cândido Sales-BA.	22
Figura 2 - Gráfico das Variâncias dos fatores com a rotação varimax	24
Figura 3 - Gráfico de correlação das variáveis selecionas na análise fatorial dos dados de genótipos de mandioca na Cidade de Cândido Sales – BA.	25
Figura 4 - Gráfico de Distribuição das variáveis predominantes em cada genótipos de mandioca na cidade de Cândido Sales – BA.	27
Figura 5 - Dendrograma gerado da matriz de dissimilaridade da distância de Chebyshev utilizando o método de ligação média com grupos sugeridos no índice de Ratkowsky.	28
Figura 6 - Dendrograma gerado da matriz de dissimilaridade da distância de Mahalanobis utilizando o método de ligação média com grupos sugeridos no índice de Ratkowsky.	30

LISTA DE TABELAS

Tabela 1 - Genótipos de mandioca com todas as variáveis	20
Tabela 2 - Autovalores, proporção da variância explicada e proporção da variância explicada acumulada, para os fatores com a rotação varimax.....	23
Tabela 3 - Resultado da análise descritiva do banco de dados para variáveis selecionadas na análise fatorial nos genótipos de mandioca da cidade de Cândido Sales BA.....	24
Tabela 4 - Grupos gerados da matriz de dissimilaridade da distância de Chebyshev, utilizando o método de ligação média com grupos sugeridos no índice de Ratkowsky.	29
Tabela 5 - Grupos gerados da matriz de dissimilaridade da distância de Mahalanobis, utilizando o método de ligação média com grupos sugeridos no índice de Ratkowsky.	31

SUMÁRIO

1	INTRODUÇÃO	13
2	REVISÃO DE LITERATURA	14
2.1	MANDIOCA (<i>Manihot esculenta crantz</i>)	14
2.2	ANALISE FATORIAL	16
2.3	ANÁLISE DE AGRUPAMENTOS	18
3	MATERIAL E MÉTODOS	19
4	RESULTADOS E DISCUSSÃO	21
5	CONCLUSÕES	32
6	REFERÊNCIAS	33

1 INTRODUÇÃO

Diversas áreas do conhecimento científico colhem e analisam múltiplas informações com intenção de conhecer o comportamento de determinados fenômenos. Essas informações são, posteriormente, organizadas e transformadas em base de dados.

A forma como esses dados são trabalhados é de extrema importância para que os resultados representem a realidade do fenômeno abordado. A estatística multivariada aliada com os avanços tecnológicos vem proporcionando planejamento, execução e análise de pesquisas mais eficientes.

Os primeiros pesquisadores a explorar informações no âmbito multivariado foram: Pearson, Fisher, Hotelling, Wilks e Bartlett.

A Estatística multivariada é conjunto de métodos e técnicas estatísticas que analisam variáveis que possuem diversas informações simultaneamente em uma única unidade experimental (JOHNSON; WICHERN, 2014).

Dentre os diversos objetivos da estatística multivariada temos a simplificação onde o objeto de estudo é apresentado de forma mais simples possível sem perder suas características principais, indicação de grupos de objetos ou indivíduos com atributos semelhantes com base em critério confiável, verificação da existência de relação entre variáveis e etc.

Existem várias técnicas da estatística multivariada que podem ser utilizadas dependendo do fenômeno e o que o pesquisador deseja verificar. Alguns exemplos são: análise fatorial, análise de agrupamento, medidas de similaridade e dissimilaridade, métodos de ligação, dendrograma, coeficiente cofenético e etc.

Podemos definir a análise fatorial como método estatístico multivariado no qual o objetivo principal é sumarizar relações em conjunto de informações, numa matriz de dados, ou seja, a análise fatorial procura resolver o problema de identificar relações entre número grande de variáveis, resumindo e reduzindo o número de variáveis em um sistema de dimensões similares denominados fatores (EVERITT; HOTHORN, 2011).

A análise de agrupamento também conhecida como análise de conglomerados ou análise de cluster, é conjunto de técnicas cujo o objetivo é possibilitar a divisão da amostra, ou população em sistema, onde seja possível visualizar grupos constituídos

de elementos similares (homogêneos) dentro dos grupos e diferentes (heterogêneos) entre os grupos.

A estrutura de grupos criada na análise de agrupamento depende inicialmente de uma medida que mensure a distância ou semelhança entre dois elementos, objetos e indivíduos. Essa medida pode ser classificada em medida de similaridade (quanto maior o valor mais próximo ou semelhantes são os elementos) e medida de dissimilaridade (quanto maior o valor mais distante ou diferente são os elementos). Ao término dessa etapa uma matriz é gerada com as distâncias ou semelhanças entre os elementos.

Um método é utilizado na matriz de distâncias afim de obter um esquema com grupos que pode ser visualizado de forma mais simples através de uma representação gráfica chamada dendrograma.

Os métodos que organizam os grupos podem ser classificados em dois, métodos hierárquicos e métodos não hierárquicos. Cada qual possui seus critérios e restrições na criação dos grupos, porém a finalidade é a mesma: manter a homogeneidade dentro dos grupos e heterogeneidade entre grupos.

A confirmação do melhor método de construção de grupos é realizada pela verificação do coeficiente cofenético, quanto mais elevado o valor do coeficiente melhor é a visualização dos grupos no dendrograma.

A fim de contribuir com a agricultura da mandioca na cidade de Cândido Sales – BA, esse trabalho foi realizado para ajudar a compreender as diferenças genéticas de 28 genótipos de mandioca utilizando a estatística multivariada, método de análise fatorial os agrupamentos e medidas de dissimilaridade entre outras, para obter grupos divergentes e demonstrar as principais características presente em cada grupo de genótipos, com esses grupos os programas de melhoramento podem obter cruzamentos com espécies superiores e manter um banco de informações sobre os genótipos da cidade.

2 REVISÃO DE LITERATURA.

2.1 MANDIOCA (*Manihot esculenta crantz*)

A mandioca (*Manihot esculenta crantz*) é espécie de *Euphorbiaceae* popularmente conhecida por macaxeira ou aipim, uma das principais fontes de

carboidratos para milhões de pessoas ao redor do mundo em especial nos países em desenvolvimento. O Brasil é o país que possivelmente originou essa espécie, a qual é cultivada praticamente em todo território brasileiro, desempenhando papel importante na indústria e alimentação (ZAGO et al., 2017).

O estado da Bahia é o terceiro maior produtor e obteve uma produção de 1.956.103/ha de mandioca no ano de 2018 e a cidade de Cândido Sales - BA possui uma área plantada de 6.000 hectares (IBGE, 2016). Entretanto, houve uma queda na produção em Cândido Sales e no estado da Bahia. Uma das causas da queda de produtividade de mandioca no Brasil, é a falta de cultivares mais resistente, aos vários solos e climas existente no Brasil e a falta de informação sobre manuseio desse plantio dos pequenos agricultores (LOPES et al., 2010).

Por ser uma espécie rústica e de plantio simples, geradora de renda no meio rural, ofertando aos agricultores oportunidades de trabalho, torna-se importante o aprimoramento desse cultivo. Os programas de melhoramento genético auxiliam os agricultores proporcionando cultivos mais resistentes do processo de plantio até a colheita. Uma etapa fundamental nos programas de melhoramento é a compreensão da divergência genética para poder realizar os cruzamentos entre as espécies. A avaliação da divergência genética de espécies populares como a mandioca, auxilia a conservar as culturas regionais e manter vivas as espécies selvagens (FAO, 2013).

Alguns procedimentos vem sendo bastantes usadas para caracterizar a divergência genética, como os procedimentos biométricos que podem ser por análise dialélica, onde é preciso realizar combinações entre as variedades, e depois observar cada uma delas. Se os números de genótipos forem grandes isso pode se tornar inviável. Já nos meios preditivos existe a vantagem de se conhecer os fatores que indicam divergência entre os indivíduos antes de realizar os cruzamentos, com isso facilitando a escolha dos genótipos mais propícios a gerar espécies superiores.

No método preditivo destaca-se a análise multivariada, onde não é necessário ter cruzamentos prévios entre espécies (FONSECA et al., 2006). A utilização da análise multivariada combinada com a caracterização morfológica de cultivares tem proporcionado bons resultados em estudos de divergência genética, com romã (MAHAJAN; JAVED; KAPOOR, 2018), azeite (ABUZAYED; FRARY; DOGANLAR, 2018) e batata doce (ANDRADE et al., 2017).

2.2 ANÁLISE FATORIAL

A aplicação da análise fatorial (AF) em estudos agronômicos teve crescimento considerável graças aos avanços computacionais que facilitaram as análises e interpretação das relações entre as variáveis. A principal finalidade da AF é explicar as relações entre variáveis utilizando o menor número possível de fatores.

As etapas da AF podem ser resumidas em: obter a matriz de correlações entre todas as variáveis, determinar os fatores que são suficientes para explicar as relações das variáveis, realizar rotações dos fatores até encontrar as relações de forma mais simples e objetiva possível, calcular os escores fatoriais e obter os resultados.

Seja $X_{(px1)}$ um vetor de variáveis com vetores de média μ e variância Σ , o modelo fatorial ortogonal pressupõe que o vetor $X_{(px1)}$ depende linearmente de um conjunto de variáveis que são abstratas F_1, F_2, \dots, F_m :

$$\begin{aligned} X_1 - \mu_1 &= \ell_{11}F_1 + \ell_{12}F_2 + \dots + \ell_{1m}F_m + \varepsilon_1 \\ X_2 - \mu_2 &= \ell_{21}F_1 + \ell_{22}F_2 + \dots + \ell_{2m}F_m + \varepsilon_2 \\ &\vdots \\ X_p - \mu_p &= \ell_{p1}F_1 + \ell_{p2}F_2 + \dots + \ell_{pm}F_m + \varepsilon_p \end{aligned} \quad (1)$$

em que ℓ_{ij} é um coeficiente denominado carga da i -ésima variável sobre o j -ésimo fator. Os fatores podem ser ortogonais (não correlacionados) ou fatores oblíquos (correlacionados) Assim, segundo Johnson e Winchern (2014) o modelo matricial é descrito como:

$$\mathbf{X}_{(px1)} - \boldsymbol{\mu}_{(px1)} = \mathbf{L}_{(pxm)}\mathbf{F}_{(mx1)} + \boldsymbol{\varepsilon}_{(px1)} \quad (2)$$

As estruturas, propriedades e rotações dos fatores de análise fatorial foram descritas por (JOHNSON e WICHERN, 2014) e (GRANATE et al., 2001). O teste de esfericidade de Bartlett proposto por (BARTLETT; WHITE, 1934), conjuntamente com a medida de adequação da amostra de Kaiser Meyer Olkin (KMO) proposto por (HILL, 1980) é utilizada para a validação da análise fatorial.

O teste de esfericidade de Bartlett considera como H_0 que a matriz de correlação da população é uma matriz identidade o que sugere que o modelo fatorial é inapropriado a estatística do teste é dado por:

$$X^2 = - \left[(n - 1) - \frac{2p + 5}{6} \right] \ln|R| \quad (3)$$

Com uma distribuição qui-quadrado com $v = \frac{p(p-1)}{2}$ graus de liberdade, em que, n é o tamanho da amostra, p número de variáveis, $|R|$ é o determinate da matriz de correlação.

A medida de adequacidade da amostra Kaiser Meyer Olkin (KMO) é calculada por:

$$KMO = \frac{\sum \sum_{j \neq k} r_{jk}^2}{\sum \sum_{j \neq k} r_{jk}^2 - \sum \sum_{j \neq k} q_{jk}^2} \quad (4)$$

em que r_{jk}^2 é o quadro dos elementos da matriz de correlação original fora da diagonal, q_{jk}^2 é o quadrado das correlações parcial entre variáveis. Considerando KMO aceitável maior ou igual a 0,5.

2.3 ANÁLISE DE AGRUPAMENTOS

A análise de agrupamentos (A.G.), organiza elementos com características homogêneas dentro dos grupos e heterógena entre grupos, utilizando método de ligação, visando compreender como se relacionar as variáveis de uma amostra ou população.

A análise de agrupamento pode ser utilizada em diversos fenômenos, na identificação de sistemas de manejo (ROBOREDO et al., 2018), bioprospecção (SILVA; BARBOSA; VIDAL, 2018) e análise da água (BASSO; CARVALHO, 2007).

Um conceito essencial na A.G. é o de medida, que possa informar quanto um objeto (indivíduo) está próximo ou semelhante a outro. Essas medidas são denominadas medidas de similaridade e dissimilaridade. Na literatura existe várias medidas de dissimilaridade que podem ser usadas em variáveis quantitativas, sendo as mais conhecidas: distância Euclidiana; distâncias Euclidianas quadrada, ponderada, média e padronizada, distância de Nei; distância de Minkowski; distância de Mahalanobis e distância de Chebychev.

Distância de Mahalanobis - a distância entre dois objetos a_i e a_j , $i \neq j$ é dada por:

$$D(a_i, a_j) = (a_i - a_j)' S^{-1} (a_i - a_j) \quad (5)$$

Em que S é a matriz de covariância.

Distância de Chebychev – a distância entre dois objetos a_i e a_j é o valor máximo para todas as variáveis, é definida por:

$$D(a_i, a_j) = \max_v |a_{iv} - a_{jv}| \quad (6)$$

Os procedimentos de ligação denominados hierárquicos aglomerativos utilizam a matriz de dissimilaridade ou similaridade com o objetivo de organizar grupos homogêneos, em que os grupos vão sendo formados passo a passo até ter uma estrutura denominada dendrograma, onde é possível visualizar os grupos formados.

No método de Ligação Média a matriz $\mathbf{D} = \{d_{ij}\}$ inicial pode ser de distâncias ou semelhanças, onde os objetos são selecionados e agrupados, os elementos A e B por exemplo são similares e formam um grupo (AB) que será comparado com o grupo C, essa distância ou semelhança é medida por:

$$d_{(AB)C} = \frac{\sum_i \sum_j d_{ij}}{N_{AB} \cdot N_C} \quad (7)$$

A distância entre o indivíduo i do grupo AB para j pertencente ao grupo C é $d_{(AB)C}$, e N_{AB} , N_C o número de elementos em AB e C respectivamente.

3 MATERIAL E MÉTODOS

Este estudo foi realizado com base nos dados obtido por Guimarães (2013) na localidade da região de Bomba, com 15°18'13" de latitude Sul e de 41°17'32" de longitude Oeste, no município de Cândido Sales – Bahia. Os dados são oriundos de um experimento organizado em blocos casualizados com 28 tratamentos e três repetições, com parcelas de 8,4 m de comprimento e 3,0 m de largura, com espaço de plantio de 1,0 m entre linhas e 0,60 m entre plantas e, 26 plantas úteis, totalizando 15,6 m². No presente trabalho, foi aplicado a análise fatorial com a finalidade de reduzir as variáveis observadas nos 28 genótipos de mandioca. A variedade Sergipe aparece em dois tratamentos, um com manivas (segmentos do caule) de outras regiões (Sergipe) e outra com manivas criadas do método de multiplicação rápida (Sergipe MR) Guimarães (2013)

Na Tabela 1 estão descritos os dados referentes aos 28 genótipos de mandioca e as 23 variáveis observadas no experimento, dados que, posteriormente, foram utilizados na análise fatorial visando a redução de variáveis, e aplicação das medidas de dissimilaridade para a classificação Hierárquica, resultando na formação de grupos.

Tabela 1 - Genótipos de mandioca com todas as variáveis

Genótipos	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	V17	V18	V19	V20	V21	V22	V23
Amansa Burro	1,91	1,83	8,59	0,402	5,08	5,11	0,467	2,71	54,48	2,360	39,31	38,33	0,483	1,095	46,86	0,605	10,76	12,74	31,32	26,67	3,4	23,35	2,98
Aramaris	1,54	1,97	7,34	0,275	5,6	5,59	0,607	4,93	67,41	2,239	39,38	40,24	0,483	0,847	39,95	0,514	6,04	12,84	30,6	25,95	3,33	22,39	2,87
Bom Jardim	1,81	2,02	9,22	0,404	5,13	5,53	0,51	3,16	52,65	2,521	44,6	44,54	0,433	1,152	30,1	0,388	7,78	8,87	30,75	26,1	2,32	22,59	2
Bromadeira	1,6	1,94	6,32	0,301	5,55	5,14	0,54	3,19	64,33	1,231	26,96	24,56	0,583	1,117	27,21	0,351	4,66	9,81	31,49	26,84	2,7	23,57	2,4
Caipira	1,77	2,03	10,73	0,363	5,38	5,92	0,577	4,89	69,3	2,999	61,84	57,77	0,367	0,888	46,64	0,599	8,66	19,72	31,28	26,63	5,26	23,3	4,6
Caitite	1,98	1,75	12,78	0,548	5,06	5,23	0,433	2,89	54,23	2,125	41,31	40,25	0,783	1,148	59,76	0,774	14,15	16,64	27,56	22,91	3,82	18,33	3,07
Caravela	1,36	1,78	9,3	0,422	5,49	5,58	0,68	2,98	65,68	4,001	42,21	44,87	0,433	1,02	40,22	0,519	6,71	12,9	32,1	27,45	3,55	24,4	3,16
Kiriris	1,96	2,08	10,48	0,308	5,22	5,37	0,507	5,42	71,76	2,758	32,23	38,69	0,3	1,083	30,43	0,389	6,4	16,18	29,44	24,79	4,08	20,84	3,47
Lagoão	1,39	1,94	6,37	0,183	5,8	5,84	0,77	4,79	66,93	3,180	46,15	48,93	0,417	1,021	43,31	0,561	5,92	14,23	34,13	29,48	4,21	27,11	3,88
Lavra Velha	1,35	1,8	4,59	0,359	5,96	4,51	0,473	3,83	53,42	2,453	44,81	41,77	0,433	1,03	39,09	0,503	7,49	8,12	28,5	23,85	1,94	19,58	1,59
Malacacheta MR	1,65	1,84	7,25	0,424	5,54	5,33	0,6	0	68,48	1,370	33,44	35,73	0,383	1,164	27,05	0,351	10,58	22,5	28,18	23,53	5,34	19,16	4,37
Mulatinha	1,82	1,81	10,29	0,574	5,6	5,74	0,707	2,06	43,45	3,309	57,11	61,3	0,517	1,134	43,46	0,565	20,29	15,77	32,8	28,15	4,44	25,33	3,99
Parazinha	2,01	1,95	10,93	0,456	5,36	5,01	0,5	4,26	63,55	4,070	42,41	36,19	0,567	1,088	53,74	0,695	9,85	17,01	31,39	26,74	4,54	23,45	3,98
Peru	1,65	1,87	6,94	0,565	4,89	5,17	0,407	2,61	53,79	2,303	60,01	57,64	0,343	1,088	56,83	0,734	12,25	13,82	30,26	25,61	3,55	21,94	3,05
Poti Branca	2,42	2,08	16,24	0,502	5,31	5,26	0,52	2,26	53,95	4,110	50,96	56,67	0,47	1,137	59,38	0,768	16,52	19,44	32,28	27,63	5,32	24,63	4,72
Salangor	1,72	2,22	9,63	0,301	6,35	4,17	0,49	4,65	48,25	4,322	64,01	65,17	0,56	1,051	40,85	0,527	12,02	11,13	28,91	24,26	2,7	20,14	2,24
Sergipana	1,57	2,06	9	0,387	5,58	4,74	0,45	6,88	53,24	2,769	41,26	32,19	0,45	1,251	26,59	0,345	9,08	10,06	28,2	23,55	2,44	19,19	2,03
Sergipe	1,66	1,97	10,51	0,387	5,55	5,41	0,6	0,65	55,77	3,015	52,98	40,11	0,463	1,125	34,65	0,449	12,44	15,02	31	26,35	3,91	22,92	3,39
Sergipe MR	1,77	1,98	10,09	0,519	5,54	5,06	0,57	1,31	62,42	3,699	59,15	53,48	0,243	1,114	42,3	0,548	10,06	17,5	30,26	25,61	4,61	21,94	4,01
Simbé	1,67	1,86	9,74	0,502	5,03	5,02	0,427	2,23	56,38	3,171	64,56	56,11	0,477	0,961	38,54	0,497	11,42	14,74	30,1	25,44	3,71	21,72	3,15
Tapioqueira	1,81	1,99	6,63	0,238	5,23	5,59	0,507	3,31	67,25	2,427	40	39,11	0,51	0,772	22,34	0,285	6,72	13,74	30,28	25,63	3,54	21,97	3,03
Tussuma	2,06	2,32	13,52	0,323	6,12	5,49	0,75	5,93	49,94	3,295	62,8	59,48	0,533	1,012	59,2	0,764	12,58	12,99	30,53	25,88	3,44	22,29	3
Verdinha	1,3	1,99	7,4	0,271	5,5	5,13	0,54	2,5	71,59	2,555	58,37	53,66	0,583	1,008	58,54	0,756	9,12	22,86	31,26	26,61	6,14	23,27	5,39
2006-4	1,54	1,94	5,35	0,406	5,31	5,7	0,533	3,36	73,92	616	26,31	23,97	0,417	1,212	25,71	0,326	6,63	18,08	23,16	18,51	3,41	12,46	2,34
2006-5	1,95	1,88	7,29	0,345	5,14	5,21	0,483	1,96	74,58	664	20,51	22,11	0,5	1,172	39,54	0,513	10,08	29,27	28,85	24,2	7,11	20,06	5,91
2006-8	1,79	1,79	8,82	0,565	4,78	4,64	0,273	0	62,22	350	21,78	20,98	0,5	1,323	23,57	0,304	10,91	18,05	26,69	22,04	3,98	17,18	3,1
2006-10	1,5	2,09	6,21	0,402	5,05	4,72	0,38	3,58	74,98	907	30,33	26,81	0,4	1,221	35,19	0,456	7,63	22,93	26,82	22,17	5,07	17,35	3,96
2006-12	1,59	1,81	6,16	0,404	5,07	5,02	0,42	2,44	67,4	1,014	33,35	31,42	0,517	1,036	35,86	0,462	7,64	15,75	25,22	20,57	3,17	15,22	2,3

V1 = altura de planta, V2 = diâmetro do caule, V3 = número de manivas sementes por planta, V4 = número de hastes por plantas, V5 = comprimento das raízes tuberosas, V6 = diâmetro das raízes tuberosas, V7 = peso das raízes tuberosas, V8 = mortalidade de plantas, V9 = índice de colheita, V10 = área foliar total ao final do primeiro ciclo, V11 = porcentagem de radiação fotossinteticamente ativa absorvida pela copa entre plantas, V12 = porcentagem de radiação fotossinteticamente ativa absorvida pela copa entre linhas, V13 = potencial hídrico foliar antemanhã, V14 = potencial hídrico foliar ao meio-dia, V15 = área foliar total no momento da colheita, V16 = índice de área foliar no momento da colheita, V17 = produtividade da parte aérea, V18 = produtividade de raízes tuberosas, V19 = massa seca de raízes tuberosas, V20 = teor de amido de raízes tuberosas, V21 = produtividade de amido de raízes tuberosas, V22 = teor de farinha em raízes tuberosas e V23 = produtividade de farinha em raízes tuberosas.

O índice que indica o número de grupos no dendrograma é o índice de Ratkowsky, Segundo (Charrad et al., 2014) esse índice determina o número ótimo de grupos com base na seguinte divisão:

$$Ratkowsky = \frac{\bar{S}}{q^{1/2}} \quad (8)$$

em que:

$$\bar{S}^2 = \frac{1}{p} \sum_{j=1}^p \frac{BGSS_j}{TSS_j} \quad (8.1)$$

$$BGSS_j = \sum_{k=1}^q n_k (c_{kj} - \bar{x}_j) \quad (8.2)$$

$$TSS_j = \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2 \quad (8.3)$$

A análise foi realizada no software R versão 3.4.0, aplicando análise descritiva nos dados para verificar suas características. Posteriormente a análise fatorial indicou 6 fatores restando 13 variáveis das 23 observadas.

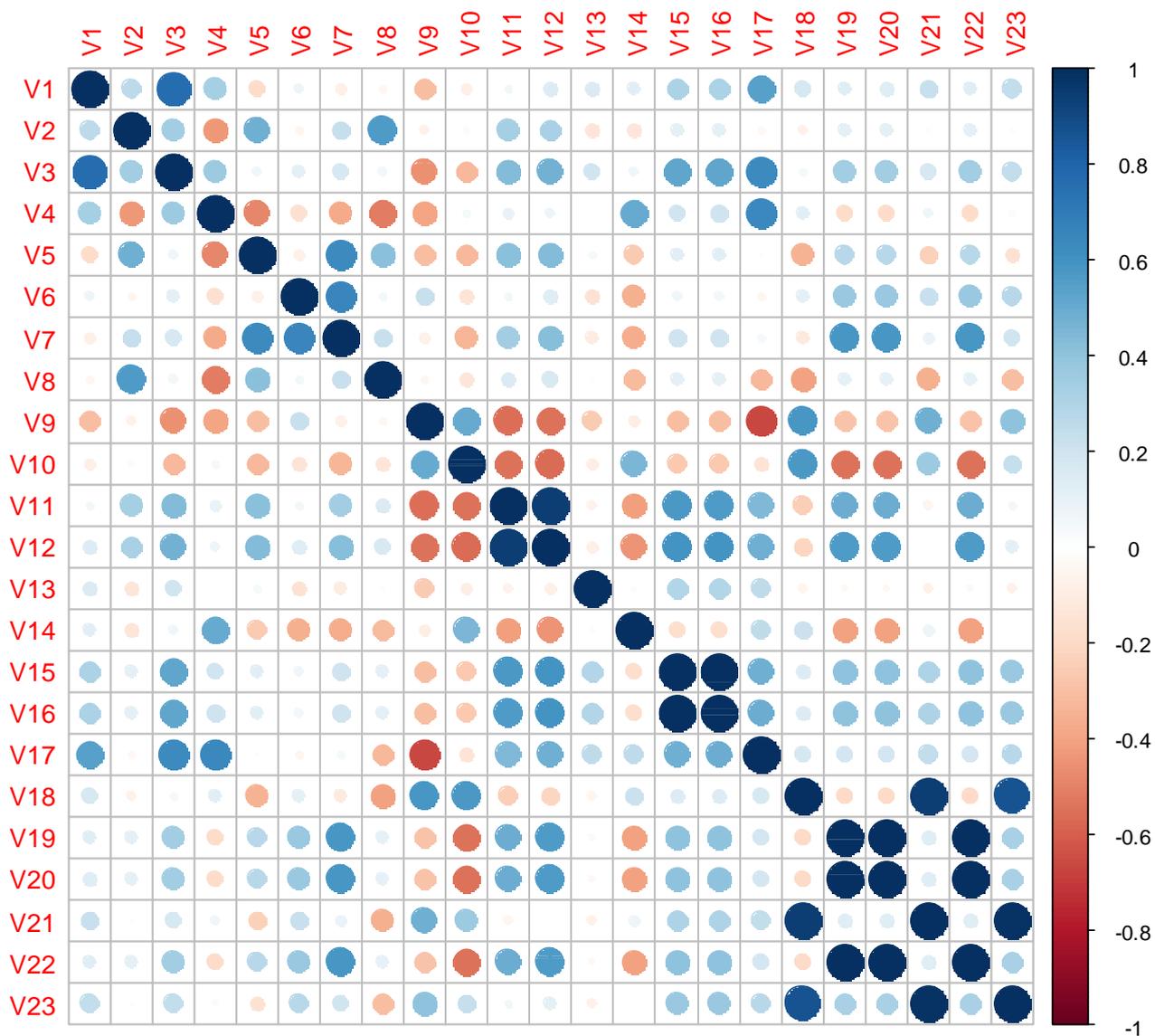
Utilizando como pacote principal o NbClust que pode ser encontrado em Charrad et al., (2014) foi realizada diversas combinações de distâncias com métodos de ligação, observando os dendrogramas e coeficiente cofenético para verificar a melhor junção de técnicas os melhores resultados foram as distâncias de Chebyshev e Mahalanobis com o método de ligação Média, outros resultados podem ser encontrados em .

4 RESULTADOS E DISCUSSÃO

Pode-se observar na Figura 1 as correlações das variáveis dos genótipos de mandioca da cidade de Cândido Sales – BA, considerando todas as variáveis. A correlação entre V1 = Altura das plantas e Número de manivas-sementes por planta (V3), e entre (V1) e produtividade da parte aérea (V17) foi positiva, ou seja, uma relação crescente entre (V1) com (V3) e (V17). As variáveis Diâmetro do caule (V2) e

mortalidade de plantas (V8) apresentaram correlação positiva, ou seja, a medida que o diâmetro do caule cresce a mortalidade aumenta.

Figura 1 - Gráfico de correlações das variáveis observadas nos genótipos de mandioca da cidade de Cândido Sales-BA.



V1 = altura de planta, V2 = diâmetro do caule, V3 = número de manivas sementes por planta, V4 = número de hastes por plantas, V5 = comprimento das raízes tuberosas, V6 = diâmetro das raízes tuberosas, V7 = peso das raízes tuberosas, V8 = mortalidade de plantas, V9 = índice de colheita, V10 = área foliar total ao final do primeiro ciclo, V11 = porcentagem de radiação fotossinteticamente ativa absorvida pela copa entre plantas, V12 = porcentagem de radiação fotossinteticamente ativa absorvida pela copa entre linhas, V13 = potencial hídrico foliar antemanhã, V14 = potencial hídrico foliar ao meio-dia, V15 = área foliar total no momento da colheita, V16 = índice de área foliar no momento da colheita, V17 = produtividade da parte aérea, V18 = produtividade de raízes tuberosas, V19 = massa seca de raízes tuberosas, V20 = teor de amido de raízes tuberosas, V21 = produtividade de amido de raízes tuberosas, V22 = teor de farinha em raízes tuberosas e V23 = produtividade de farinha em raízes tuberosas.

De acordo com a Tabela 3 as variáveis foram reduzidas para seis fatores com treze variáveis, com proporção da variância explicada acumulada de 92%, ou seja, com treze variáveis é possível explicar a variação dos dados.

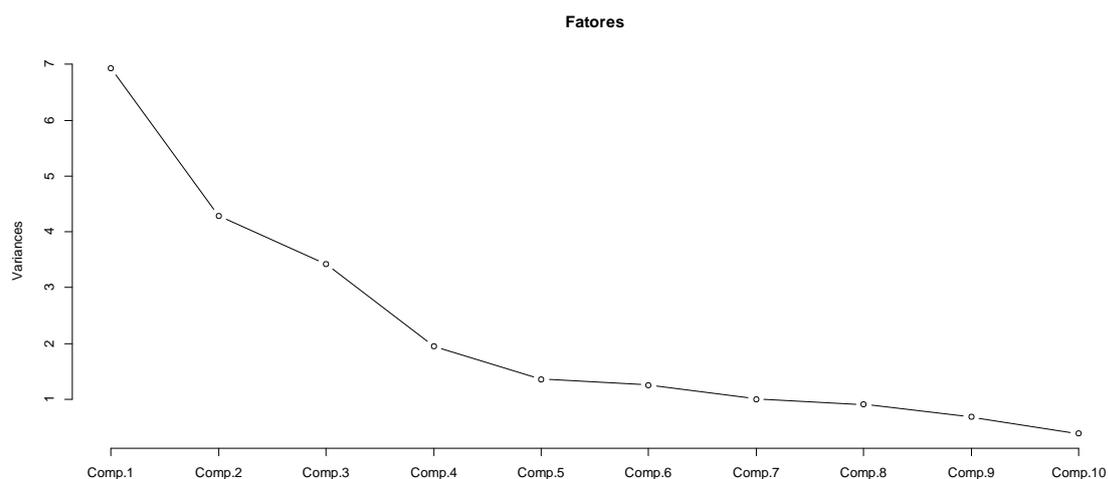
Utilizando a rotação varimax nos fatores para facilitar a interpretação da análise fatorial, posteriormente foi aplicado o teste de esfericidade de Bartlett que resultou em um valor $p < 0,00$ com $\alpha=0,05$, esse resultado indica que a matriz de correlação não era identidade como a Figura 1 demonstra. O critério de Kayser Meyer Olkin resultou em um $KMO = 0,51$ que para (HAIR et al., 1998) é considerado aceitável, com isso a aplicação do Análise fatorial foi correta.

Tabela 2 - Autovalores, proporção da variância explicada e proporção da variância explicada acumulada, para os fatores com a rotação varimax.

Fator	Autovalor	Proporção da variância explicada	Proporção da variância explicada acumulada
1	4,56	0,24	0,24
2	3,89	0,20	0,44
3	3,75	0,19	0,63
4	2,81	0,15	0,78
5	2,73	0,14	0,92
6	1,48	0,08	1

Confirmando o que a Tabela 3 demonstra, na Figura 2 é possível observa a redução gradativa nos fatores para a variância explicada após o fator 6.

Figura 2 - Gráfico das Variâncias dos fatores com a rotação varimax



Na Tabela 4 observa-se a análise descritiva das variáveis que ficaram no estudo após análise fatorial com a rotação varimax, com o menor valor encontrado 0,27 para variável peso das raízes tuberosas (V7). O maior valor 74,98 foi observado para variável índice de colheita (V9). A menor mediana 0,51 foi encontrada para variável, peso das raízes tuberosas (V7), enquanto a maior mediana 62,98 para variável índice de colheita (V9).

Tabela 3 - Resultado da análise descritiva do banco de dados para variáveis selecionadas na análise fatorial nos genótipos de mandioca da cidade de Cândido Sales BA.

Variáveis	V1	V2	V5	V6	V7	V9	V17	V18	V19	V20	V21	V22	V23
Min.	1,30	1,75	4,78	4,78	0,27	43,45	4,66	8,12	23,16	18,51	1,94	12,46	1,59
1º Qua.	1,56	1,84	5,12	5,12	0,46	53,91	7,30	12,88	28,43	23,77	3,39	19,48	2,75
Mediana	1,70	1,95	5,37	5,37	0,51	62,98	9,49	15,38	30,27	25,62	3,77	21,95	3,13
Média	1,72	1,95	5,40	5,40	0,53	61,48	9,80	15,81	29,76	25,11	3,97	21,27	3,36
3º Qua..	1,84	2,02	5,56	5,56	0,58	67,68	11,57	18,06	31,29	26,64	4,56	23,31	3,99
Máx.	2,42	2,32	6,35	6,35	0,77	74,98	20,29	29,27	34,13	29,48	7,11	27,11	5,91

V1 = altura de planta, V2 = diâmetro do caule, V6 = diâmetro das raízes tuberosas, V7 = peso das raízes tuberosas, V9 = índice de colheita, V17 = produtividade da parte aérea, V18 = produtividade de raízes tuberosas, V19 = massa seca de raízes tuberosas, V20 = teor de amido de raízes tuberosas, V21 = produtividade de amido de raízes tuberosas, V22 = teor de farinha em raízes tuberosas e V23 = produtividade de farinha em raízes tuberosas.

A

Figura 3 demonstra as correlações entre as variáveis selecionadas na análise fatorial. A (V1) é fortemente correlacionada positivamente com a V17, indicando que a medida que a altura da planta aumenta (V1), cresce o número de manivas sementes

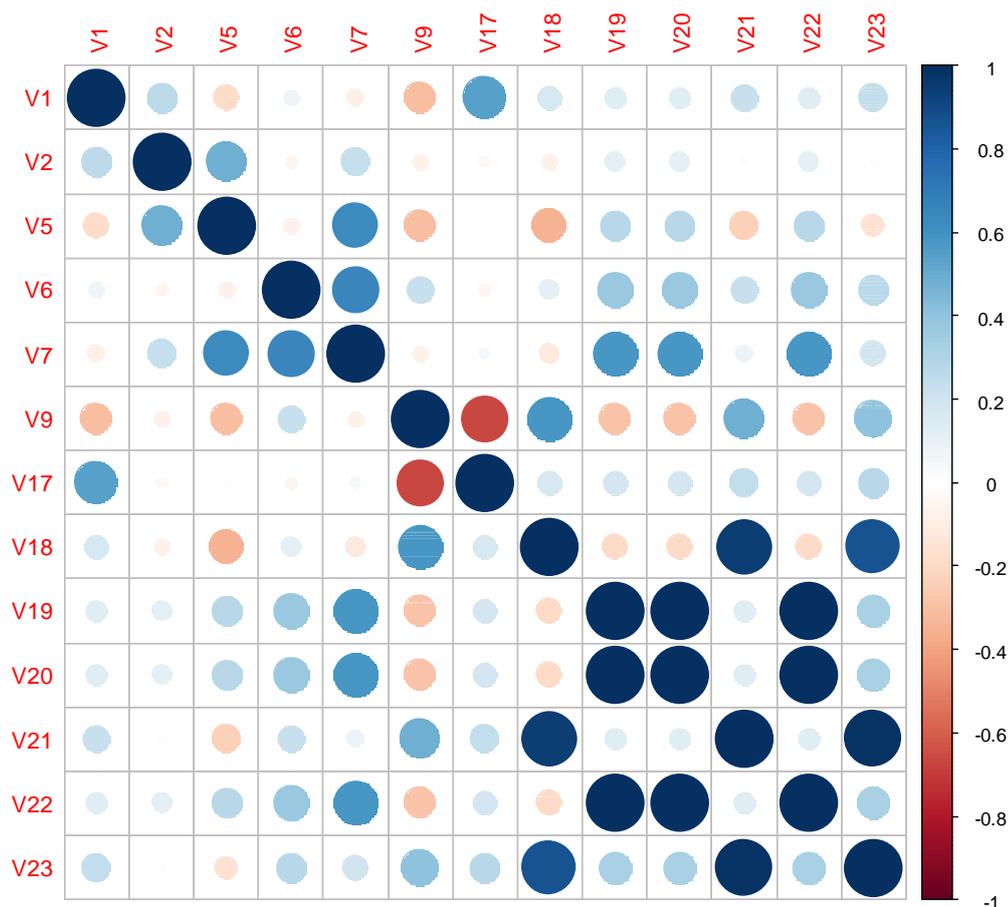
por planta (V17). A variável diâmetro do caule (V2) é correlacionada positivamente com comprimento da raiz tuberosa (V5).

Ainda na

Figura 3, a variável índice de colheita (V9) se correlaciona negativamente com a variável produtividade da parte aérea (V17), indicando que a medida que a produtividade da parte aérea aumenta, diminui o índice de colheita.

As seis últimas variáveis possuem um grau de correlação positiva (0,8 entre 1,0) essas relações possuem esse comportamento devido a relação biológica das variáveis. Outro fator que explica essa relação é que as variáveis contêm apenas características das raízes das espécies de mandioca.

Figura 3 - Gráfico de correlação das variáveis selecionas na análise fatorial dos dados de genótipos de mandioca na Cidade de Cândido Sales – BA.



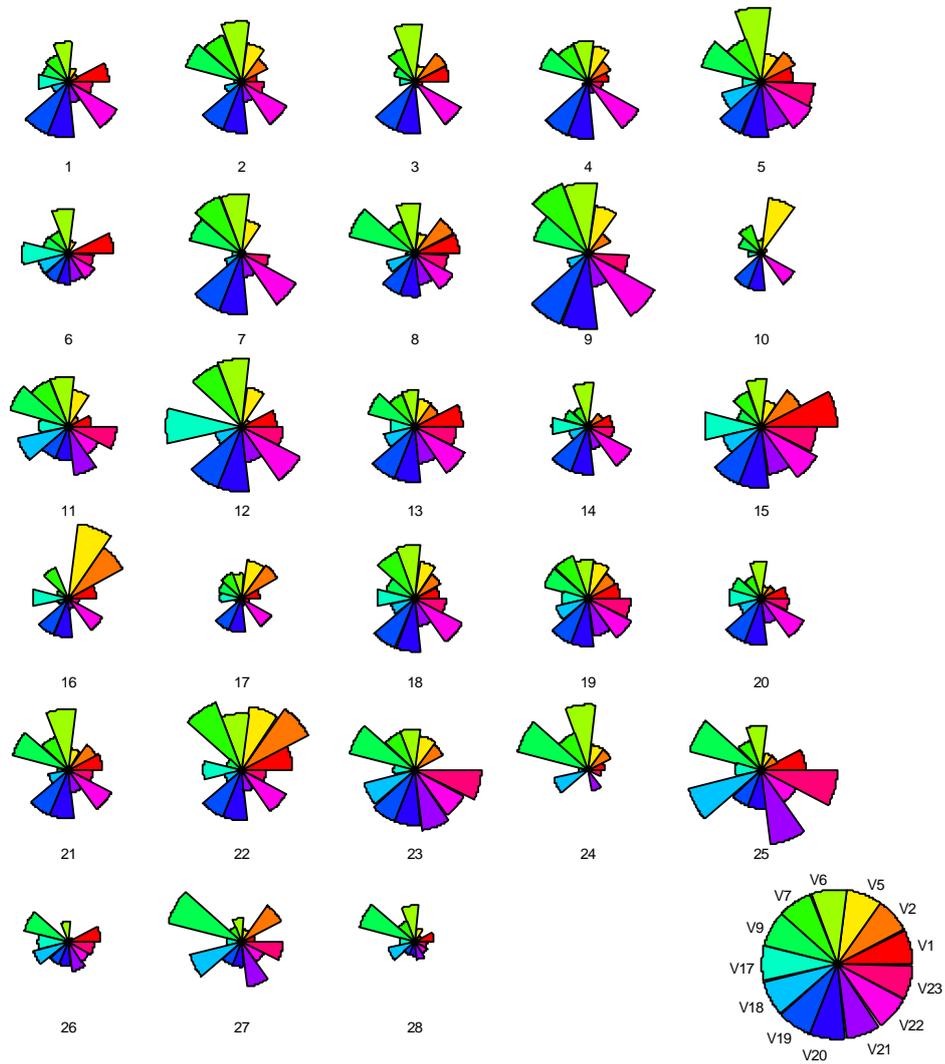
V1 = altura de planta, V2 = diâmetro do caule, V6 = diâmetro das raízes tuberosas, V7 = peso das raízes tuberosas, V9 = índice de colheita, V17 = produtividade da parte aérea, V18 = produtividade de raízes tuberosas, V19 = massa seca de raízes tuberosas, V20 = teor de amido de raízes tuberosas, V21 = produtividade de amido de raízes tuberosas, V22 = teor de farinha em raízes tuberosas e V23 = produtividade de farinha em raízes tuberosas

Na Figura 4 foi possível ver como as variáveis estão distribuídas nos genótipos de mandioca que estão enumerados de 1 a 28, os quais estão descritos na Tabela 1. Cada variável (V1, V2, V5, V6, V7, V9, V17, V18, V19, V20, V21, V22, V23) pode ser identificadas por cor. Visualmente não foi tão simples definir quais genótipos possuem características mais similares a ponto de formar grupos com elementos homogêneos e heterogêneos entre grupos.

Observou-se que os genótipos 1 ao 5 respectivamente Amansa Burro, Aramaris, Bom Jardim, Bromadeira e Caipira possuem as características de massa seca e teor de amido que são respectivamente as variáveis V19 e V20. Os genótipos Lagoão (9), Lavra Velha (10), Verdinha (23) e 2006 - 4 (24) tem a característica similar de ter uma baixa altura de planta (V1). Ainda na Figura 4 temos as variáveis, V1 = altura de planta, V2 = diâmetro do caule, V6 = diâmetro das raízes tuberosas, V7 = peso das raízes tuberosas, V9 = índice de colheita, V17 = produtividade da parte aérea, V18 = produtividade de raízes tuberosas, V19 = massa seca de raízes tuberosas, V20 = teor de amido de raízes tuberosas, V21 = produtividade de amido de raízes tuberosas, V22 = teor de farinha em raízes tuberosas e V23 = produtividade de farinha em raízes tuberosas,

foi aplicado as técnicas de agrupamentos multivariados que vão ser descritas nas Figura 5 e Figura 6 para facilitar as formações dos grupos de genótipos de mandioca.

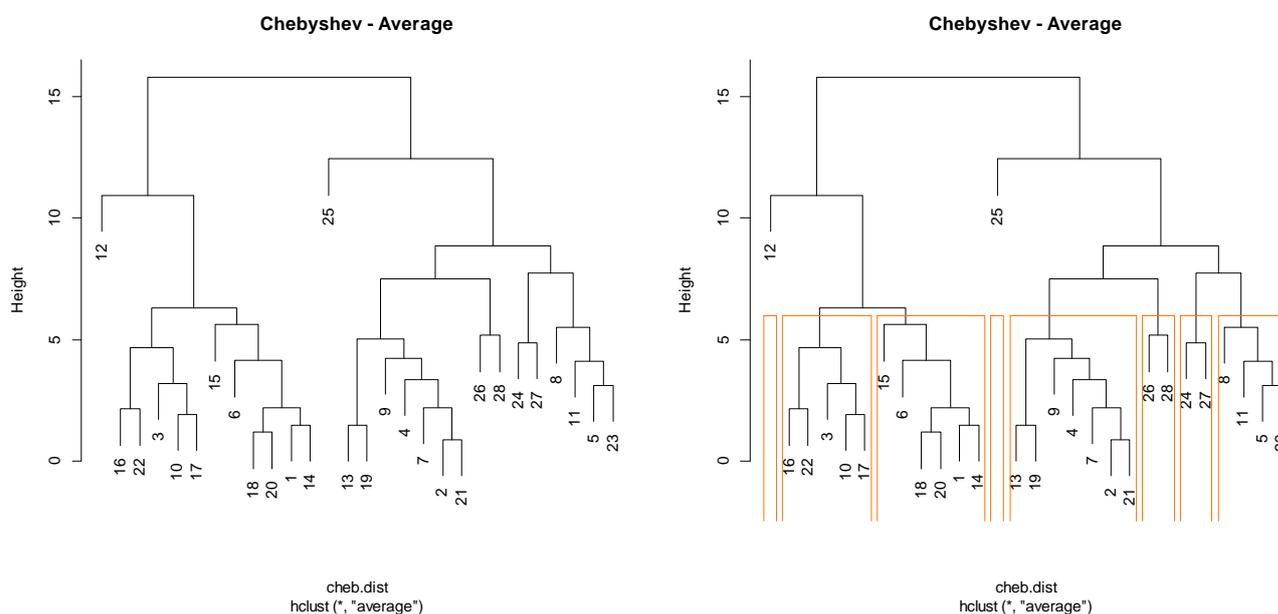
Figura 4 - Gráfico de Distribuição das variáveis predominantes em cada genótipos de mandioca na cidade de Cândido Sales – BA.



O dendrograma da

Figura 5 foi o resultado da distância de Chebyshev combinado com o método de ligação média (Average). Como critério para a formação dos grupos o índice de Ratkowsky que indicou oito grupos com o coeficiente cofenético de 0,80. Os grupos podem ser vistos na Tabela 5. A utilização de métodos de agrupamentos após a redução de variáveis por análise fatorial pode ser vista em (ALEIXO; DE SOUZA; FERRAUDO, 2007)

Figura 5 - Dendrograma gerado da matriz de dissimilaridade da distância de Chebyshev utilizando o método de ligação média com grupos sugeridos no índice de Ratkowsky.



De acordo com as características descritas na Figura 4 e Figura 5 os grupos possuem as seguintes formações e características: G1, formado com o genótipo Mulatinha, possui características balanceadas entre parte superior da planta e inferior, na produção da parte aérea da mandioca e raiz com grande diâmetro e teor de amido na raiz, com pouco diâmetro de caule e baixo teor de matéria seca. G2, formado com os genótipos Salangor (16), Tussuma (22), Bom jardim (3), Lavra velha (10) e Sergipana (17) apresentam bom teor de farinha nas raízes tuberosas, mesmo com uma baixa produtividade de raízes tuberosas, pouco teor de amido e baixa produtividade da parte aérea.

No G3, formado com os genótipos Poti Branco (15), Caitite (6), Sergipe (18), Simbé (20), Amansa burro (1) e Peru (14), tem como principais características o diâmetro da raiz tuberosas, a porcentagem de massa seca nas raízes tuberosas, o teor de amido nas raízes tuberosas, o teor de farinha em raízes tuberosas, com baixos valores de diâmetro do caule e o comprimento das raízes tuberosas. O G4, formado apenas pelo genótipo 2006-5 (25), possui as características semelhantes do grupo G3, visto que o genótipo 2006-5 (25) foi melhorado geneticamente na Embrapa. O G5 constituído pelos genótipos Parazinha (13), Sergipe MR (19), Lagoão (9), Bromadeira (4), Caravela (7), Aramaris (2) e Tapioqueira (21), possui valores altos de diâmetro

das raízes tuberosas, peso de raízes tuberosas, índice de colheita, porcentagem de massa seca nas raízes tuberosas, teor de amido nas raízes tuberosas e teor de farinha nas raízes tuberosas, com baixo valores de diâmetro do caule, produtividade de raízes tuberosas e produtividade de parte aérea. O G6, formado com os genótipos 2006-8 (26) e 2006-12 (28), tem características de bom índice de colheita e produtividade de raízes tuberosas com baixos valores, principalmente, de diâmetro do caule e de comprimento das raízes tuberosas.

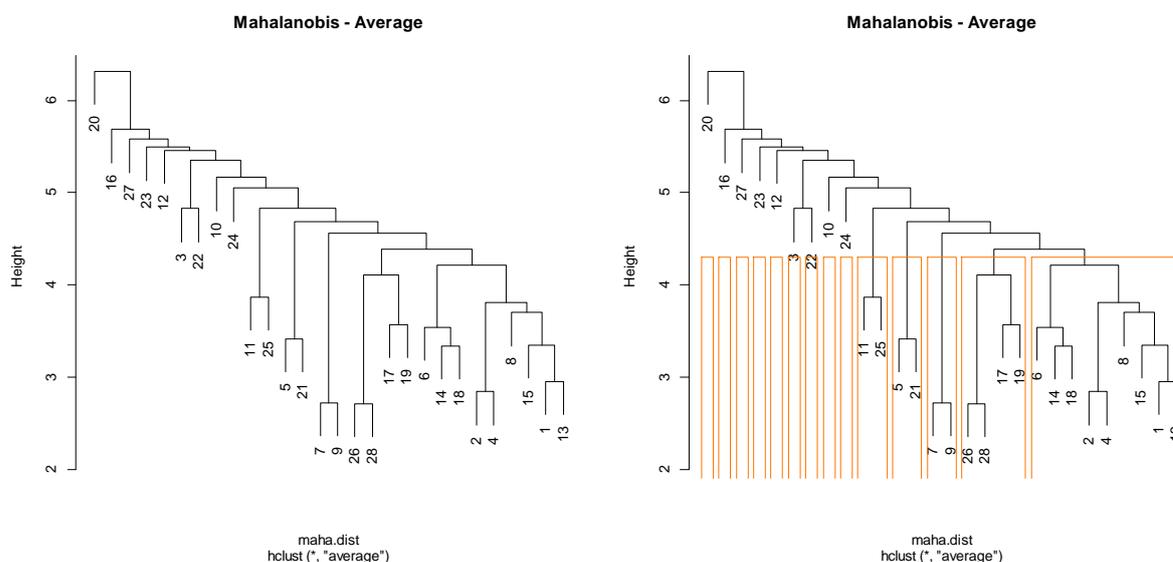
O G7, composto pelos genótipos 2006-4 (24) e 2006-12 (27), apresentou melhores características para índice de colheita e produtividade da raiz tuberosas, pouca característica de altura de planta e produtividade de parte aérea. O último grupo, o G8, formado por Kiriris (8), Malacacheta MR (11), Caipira (5) e Verdinha (23), teve como característica um baixo valor de produtividade da parte aérea e a presença de todas as demais variáveis.

Tabela 4 - Grupos gerados da matriz de dissimilaridade da distância de Chebyshev, utilizando o método de ligação média com grupos sugeridos no índice de Ratkowsky.

Grupos	Genótipos	Característica Dominante	Característica Não Dominante
G1	12	V1; V5; V6; V7; V17; V18; V19; V20; V21 V22; V23.	V2; V9.
G2	16; 22; 3; 10; 17	V1; V2; V5; V6; V7; V19; V20; V22; V23.	V9; V17; V18; V21.
G3	15; 6; 18; 20; 1; 14	V1; V6; V7; V9; V17; V18; V19; V20; V21; V22; V23.	V2; V5.
G4	25	V1; V6; V7; V9; V17; V18; V19; V20; V21; V22; V23.	V2; V5.
G5	13; 19; 9; 4; 7; 2; 21	V1; V5 V6; V7; V9; V19; V20; V21; V22; V23.	V2; V17; V18.
G6	26; 28	V1; V6; V7; V9; V17; V18; V19; V20; V21; V22; V23	V2; V5.
G7	24; 27	V2; V5 V6; V7; V9; V18; V19; V20; V21; V22; V23.	V1; V17.
G8	8; 11; 5; 23	V1; V2; V5 V6; V7; V9; V18; V19; V20; V21; V22; V23.	V17.

O dendrograma da Figura 6 foi o resultado da distância de Mahalanobis combinado com o método de ligação média (Average). Como critério para a formação dos grupos o índice de Ratkowsky que indicou 14 grupos com o coeficiente cofenético de 0,79. Na configuração sugerida da distância de Mahalanobis existe uma quantidade de grupos unitários maior comparado com a distância de Chebyshev.

Figura 6 - Dendrograma gerado da matriz de dissimilaridade da distância de Mahalanobis utilizando o método de ligação média com grupos sugeridos no índice de Ratkowsky.



Observando o coeficiente cofenético do dendrograma na Figura 5 de 0,80 e o coeficiente cofenético na Figura 6 de 0,79 são similares, porém o dendrograma da distância de Chebyshev com o método de ligação média demonstra ter ajustado de forma mais coerente as informações do grupo, para definir os grupos divergentes de genótipos de mandioca.

Tabela 5 - Grupos gerados da matriz de dissimilaridade da distância de Mahalanobis, utilizando o método de ligação média com grupos sugeridos no índice de Ratkowsky.

Grupos	Genótipos
G1	20
G2	16
G3	27
G4	23
G5	12
G6	3
G7	22
G8	10
G9	24
G10	11; 25
G11	5; 21
G12	7;9
G13	26; 28; 17; 19
G14	6; 14; 18; 2; 4; 8; 15; 1; 13

5 CONCLUSÕES

- O uso das distâncias de Chebyshev e Mahalanobis com o método de ligação média apresenta grupos dos mais variados, com genótipos de características similares dentro dos grupos e divergentes entre grupos demonstrando que é possível cruzar esses grupos em programas de melhoramento afim de obter genótipos mais promissores para a mandioca.
- Nessa situação experimental a distância de Chebyshev com o método de ligação média e o índice de Ratkowsky demonstrou ser melhor para definir os grupos e suas características
- Os grupos G3, G4, G6 não são indicados para cruzamentos entre si, pois apresentam características similares, G1, G2, G5 e G8 são divergentes entre si e com os grupos G3, G4 e G6.
- A análise multivariada demonstrou ser uma excelente ferramenta no auxílio preditivo dos genótipos de mandioca para programas de melhoramentos e de conservação de espécies.

6 Referências

ABUZAYED, M.; FRARY, A.; DOGANLAR, S. Genetic diversity of some Palestinian and Turkish olive (*Olea europaea* L.) germplasm determined with SSR markers. **IUG Journal of Natural Studies**, Gaza, 2018. 10 - 17.

ALEIXO, S. S.; DE SOUZA, G.; FERRAUDO, A. S. Técnicas de análise multivariada na determinação de grupos homogêneos de produtores de leite. **Revista Brasileira de Zootecnia**, Viçosa, 7 Maio 2007. 2168-2175.

ANDRADE, E. K. V. D. et al. ACTA SCIENTIARUM. AGRONOMY. **Genetic dissimilarity among sweet potato genotypes using**, Maringá, 2017. 447-455.

BARTLETT, M. S.; WHITE, F. P. A note on tests of significance in multivariate analysis. **Mathematical Proceedings of the Cambridge Philosophical Society**, Cambridge, 1939. 180 - 185.

BASSO, E. R.; CARVALHO, S. L. D. AVALIAÇÃO DA QUALIDADE DA ÁGUA EM DUAS REPRESAS E UMA LAGOA NO MUNICÍPIO DE ILHA SOLTEIRA (SP). **Holos Environment**, Rio Claro, 13 Julho 2007. 16 - 29.

CHARRAD, M. et al. NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. **Journal of Statistical Software**, Innsbruck, 4 Outubro 2014. 1 - 36.

EVERITT, B.; HOTHORN, T. An Introduction to Applied Multivariate Analysis with R. In: EVERITT, B.; HOTHORN, T. **An Introduction to Applied Multivariate Analysis with R**. [S.l.]: Springer, 2011. Cap. 7.

FAO. Produzir mais com menos. **Food and Agriculture Organization the United Nations**, 2013. Disponível em: <<http://www.fao.org/ag/save-and-grow/cassava/pt/index.html>>. Acesso em: 27 2017 2017.

FONSECA, F. A. D. et al. Divergência genética em café conilon. **PESQUISA AGROPECUÁRIA BRASILEIRA**, Brasília, Abril 2006. 599-605.

GRANATE, M. J. et al. A análise de fatores na predição de ganhos por seleção em milho (*Zea mays* L.). **Acta Scientiarum. Agronomy**, Maringá, 2001. 1271-1279.

GUIMARÃES, D. G. **AVALIAÇÃO DE GENOTIPOS DE MANDIOCA EM CÂNDIDO SALES – BA**. Vitória da Conquista: [s.n.], 2013. 1 - 101 p. Disponível em: <http://www.uesb.br/ppgagronomia/_old_/banco-de-dissertacoes/2013/Douglas-Goncalves-Guimaraes.pdf>. Acesso em: 28 Junho 2018.

HAIR, J. F. J. et al. **Multivariate Data Analysis**. Tradução de Adonai Schlup San't Anna e Anselmo Chaves Neto. [S.l.]: [s.n.], 1998.

HILL, R. S. A STOPPING RULE FOR PARTITIONING DENDROGRAMS. **The University of Chicago Press Journals**, Chicago, 1980.

IBGE. www.ibge.gov.br. **sidra.ibge.gov.br**, 2016. Disponível em: <<https://sidra.ibge.gov.br/tabela/5457#resultado>>. Acesso em: 1 Novembro 2017.

JOHNSON, R.; WICHERN, D. Applied Multivariate Statistical Analysis. In: JOHNSON, R.; WICHERN, D. **Applied Multivariate Statistical Analysis**. 6. ed. New Jersey: Pearson, 2014. Cap. 9, p. 482.

JOHNSON, R.; WICHERN, D. **Applied Multivariate Statistical Analysis**. 6. ed. Nova Jersey: Pearson, 2014.

LOPES, A. C. et al. COMPLEMENTAÇÃO DA IRRIGAÇÃO E ÉPOCAS DE COLHEITA DE MANDIOCA CV. COQUEIRO NO PLANALTO DE CONQUISTA, BA. **CIÊNCIA E AGROTECNOLOGIA**, Lavras, Junho 2010. 579-587.

MAHAJAN, R.; JAVED, A.; KAPOOR, N. Characterization of genetic diversity of wild pomegranate collected from Himachal Pradesh, India. **Annals of Plant Sciences**, Andhra Pradesh, 30 jan. 2018. 2042-2046.

ROBOREDO, D. et al. CLUSTERIZAÇÃO DE SISTEMAS DE MANEJOS E A CONSTRUÇÃO DE INDICADORES DE SUSTENTABILIDADE UTILIZANDO A METODOLOGIA MESMIS NO TERRITÓRIO PORTAL DA AMAZÔNIA. **RA'E GA: O ESPAÇO GEOGRÁFICO EM ANÁLISE**, Curitiba, Fevereiro 2018. 23 - 42.

SILVA, T. D. C. D. M.; BARBOSA, N. P. S. R.; VIDAL, E. Bioprospecção de leveduras oleaginosas do Agreste Pernambucano: síntese de lipídios para a produção de biodiesel a partir do resíduo glicérol bruto. **Journal of Environmental Analysis and Progress**, Recife, 30 Janeiro 2018. 92 - 99.

ZAGO, B. W. et al. Morphological diversity of cassava accessions of the south-central mesoregion of the State of Mato Grosso, Brazil. **Genetics and Molecular Research**, Ribeirão Preto, 17 Agosto 2017. 10.