

RAQUEL CORDEIRO SANTOS

**ANÁLISE DE COMPONENTES PRINCIPAIS E ANÁLISE DE FATORES EM
ACESSOS DE *Capsicum annuum* L. NO ESTUDO DE VARIABILIDADE
GENÉTICA**

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

Orientador: Paulo Roberto Cecon

**VIÇOSA - MINAS GERAIS
2023**

**Ficha catalográfica elaborada pela Biblioteca Central da Universidade
Federal de Viçosa - Campus Viçosa**

T

Santos, Raquel Cordeiro, 1989-
S237a Análise de componentes principais e análise de fatores em
2023 acessos de *Capsicum annuum* L. no estudo de variabilidade
genética / Raquel Cordeiro Santos. – Viçosa, MG, 2023.
1 dissertação eletrônica (47 f.): il. (algumas color.).

Orientador: Paulo Roberto Cecon.

Dissertação (mestrado) - Universidade Federal de Viçosa,
Departamento de Estatística, 2023.

Referências bibliográficas: f. 44-47.

DOI: <https://doi.org/10.47328/ufvbbt.2023.632>

Modo de acesso: World Wide Web.

1. Análise multivariada. 2. Pimenta - Melhoramento
genético - Métodos estatísticos. I. Cecon, Paulo Roberto, 1955-
II. Universidade Federal de Viçosa. Departamento de Estatística.
Programa de Pós-Graduação em Estatística Aplicada e
Biometria. III. Título.

CDD 22. ed. 519.535


RAQUEL CORDEIRO SANTOS

**ANÁLISE DE COMPONENTES PRINCIPAIS E ANÁLISE DE FATORES EM
ACESSOS DE *Capsicum annuum* L. NO ESTUDO DE VARIABILIDADE
GENÉTICA**


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

APROVADA: 30 de agosto de 2023

Assentimento:

Documento assinado digitalmente
 RAQUEL CORDEIRO SANTOS
Data: 20/10/2023 16:34:19-0300
Verifique em <https://validar.iti.gov.br>

Raquel Cordeiro Santos
Autor

Documento assinado digitalmente
 PAULO ROBERTO CECON
Data: 23/10/2023 08:17:05-0300
Verifique em <https://validar.iti.gov.br>

Paulo Roberto Cecon
Orientador

*Aos que se entregam de coração e alma
para sonhar e realizar.*

AGRADECIMENTOS

Primeiramente à Deus, pois sem Ele nada seria possível.

Aos meus pais, que lhes devo a vida e cada passo que continuo a dar rumo aos meus sonhos.

Ao meu irmão Rafael, por acreditar e apoiar minhas escolhas.

Ao meu orientador Paulo Roberto Cecon, que com carinho, orientou-me para além desse trabalho e entendeu minhas decisões e meu momento de transição profissional.

À Universidade Federal de Viçosa e ao PPESTBIO pela oportunidade de realizar a pós-graduação e aperfeiçoamento de meus conhecimentos.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pela concessão da bolsa de estudos. O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

À Empresa Brasileira de Correios e Telégrafos que oportunamente, deu-me o tempo necessário à realização do mestrado.

Aos colegas do mestrado, que sempre foram companheiros, mesmo em tempos tão difíceis.

Aos membros da banca examinadora pela disponibilidade e sugestões valiosas para este trabalho.

Aos professores do PPESTBIO e todos aqueles que fizeram parte da minha trajetória acadêmica, por serem disseminadores do saber e por me fazer chegar aonde cheguei.

Aos amigos e pessoas próximas que compreenderam minhas ausências e necessidade enquanto me dedicava aos estudos e escrita desse trabalho.

Aos pesquisadores dos grupos de estudos LAPEA e GESTBIO por serem amigos, incentivadores e animados para fazer ciência.

Muito obrigada!

*“Tudo vale a pena
quando a alma não é pequena.”*

(Fernando Pessoa)

RESUMO

SANTOS, Raquel Cordeiro, M.Sc., Universidade Federal de Viçosa, agosto de 2023. **Análise de componentes principais e análise de fatores em acessos de *Capsicum annuum* L. no estudo de variabilidade genética.** Orientador: Paulo Roberto Cecon.

As pimenteiras pertencem a família Solanácea e ao gênero *Capsicum*, a qual representa grande variabilidade de plantas. As pimentas são consumidas em todo o mundo e em razão de seus diversos usos e demanda de mercado é necessário buscar opções de maior qualidade e produtividade, o que pode ser feito pela escolha de genótipos superiores. O objetivo dessa pesquisa foi estudar a variabilidade genética de acessos de pimenta (*C. annuum*) por meio de técnicas de análise de componentes principais e análise de fatores. Utilizando o delineamento experimental inteiramente casualizado, com quatro repetições foram analisados nove acessos de *Capsicum annuum* L., são eles: Pimenta Vulcão, Pimenta Cayene, Pimenta Peter, Pimenta Picante para vaso, Pimenta Jamaica Yellow, Pimenta Doce Italiana, Pimentão Quadrado, Pimentão Cascadura Ikeda e Pimentão Rubi Gigante. As características avaliadas foram: peso total do fruto (PT, g), comprimento do fruto (COM, mm), largura do fruto (LAR, mm), espessura do pericarpo (ESP, mm), número de sementes por fruto (NS), massa da matéria total do fruto fresca (MF, g), Massa da matéria total do fruto maduro seco (MS, g), Teor de sólidos solúveis (BRIX), Teor de Vitamina C (VIT, $\text{mg}100_{\text{g}}^{-1}$). Foi aplicada a técnica de análise multivariada denominada de componentes principais (ACP), que permitiu reduzir a dimensão da amostra, pela qual foram retidos dois componentes que juntos explicaram mais de 90% da variabilidade contida nos dados. A análise de divergência genética entre os acessos foi realizada por meio da técnica de análise multivariada denominada de análise de fatores (AF), seguido da rotação Varimax, levando a identificação de três grupos geneticamente divergentes o que permite abordagens mais direcionadas para otimizar as características-chave relacionadas ao mercado de pimentas.

Palavras-chave: Análise multivariada. Variabilidade genética. Pimenta.

ABSTRACT

SANTOS, Raquel Cordeiro, M.Sc., Universidade Federal de Viçosa, August 2023. **Principal component analysis and factor analysis in accessions of *Capsicum annuum* L. for the study of genetic variability.** Advisor: Paulo Roberto Cecon.

Pepper plants belong to the Solanaceae family and the *Capsicum* genus, which exhibits a wide range of plant diversity. Peppers are grown worldwide, and due to their various uses and market demand, it is necessary to explore options for higher quality and productivity, which can be achieved by selecting superior genotypes. The objective of this research was to study the genetic variability of pepper accessions (*C. annuum*) through principal component analysis and factor analysis techniques. Using a completely randomized experimental design with four replications, nine accessions of *Capsicum annuum* L. were analyzed, including: Volcano Pepper, Cayenne Pepper, Peter Pepper, Potted Hot Pepper, Jamaica Yellow Pepper, Italian Sweet Pepper, Square Bell Pepper, Ikeda Cascadura Bell Pepper, and Giant Ruby Bell Pepper. The evaluated characteristics included: total fruit weight (TF, g), fruit length (FL, mm), fruit width (FW, mm), pericarp thickness (PT, mm), number of seeds per fruit (NS), fresh total fruit matter weight (FM, g), mature dry total fruit matter weight (DM, g), soluble solids content (BRIX), and Vitamin C content (VIT, mg per 100g). The multivariate analysis technique known as principal component analysis (PCA) was applied, which allowed for the reduction of sample dimensionality, retaining two components that together explained more than 90% of the variability in the data. The genetic divergence analysis among the accessions was carried out through the multivariate technique called factor analysis (FA), followed by Varimax rotation, leading to the identification of three genetically divergent groups. This allows for more targeted approaches to optimize key characteristics related to the pepper market.

Keywords: Multivariate analysis. Genetic variability. Pepper.

LISTA DE ILUSTRAÇÕES

Figura 1- Gráfico <i>scree-plot</i> para determinação do número de componentes principais a serem retidos com base nos autovalores obtidos.....	32
Figura 2 - Importância das variáveis para o primeiro componente principal.....	34
Figura 3 - Importância das variáveis para o segundo componente principal.....	34
Figura 4 – Gráfico de correlação entre os dois primeiros componentes e as variáveis analisadas	35
Figura 5 - Gráfico <i>scree-plot</i> para determinação do número de fatores a serem retidos, com linha de referência em 1.....	38
Figura 6 – Dispersão gráfica dos acessos em relação aos dois primeiros fatores.....	41

LISTA DE TABELAS

Tabela 1 – Análise descritiva para as variáveis estudadas.....	30
Tabela 2 – Dados originais provenientes do experimento.....	31
Tabela 3 – Dados padronizados.....	31
Tabela 4 - Informações importantes da análise de componentes principais.....	32
Tabela 5 - Coeficiente associados às características e seus coeficientes de correlação com os dois primeiros componentes principais.....	33
Tabela 6 – Informações importantes da análise de fatores.....	37
Tabela 7 – Cargas fatoriais antes e depois da rotação ortogonal.....	39
Tabela 8 – Variância explicada pelos fatores.....	40

LISTA DE SIGLAS E ABREVIATURAS

ACP Análise de Componentes Principais
CAPES Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
COM Comprimento do fruto maduro
CP componente principal
ESP Espessura do pericarpo
AF Análise de Fatores
FAO Organização das Nações Unidas para a Alimentação e a Agricultura
GESTBIO Grupo de Estudos em Estatística aplicada e Biometria
LAR Largura do fruto maduro
LAPEA Laboratório de Análises e Pesquisas em Estatística Aplicada
NS Número de sementes por frutos
PPESTBIO Programa de Pós-Graduação em Estatística aplicada e Biometria
PT Peso total do fruto
TRAT Tratamento
UFV Universidade Federal de Viçosa
VAR Variável
VIT Teor de vitamina C

SUMÁRIO

1. INTRODUÇÃO	12
2. REVISÃO DE LITERATURA	15
2.1 Origem e aspectos gerais do gênero <i>Capsicum annuum</i>	15
2.2 Mercado mundial e nacional de <i>Capsicum</i>	16
2.3 Divergência Genética	17
2.4 Análise multivariada	18
2.5 Componentes Principais	19
2.5.1 Determinação do número de componentes a serem retidos.....	22
2.6 Análise de Fatores	23
2.6.1 Obtenção dos fatores pelas suas cargas fatoriais	25
2.6.2 Escolha do número de fatores	25
2.6.3 Rotação dos Fatores.....	26
3. MATERIAL E MÉTODOS.....	27
3.1 Descrição do experimento.....	27
3.2 Estatística descritiva dos dados	28
3.3 Obtenção dos componentes principais	28
3.4 Obtenção dos fatores e dissimilaridade entre acessos	29
4. RESULTADOS E DISCUSSÕES	30
4.1 Análise descritiva das variáveis em estudo.....	30
4.2 Componentes principais.....	30
4.3 Análise de Fatores	37
5. CONCLUSÕES.....	43
6. REFERÊNCIAS BIBLIOGRÁFICAS.....	44

1. INTRODUÇÃO

Pimentas são depois do sal, o condimento mais utilizado no mundo e encontrado em quase todos os lugares atualmente (Portal São Francisco, 2023). Elas pertencem ao gênero *Capsicum* e a família Solonácea e compreende cerca de 30 espécies já identificadas (RIBEIRO; HENZ, 2007), apresentando diversidade morfológica, em relação à cor dos frutos, forma, textura, tamanho, aroma, e até mesmo o grau de pungência dos frutos, o que torna o mercado de pimentas um segmento diversificado, devido à quantidade de subprodutos que podem ser produzidos (SANTOS, 2018).

As pimentas também se destacam como um importante segmento do mercado de hortaliças frescas no Brasil, com ênfase nas espécies *C. annuum* L. (pimentão), *C. chinense* Jacq. (pimenta-de-cheiro), *C. frutescens* L. (pimenta malagueta) e *C. baccatum* L. (dedo-de-moça), que são muito consumidas e produzidas, de excelente adaptação às condições de clima tropical e com ampla diversidade genética (ZENI; BOSIO, 2011).

Além disso, as pimentas podem ser usadas em diferentes contextos como temperos, condimentos, corantes, conservantes e até usos medicinais apresentando vários benefícios à saúde por serem fonte de antioxidantes naturais como vitamina C, os carotenóides, os quais têm atividade provitamina A, vitamina E, vitaminas do complexo B além de compostos fenólicos (PINTO *et al.*, 2013).

Devido aos diversos usos apresentados pelas pimentas e para atender o mercado consumidor que tem vindo a crescer, existe a necessidade de melhorar qualitativa e quantitativamente a produtividade das culturas de forma associada a resistência à pragas e doenças (BENTO *et al.*, 2007), o que pode ser feito através de programas de melhoramento.

O melhoramento genético das pimentas visa a seleção, dentro de populações já existentes, de características desejáveis para atender necessidades específicas, permitindo melhorias na produção, maior valor comercial, mais sabor, maior resistência a doenças e pragas, qualidade dos frutos para comercialização como: aparência, resistência ao armazenamento prolongado, processamento e qualidade nutricional (CASALI, 1984). O trabalho de melhoramento genético, ao ser feito de forma responsável, pode garantir a competitividade do Brasil no mercado global.

Uma das maiores características do gênero *Capsicum* é a sua alta variabilidade de tipos de frutos, cores, formas, tamanhos e intensidade de sabor picante (Capsaicina) (BLAT, 2004), o que a torna essencial para o trabalho de melhoramento, esta diversidade facilita o desenvolvimento de novas variedades e híbridos para a obtenção de linhagens superiores, logo é fundamental o conhecimento da diversidade entre os acessos, (RÊGO *et al.*, 2013), que permitirá selecionar acessos geneticamente mais divergentes para cruzamentos, possibilitando o aparecimento de cultivares superiores.

A obtenção de genótipos com desempenho superior pode ser feita com o uso de técnicas estatísticas que, permitem uma maior eficácia na avaliação do material genético e na avaliação de divergência, considerando-se também, a contribuição e a importância relativa dos caracteres que contribuem para a variância total existente nas populações (RÊGO *et al.*, 2013).

Assim as técnicas multivariadas são bastante utilizadas pelos programas de melhoramento por estimar a divergência genética e o grau de dissimilaridade entre acessos dentro de uma população melhorada (PEREIRA; CRUZ, 2003). Dentre as técnicas estatísticas multivariadas, destacam-se a Análise de Componentes Principais (ACP) e a Análise de Fatores (AF), ambas concebidas para o tratamento de medidas compostas contendo múltiplas características mensuradas em uma mesma unidade amostral.

A ACP objetiva reduzir a dimensionalidade dos dados, promovendo uma transformação linear das variáveis originais em um novo conjunto de variáveis não correlacionadas denominadas de componentes principais. Esses componentes são ordenados por sua capacidade de capturar a maior parte da variância total dos dados, permitindo assim a visualização e compreensão da estrutura subjacente deles.

Por sua vez, a AF se assemelha à ACP em alguns aspectos, mas difere em seus propósitos. A análise fatorial busca identificar fatores latentes que explicam a covariância entre as variáveis observadas. Esses fatores latentes representam características não diretamente observáveis, mas que influenciam conjuntamente as múltiplas variáveis mensuradas. Através da análise dos fatores, visa compreender as relações subjacentes e estruturais entre as variáveis, permitindo uma abordagem mais profunda na investigação dos fenômenos estudados.

Assim, este estudo, teve como objetivo estudar a variabilidade genética por meio da análise de componente principal e análise fatores.

2. REVISÃO DE LITERATURA

2.1 Origem e aspectos gerais do gênero *Capsicum annum*.

A *Capsicum annum* L. é uma espécie de planta originária das Américas Central e do Sul. Acredita-se que essa espécie tenha sido cultivada por milhares de anos por povos indígenas antes mesmo da chegada dos europeus ao continente. Em sua variedade primitiva tinham frutos pequenos e encarnados, muito atrativos para os pássaros (SANTOS, 2009).

De acordo com Longatti citando Reifschneider (2000), acredita que a partir de explorações arqueológicas em Tehuacán, no México, que as pimentas já eram consumidas há aproximadamente nove mil anos antes de Cristo, por populações pré-colombianas, e foram introduzidas no continente europeu como produto de luxo e, a partir daí na África e na Ásia, após a chegada de Cristóvão Colombo à América no século XV (FERRÃO, 2005). Percebe-se que as rotas de navegação promoveram a dispersão de espécies de pimentas pelo mundo, possibilitando o consumo crescente por todos os povos de todas as origens e sob diversas formas de utilização.

Santos (2009) lembra que as caracterizações das diferentes espécies envolvem um vasto conjunto de descritores, mas ainda assim, é conveniente ter a noção de que existe um número restrito de elementos particulares que se enquadram numa descrição diagnóstica de base, e que permitem, numa primeira abordagem visual, identificar o indivíduo presente.

Sendo assim, a espécie *Capsicum annum*, de forma geral, apresenta uma flor por nó. Na ântese, os pedicelos podem ser eretos, horizontais ou pendentes. A corola é branca (raramente violeta), sem máculas na base dos lobos das pétalas. As anteras são geralmente azuladas. Os cálices dos frutos maduros são geralmente pouco dentados e não possuem constrição anelar na união com o pedicelo. Os frutos são de várias cores e formas, geralmente pendentes, persistentes, com polpa firme. As sementes são cor de palha. O sistema radicular é pivotante, com um número elevado de ramificações laterais, podendo chegar a profundidades de 70-120 cm. (RIBEIRO; HENZ, 2007).

A palavra *Capsicum* tem origem do grego *kapsō*, que significa picar e *kapsakes* que por sua vez significa cápsula (NUEZ *et al.*, 1996). A espécie

Capsicum annuum é considerada uma das mais importantes por abranger quase todas as variedades cultivadas, sendo o pimentão a variedade mais comum dessa espécie, podendo apresentar-se em duas formas doces.

Os pimentões são cultivados em diferentes variedades, incluindo verde, vermelho, amarelo e laranja, e cada variedade tem suas próprias características distintas de sabor e cor. Além disso, os pimentões são ricos em vitaminas e minerais, incluindo vitamina C, vitamina A e potássio, e são considerados uma opção saudável para incluir na dieta.

2.2 Mercado mundial e nacional de *Capsicum*

Capsicum annuum L. é uma das espécies vegetais mais importantes do mundo, tanto em termos de uso alimentar quanto de valor comercial. A produção e o comércio mundial de *Capsicum* tem crescido significativamente nos últimos anos, impulsionados pela crescente demanda por alimentos picantes em muitas partes do mundo.

De acordo com dados da FAO (Organização das Nações Unidas para Agricultura e Alimentação), a produção mundial de pimentas e pimentões em 2020 foi de cerca de 38 milhões de toneladas, com a China sendo o maior produtor, seguido por México e Indonésia. (FAO, 2021)

O mercado mundial de *Capsicum* continuou a crescer em 2022 e de acordo com relatórios recentes da Grand View Research, o mercado global de pimentas e pimentões foi avaliado em cerca de US\$ 5,5 bilhões em 2020 e deve crescer a uma taxa composta anual de cerca de 6,6% entre 2021 e 2028.

No Brasil, a produção de pimentas e pimentões tem crescido, mas não tem uma estimativa exata da produção de pimenta, porque em grande parte a pimenta é cultivada por pequenos agricultores de diversas regiões brasileiras, mas acredita-se que a área cultivada anualmente chega próximo de cinco mil ha com uma produção de 75 mil toneladas (RIBEIRO *et al.*, 2022), tendo Minas Gerais, Goiás, São Paulo, Ceará, Rio Grande do Sul, Bahia e Sergipe como os estados de maior concentração da produção.

O mercado nacional de *Capsicum* é bastante diversificado e inclui tanto o consumo *in natura* quanto a industrialização para a produção de molhos, conservas, temperos e outros produtos alimentícios. Além do mercado interno, parte da

produção brasileira de pimentas é exportada em diferentes formas, como páprica, pasta, desidratada e conservas ornamentais. Esta crescente demanda do mercado tem impulsionado o aumento da área cultivada e o estabelecimento de agroindústrias, tornando o agronegócio de pimentas um dos mais importantes do país (RIBEIRO *et al.*, 2022).

Concluindo, o mercado mundial e nacional de *Capsicum* é bastante promissor, com perspectivas de crescimento contínuo nos próximos anos, impulsionado pela crescente demanda por alimentos picantes e pela popularidade da culinária regional em muitas partes do mundo.

2.3 Divergência Genética

De acordo com Costa (2004), O gênero *Capsicum* apresenta uma grande diversidade genética, tanto dentro das espécies quanto entre elas. A diversidade genética é determinada pela variabilidade nos genes, que podem ser observados através de diferenças na forma, cor e tamanho dos frutos, na arquitetura das plantas e na resistência a doenças e pragas.

A divergência genética no *Capsicum* é atribuída à sua origem em um ambiente naturalmente diverso e à sua domesticação por populações humanas ao longo de milhares de anos. O Brasil é considerado o centro de diversidade de *Capsicum*, visto que a ampla variabilidade existente na espécie ainda não está inteiramente conhecida e explorada (SUDRÉ *et al.*, 2010), destacando-se, ainda, por possuir ampla diversidade em todas as categoriais, sendo contemplado com quatro das cinco espécies domesticadas: *Capsicum annuum* var. *annuum*; *C. baccatum* var. *pendulum*; *C. frutescens* e; *C. chinense* (Longatti *apud* REIFSCHEIDER, 2000).

A existência de variabilidade genética na população é condição básica para que um programa de melhoramento genético obtenha sucesso (IVOGLO *et al.*, 2008), ou seja, pode-se fazer a hibridação de genótipos selecionando características desejadas e eliminando aquelas indesejadas. Nesse contexto é necessário que a variabilidade genética e fenotípica seja adequadamente estudada para que a mesma seja conservada e efetivamente útil no desenvolvimento de cultivares adaptadas às regiões tropicais e sub-tropicais (COSTA *et al.*, 2011).

Sabendo que divergência genética está relacionada ao grau de distanciamento entre populações quanto ao conjunto de caracteres que lhes são

peculiares (FIGUEIREDO, 2004), pode-se determinar quão distante geneticamente uma população ou acesso é de outro, utilizando-se da estatística multivariada que permite unificar múltiplas informações de um conjunto de caracteres (SUDRÉ *et al*, 2005).

2.4 Análise multivariada

A análise univariada trabalha as variáveis de maneira isolada, o que torna praticamente inviável a seleção dos acessos com as melhores combinações de características, pois quando a seleção é feita com base em uma ou poucas características, geralmente resulta-se em materiais ou recomendações com desempenho não tão favorável para características que não foram consideradas e conseqüentemente os materiais ou recomendações podem ser rejeitadas pelos destinatários finais.

Assim, recorre-se às análises multivariadas, as quais trabalham as variáveis de forma conjunta e permitem a combinação de múltiplas informações contidas na unidade experimental, possibilitando a caracterização dos acessos com base em um conjunto de variáveis independentes (FERRÃO *et al.*, 2011).

As técnicas de análise multivariada apresentam-se como instrumentos poderosos na análise de dados em diversas áreas, como na biologia, na agricultura, na medicina, na economia, entre outras, em razão de permitirem combinar as informações múltiplas, tomadas nas unidades de amostragem (GODOI, 1985). Essas técnicas possibilitam prever fenômenos genéticos, baseando-se, simultaneamente, em um complexo de variáveis de importância tanto genética quanto evolutiva.

A análise estatística multivariada pode ser aplicada a qualquer delineamento experimental e tem como objetivo medir, explicar e prever o grau de relacionamento entre variáveis (VIALI, 2005). Ela simplifica o processo, tornando mais eficiente e completo o que seria feito por inúmeras análises univariada ou bivariadas.

Em um programa de melhoramento, o estudo da diversidade genética por meio de técnicas multivariadas é de primordial importância, principalmente no planejamento de programas e na definição de estratégias de trabalhos (IVOGLO *et al.*, 2008), e também porque raras são as vezes em que se trabalha com variáveis

independentes umas das outras e as técnicas multivariadas permite examinar adequadamente as relações de inúmeras variáveis ao mesmo tempo e obter um entendimento completo.

Na literatura, os métodos multivariados estão divididos em diferentes técnicas as quais são utilizadas dependendo das necessidades da pesquisa, precisão desejada ou mesmo das características do conjunto de dados. Neste trabalho o foco é para a Análise de Componentes Principais e Análise de Fatores no estudo da variabilidade genética de pimentas.

2.5 Componentes Principais

Concebida por Karl Pearson em 1901 e estudada por Harold Hotelling em 1933 (LOIOLA, 2020), a Análise de componentes principais (ACP), também conhecida como PCA (do inglês *Principal component analysis*), é uma técnica estatística que tem por finalidade básica transformar um conjunto p de variáveis correlacionadas em um conjunto menor de k variáveis não correlacionadas ($k < p$), a partir de combinações lineares das variáveis originais.

Dado o objetivo da ACP, esta técnica é comumente aplicada para encontrar padrões em dados de alta dimensão e expressá-los de maneira a destacar suas semelhanças e diferenças, o que simplifica a interpretação dos dados. Seu uso tem sido alargado conforme os avanços computacionais e o desenvolvimento de programações que possibilitam cálculos matriciais que diagonalizam matrizes simétricas positivas semi-definidas (KIRCH *et al.*, 2019).

A ideia por trás da PCA é encontrar a direção na qual os dados apresentam a maior variância, ou seja, a direção na qual os dados se espalham mais. Essa direção é definida como o primeiro componente principal. Em seguida, é encontrada a direção que maximiza a variância dos dados, mas que é ortogonal (perpendicular) à primeira direção. Essa direção é o segundo componente principal, e assim por diante, até que sejam encontrados todos os componentes principais.

Uma vez que, cada componente principal é uma combinação linear de todas as variáveis originais, os quais são estimados para reter o máximo de informações em termos da variação total contida nos dados, tem-se os autovalores que representam as variâncias dos componentes principais e os autovetores

representando os dados originais transformados. (REGAZZI; CRUZ, 2020; VARELLA 2008)

Na prática construção dos componentes principais é feita a partir de uma matriz de covariância ou de correlação dos dados originais. A matriz de covariância é uma medida da variabilidade conjunta entre as variáveis, enquanto a matriz de correlação é uma medida da relação linear entre as variáveis. No entanto, se houver uma discrepância acentuada entre as variáveis, os componentes principais obtidos via matriz de covariâncias têm pouca utilidade, devido a influência que as variáveis de maior variância causam nos componentes principais e neste caso deve-se fazer a padronização de cada variável (MINGOTI, 2020).

A teoria a seguir está de acordo com o exposto em Regazzi e Cruz (2020). Considere uma amostra aleatória de n elementos, em que, para cada elemento da amostra, denotados por X_1, X_2, \dots, X_p , foram observadas p variáveis de interesse. Este conjunto de $n \times p$ medidas, dá origem a uma matriz X , de ordem $n \times p$, dada por:

$$X = \begin{bmatrix} X_{11} & \cdots & X_{1p} \\ \vdots & \ddots & \vdots \\ X_{n1} & \cdots & X_{np} \end{bmatrix}$$

A correlação linear amostral entre as i -ésimas e j -ésimas variáveis pode ser estimada através da equação:

$$r_{ij} = \frac{\text{Côv}(X_i, X_j)}{\sqrt{\text{Var}(X_i) \cdot \text{Var}(X_j)}}$$

em que $\text{Côv}(X_i, X_j)$ é a covariância estimada entre X_i e X_j e $\text{Var}(X_i)$ e $\text{Var}(X_j)$ é a variância estimada para X_i e X_j , respectivamente. Assim a matriz de correlação é definida por:

$$R = \begin{bmatrix} 1 & \cdots & r_{1p} \\ \vdots & \ddots & \vdots \\ r_{p1} & \cdots & 1 \end{bmatrix}$$

Percebe-se que R é uma matriz simétrica de dimensão $p \times p$. Enfatiza-se que utilizar a matriz de correlação dos dados, equivale a utilização da matriz de dados padronizados.

Deseja-se, a partir de p variáveis padronizadas R_1, R_2, \dots, R_p encontrar outro conjunto de variáveis Y_1, Y_2, \dots, Y_p , tal que as seguintes condições sejam atendidas:

- a) $Y_i = \sum_{j=1}^p a_{ij}R_j$, com $i = 1, 2, \dots, p$ e $\sum_{j=1}^p a_{ij}^2 = 1$
- b) $\text{Var}(Y_1) \geq \text{Var}(Y_2) \geq \dots \geq \text{Var}(Y_p)$
- c) $\text{Cov}(Y_i, Y_j) = 0$ para $i \neq j$
- d) $\sum_{i=1}^p \text{Var}(Y_i) = \sum_{i=1}^p \text{Var}(Z_i)$

Objetivando-se obter o vetor \bar{a} de forma que a variância de Y_1 seja maximizada, os autores sugerem o uso dos multiplicadores de Lagrange:

$$(R - \lambda_1 I) \bar{a} = 0$$

em que R é a matriz de correlação, λ_1 uma constante que representa a variância da j -ésima variável original, e I matriz identidade. Uma vez que se trata de um sistema homogêneo, podendo admitir sempre a solução trivial e todos os coeficientes seriam iguais a zero, leva à necessidade de que a solução do sistema deve ser tal que $\bar{a} \neq 0$, o que leva à condição que o determinante da matriz $[R - \lambda_1 I]$ seja nulo:

$$\det(R - \lambda_1 I) = 0.$$

Ao conhecer a equação característica acima e sendo λ_1 os valores que a satisfazem, torna-se possível determinar os autovalores λ_1 , que correspondem as raízes características da matriz R e \bar{a} o vetor característico associado, ou seja, o autovetor.

O maior autovalor estará associado ao autovetor que define o primeiro componente. O autovalor obtido é a própria variância do respectivo componente principal (FERREIRA, 2018). Portanto, como o vetor \bar{a} foi escolhido para maximizar $V(Y_1)$, tem-se que λ_1 assume, nesta condição, o valor máximo entre os elementos do conjunto de autovalores de R . (REGAZZI; CRUZ, 2020).

Já o segundo componente principal é dado pela combinação linear das variáveis originais que tem a segunda maior variância a qual é dada por:

$$V(Y_{i2}) = V(Y_2) = b' R b$$

e considerando as restrições $b'b = 1$ e $b'a = a'b = 0$, em que a primeira garante a unicidade de b e a segunda garante que Y_1 e Y_2 sejam ortogonais, da equação característica:

$$(R - \lambda_2 I) b = 0 \quad \text{e} \quad (R - \lambda_2 I) = 0$$

conclui-se que λ_2 é o segundo maior autovalor de R e b o seu autovetor associado. Fazendo o processo análogo ao que foi feito para os dois primeiros componentes principais, pode-se determinar todos os demais componentes principais.

2.5.1 Determinação do número de componentes a serem retidos

A escolha do número de componentes a serem retidos é um passo fundamental na ACP. É desejável ter $k < p$ componentes pois um dos objetivos da ACP é a redução da dimensão dos dados a serem trabalhados. No entanto, de acordo com Loiola (2020), o procedimento de decisão compreende critérios para o número exato de componentes a considerar se extrair e na maioria das vezes, a partir do mesmo número de variáveis é gerado o mesmo número de componentes, isso ocorre devido as muitas combinações lineares das variáveis originais. Nesses casos deve-se fazer a escolha de um menor número de componentes com perda mínima de informação utilizando um ou uma combinação dos métodos sugeridos na literatura.

Autores como Regazzi e Cruz (2020), sugerem o uso da importância relativa, que representa a proporção total explicada (λ_i) pela componente Y_i .

$$IR(\%) = \frac{Var(y_i)}{\sum_{i=1}^p Var(y_i)} \times 100$$

Tal que:

$$\frac{Var(y_i) + \dots + Var(y_k)}{\sum_{i=1}^p Var(y_i)} \times 100 \geq 80\%$$

Ou seja, tendo os componentes principais ordenados de forma decrescente de acordo com a variância explicada, deve-se reter os primeiros componentes que acumulem uma porcentagem de variância explicada igual ou superior a 80%.

Outro método encontrado na literatura é o critério de *scree-plot* que verifica a curva de sedimentação dos autovalores. Este critério envolve a plotagem dos autovalores em ordem decrescente no eixo vertical e o número de componentes no eixo horizontal. De acordo com Matos e Rodrigues (2019), a curva resultante normalmente mostra uma queda acentuada nos autovalores iniciais, seguida por uma queda mais gradual. O critério de *scree-plot* sugere que o número de componentes a serem retidos é determinado pelo ponto de inflexão, ou seja, o ponto em que a queda nos autovalores se torna menos acentuada, indicando que os componentes subsequentes explicam uma proporção menor de variância.

Outro método muito utilizado pelos pesquisadores é o critério de Kaiser (KAISER, 1958), que sugere reter, quando se trabalha com a matriz de correlação,

ou seja, com dados padronizados, somente os componentes principais associados a autovalores superiores a 1 ($\lambda_i > 1$) (REGAZZI; CRUZ, 2020)

De acordo com Regazzi e Cruz (2020) os escores ou valores numéricos de cada indivíduo dos componentes principais podem ser calculados por:

$$Y_{n1} = a_{11}X_{n1} + a_{12}X_{n2} + \dots + a_{1p}X_{np}$$

Em que a corresponde ao respectivo autovetor associado à variável e X é a própria variável padronizada. Os escores devem ser calculados, sempre que se desejar fazer uso do mesmo, como no caso da comparação entre tratamentos, avaliação da dissimilaridade ou agrupamento.

2.6 Análise de Fatores

Análise de fatores é uma técnica estatística multivariada utilizada para identificar padrões ou estruturas subjacentes em um conjunto de dados. Ela busca reduzir a dimensionalidade dos dados, identificando variáveis que estão altamente correlacionadas entre si e agrupando-as em fatores latentes, que explicam grande parte da variância observada nos dados.

A Análise de Fatores começou a ser desenvolvida pelo psicólogo Charles Spearman em 1904, para testar um fator de inteligência e habilidade nos jovens, quando realizados testes a diferentes áreas do conhecimento e mais tarde foi estudada por Thurstone, por acreditar que existe mais de um fator de inteligência (VICINI, 2005).

A análise de fatores pode ser vista como uma extensão da análise das componentes principais, uma vez que ambas objetivam reduzir o banco de dados e podem ser encaradas como aproximações à matriz das covariâncias. Contudo, a aproximação feita pelo modelo da análise de fatores é mais elaborada e centra-se na análise da consistência dos dados com uma estrutura pré-definida (FERREIRA, 2000 p. 86).

Enquanto a ACP gera componentes calculados como combinações lineares das variáveis originais, e busca explicar o máximo possível de variância total nas variáveis, não diferenciando variância comum da variância específica entre os itens, incluindo ambos os tipos de variância; Na análise de fatores, as variáveis originais são definidas como combinações lineares dos fatores, criando os fatores comuns ou

variáveis latentes, existindo maior interesse em explicar as covariâncias ou correlações entre as variáveis.

A análise de fatores visa encontrar fatores latentes subjacentes que explicam a covariância entre as variáveis observadas, de modo que n variáveis observadas são sintetizadas em um conjunto de m variáveis latentes, sendo $m \leq n$ (MINGOTTI, 2020). Portanto, o modelo de análise de fatores, assume-se que cada variável observada é uma combinação linear dos fatores latentes, com um termo de erro e pode ser expresso da seguinte forma (REGAZZI; CRUZ, 2020):

$$X_p - \mu_p = l_{p1}F_1 + l_{p2}F_2 + \dots + l_{pm}f_m + \mathcal{E}_p$$

Utilizando o modelo matricial:

$$X - \mu = LF + \mathcal{E}$$

$$\begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix} - \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_p \end{pmatrix} = \begin{pmatrix} l_{11} & \dots & l_{1m} \\ \vdots & \ddots & \vdots \\ l_{p1} & \dots & l_{pm} \end{pmatrix} \begin{pmatrix} F_1 \\ \vdots \\ F_m \end{pmatrix} + \begin{pmatrix} \mathcal{E}_1 \\ \vdots \\ \mathcal{E}_p \end{pmatrix}$$

Em que: X_p representa as variáveis aleatórias observadas com média μ_i , com $i = 1, 2, \dots, p$; os elementos l_{pm} são as cargas fatoriais da p -ésima variável no m -ésimo fator; F_m corresponde aos fatores latentes não observáveis; \mathcal{E}_p é o erro ou fatores específicos associados à p -ésima variável observada X_p , que captura a variabilidade única não explicada pelos fatores latentes.

Para a análise de fatores as seguintes suposições devem ser satisfeitas:

$$E(F) = 0$$

$$\text{Var}(F) = I$$

$$E(\mathcal{E}) = 0$$

$$\text{Var}(\mathcal{E}) = \psi = \text{diag}(\psi_1, \dots, \psi_p)$$

De forma que se tem um modelo fatorial ortogonal, em que os m fatores são ortogonais entre si e F e \mathcal{E} são não correlacionados

Tem-se ainda que $\text{Var}(X_i) = l_{i1}^2 + l_{i2}^2 + \dots + l_{im}^2 + \psi_i = h_i^2 + \psi_i$, em que ψ_i é a variância devido ao fator específico e h_i^2 é a comunalidade, que se refere a estimativa de variância compartilhada. A comunalidade é obtida a partir da soma em linha de cargas fatoriais quadradas. Quanto mais altos os valores, mais variância em

uma variável é explicada pelos fatores, portanto para a análise devem ser considerados um nível mínimo de 0.50 para comunalidades (Hair *et al.*, 2009).

2.6.1 Obtenção dos fatores pelas suas cargas fatoriais

As cargas fatoriais são coeficientes numéricos que representam as relações entre os fatores latentes e as variáveis observadas. Elas indicam a magnitude e a direção da influência de cada fator sobre cada variável. Essas cargas fatoriais podem ser estimadas a partir da decomposição Espectral para fatoração da matriz de correlação, no método de componente principal. Em que os autovetores representam os componentes principais e os autovalores medem a quantidade de variação explicada por cada componente.

Portanto, as cargas fatoriais l_{pm} para a p-ésima variável observada e o m-ésimo componente principal podem ser calculados por:

$$l_{pm} = e_{mp} \times \sqrt{\lambda_j}$$

Em que:

e_{mp} é o elemento da p-ésima posição do autovetor j e,

λ_j é o j-ésimo autovalor.

A interpretação das cargas fatoriais é uma etapa crucial na análise fatorial, pois permite compreender quais fatores latentes estão mais fortemente relacionados com quais variáveis. Quanto maior o valor absoluto da carga fatorial, maior é a influência do fator latente sobre a variável observada. Cargas fatoriais mais próximas de zero indicam que o fator tem pouca influência sobre a variável.

2.6.2 Escolha do número de fatores

A escolha do número de fatores é um passo fundamental na análise fatorial. É desejável ter $k < p$ fatores pois, caso contrário, embora se encontrem p variáveis não correlacionadas não se diminui o número de variáveis iniciais que é um dos objetivos da análise fatorial.

À semelhança da análise em componentes principais, para a decisão do número de fatores a reter, pode-se usar o teste de *scree-plot* dos valores próprios de S ou R versus o número de fatores. O método da importância relativa ou o critério de

Kaiser. As aplicações dos diferentes métodos podem gerar resultados diferentes, sobre o mesmo conjunto de dados, por isso cabe ao pesquisador ser parcimonioso e reter aqueles fatores comuns, advindos da aplicação dos diferentes métodos.

2.6.3 Rotação dos Fatores

Algumas vezes os resultados da análise de fatores podem ser difíceis de interpretar, devido a complexidade das cargas fatoriais. Nestes casos é recomendado fazer a rotação dos fatores de forma a simplificar a estrutura, melhor interpretação e ajuste do modelo fatorial.

Na literatura são apresentadas as rotações oblíquas e ortogonal como dois métodos gerais de rotação. As rotações oblíquas, que permitem que os fatores latentes sejam correlacionados após a rotação a exemplo do método Promax que tenta simplificar as correlações entre fatores, transformando-as em fatores não correlacionados e o Direct Oblimin, o qual não tenta ajustar as correlações entre fatores, mas simplesmente as mantém como estão.

Por outro lado, as rotações ortogonais permitem que após a rotação, os fatores mantenham independentes e não correlacionados. Exemplos de procedimentos ortogonais é a rotação Varimax, que procura maximizar a variância dos quadrados das cargas fatoriais em cada fator, resultando em fatores mutuamente exclusivos e mais fácil interpretação e a rotação Quartimax que maximiza a variância dos quadrados das cargas de cada variável em todos os fatores.

3. MATERIAL E MÉTODOS

3.1 Descrição do experimento

Para a realização deste estudo utilizou-se dados provenientes de um experimento conduzido em casa de vegetação, pertencente ao setor de olericultura do Departamento de Fitotecnia da Universidade Federal de Viçosa (UFV), no município de Viçosa, situado nas coordenadas geográficas 20° 45' de latitude sul e 42° 51' de longitude oeste, com altitude média de 650 m.

Utilizando o delineamento experimental inteiramente casualizado, com quatro repetições, sendo a unidade experimental constituída por uma planta, foram avaliados nove acessos de *Capsicum annuum* L., são eles: Pimenta Vulcão (1), Pimenta Cayene (2), Pimenta Peter (3), Pimenta Picante para vaso (4), Pimenta Jamaica Yellow (5), Pimenta Doce Italiana (6), Pimentão Quadrado (7), Pimentão Cascadura Ikeda (8) e Pimentão Rubi Gigante (9).

Foram avaliados os seguintes caracteres:

Comprimento do fruto maduro (COM) – a mensuração foi realizada a partir da inserção do pedúnculo até a ponta do fruto, utilizando-se paquímetro digital, expresso em mm;

Largura do fruto maduro (LAR) – foi mensurada na maior largura da seção transversal do fruto, utilizando-se paquímetro digital, expresso em mm;

Peso total do fruto (PT) – o fruto foi pesado em balança analítica, expresso em g;

Espessura do pericarpo (ESP) – o fruto foi cortado na transversal, em seguida foi realizada a medição na porção intermediária do fruto, utilizando-se paquímetro digital, expresso em mm;

Número de sementes por fruto (NS) – obtido pela contagem do número total de sementes por fruto;

Massa da matéria total do fruto maduro fresco (MF) – após a remoção das sementes o fruto foi pesado em balança analítica, expresso em g;

Massa da matéria total do fruto maduro seco (MS) – o material do item anterior foi conduzido à estufa de circulação forçada a 60 °C até atingir massa constante, aproximadamente 72 h, em seguida foi pesado em balança analítica fornecendo a massa da matéria total do fruto maduro seco (MTS).

Teor de sólidos solúveis (BRIX) – foi determinado extraíndo-se um suco do fruto, em seguida o refratômetro foi zerado com água destilada e após depositou-se uma a duas gotas do suco no prisma do aparelho e foi aferido o °Brix (leitura direta) a 26 °C (± 2). A escala de °Brix corresponde a um grama de açúcar (sacarose) em 100 gramas de solução (água);

Teor de Vitamina C (VIT) – a determinação do conteúdo de vitamina C ou ácido ascórbico foi realizada pelo método titulométrico, expresso em $\text{mg}100_g^{-1}$.

3.2 Estatística descritiva dos dados

A matriz de dados obtida com o experimento foi preparada para as análises, o que consistiu no cálculo das médias gerais de cada acesso, para cada uma das características avaliadas, de forma a fornecer uma representação única e eliminar o ruído aleatório presente em cada repetição individual.

Também foi feita a padronização dos dados, dividindo cada média pelo desvio padrão da variável, a fim de remover o viés e a variância dos dados devido a utilização de unidades de medidas com magnitude diferentes para as características, evitando que variáveis com grande variância influenciasse demasiado na avaliação.

A fim de ter melhor entendimento dos dados foram calculadas medidas de estatística descritiva como a média aritmética, mediana, desvio padrão, valor máximo e valor mínimo.

3.3 Obtenção dos componentes principais

A partir dos dados padronizados, foi obtida a matriz de correlações utilizando a função *cor()* do pacote base do R e utilizando a função *eigen()*, também do pacote base do R, foram obtidos os autovalores e autovetores. Em seguida foram calculados, a partir dos autovalores obtidos, os *scores* em cada componente.

Foi avaliada a contribuição de cada componente principal na explicação da variação total, o que juntamente ao gráfico de *Scree* permitiu definir o número de componentes principais a serem retidos. Para esse estudo foram considerados os *k* primeiros componentes referentes a 80% da variância total acumulada. Já a correlação entre os dois primeiros componentes e as variáveis analisadas foram mostradas a partir do gráfico de *biplot*.

3.4 Obtenção dos fatores e dissimilaridade entre acessos

Os fatores foram obtidos a partir da matriz de correlações utilizando a função *factanal()* do *software* R 4.2.2. e definindo componentes principais como método de extração de fatores. A seguir fez-se a rotação dos fatores pelo método varimax.

Foram obtidas as cargas fatoriais em relação aos componentes principais, a partir das quais foi elaborado um gráfico com a dispersão dos acessos em relação aos dois primeiros componentes principais, e de forma visual foi possível definir os grupos de dissimilaridade e fazer conclusões dos melhores cruzamentos entre acessos.

4. RESULTADOS E DISCUSSÕES

4.1 Análise descritiva das variáveis em estudo

A Análise de componentes principais e a Análise de fatores são técnicas exploratórias de dados e, portanto, a análise descritiva das variáveis (Tabela 1) é essencial para melhor compreensão dos resultados, além disso, os resultados obtidos facilitam os cálculos subsequentes.

Tabela 1 – Análise descritiva para as variáveis estudadas

Variável	Média aritmética	Mediana	Desvio-padrão	Valor máximo	Valor mínimo
PT	36,86	9,17	41,06	109,92	0,67
COM	63,65	69,32	34,92	121,57	17,87
LAR	30,88	30,71	22,23	69,96	6,93
ESP	2,31	1,32	1,62	4,97	0,73
NS	81,40	67,38	46,03	151,38	32,13
MF	35,61	8,62	40,12	107,17	0,47
MS	3,22	1,28	3,39	8,87	0,08
BRIX	8,51	8,43	1,24	10,95	7,12
VIT	103,08	98,96	24,48	148,01	76,27

PT: Peso total do fruto, COM: Comprimento do fruto maduro (COM), LAR: Largura do fruto maduro, ESP: Espessura do pericarpo, NS: Número de sementes por fruto, MF: Massa da matéria total do fruto maduro fresco, MS: Massa da matéria total do fruto maduro seco, BRIX: Teor de sólidos solúveis, VIT: Teor de Vitamina C.

4.2 Componentes principais

Conforme mencionado previamente, nesta fase, inicia-se uma análise de componentes principais dos nove tratamentos de pimentas, cada um com nove características distintas, com o objetivo de fazer a redução da dimensionalidade, visando a simplificação do conjunto de dados.

A realização do experimento permitiu obter os dados apresentados na Tabela 2, e a padronização dos mesmos estão apresentados na Tabela 3.

Tabela 2 – Dados originais provenientes do experimento

TRAT	PT	COM	LAR	ESP	NS	MF	MS	BRIX	VIT
1	1,52	27,56	8,89	0,99	42,75	1,15	0,2	9,4	114,68
2	3,49	76,82	8,03	0,91	67,38	2,8	0,44	8,44	109,85
3	8,62	53,11	20,02	1,32	32,13	8,2	1,28	9,51	148,01
4	0,67	17,87	6,93	0,73	35	0,47	0,08	10,95	128,5
5	9,17	29,53	30,71	1,27	52,25	8,62	1,11	7,46	98,96
6	52,97	121,57	35,17	2,78	151,38	50,61	4,33	7,59	91,28
7	109,92	69,32	69,96	4,97	110	107,17	8,87	7,12	76,27
8	66,05	100,4	44,32	3,49	98,47	64,08	4,91	7,65	82,75
9	79,29	76,69	53,86	4,3	143,25	77,38	7,79	8,43	77,38

PT: Peso total do fruto, COM: Comprimento do fruto maduro (COM), LAR: Largura do fruto maduro, ESP: Espessura do pericarpo, NS: Número de sementes por fruto, MF: Massa da matéria total do fruto maduro fresco, MS: Massa da matéria total do fruto maduro seco, BRIX: Teor de sólidos solúveis, VIT: Teor de Vitamina C.

Tabela 3 – Dados padronizados

TRAT	PT	COM	LAR	ESP	NS	MF	MS	BRIX	VIT
1	-0,86	-1,03	-0,99	-0,81	-0,84	-0,86	-0,89	0,72	0,47
2	-0,81	0,38	-1,03	-0,86	-0,30	-0,82	-0,82	-0,05	0,28
3	-0,69	-0,30	-0,49	-0,61	-1,07	-0,68	-0,57	0,81	1,84
4	-0,88	-1,31	-1,08	-0,98	-1,01	-0,88	-0,93	1,97	1,04
5	-0,67	-0,98	-0,01	-0,64	-0,63	-0,67	-0,62	-0,84	-0,17
6	0,39	1,66	0,19	0,29	1,52	0,37	0,33	-0,74	-0,48
7	1,78	0,16	1,76	1,65	0,62	1,78	1,67	-1,11	-1,10
8	0,71	1,05	0,60	0,73	0,37	0,71	0,50	-0,69	-0,83
9	1,03	0,37	1,03	1,23	1,34	1,04	1,35	-0,06	-1,05

PT: Peso total do fruto, COM: Comprimento do fruto maduro (COM), LAR: Largura do fruto maduro, ESP: Espessura do pericarpo, NS: Número de sementes por fruto, MF: Massa da matéria total do fruto maduro fresco, MS: Massa da matéria total do fruto maduro seco, BRIX: Teor de sólidos solúveis, VIT: Teor de Vitamina C.

Dos resultados da análise de componentes principais foram obtidos os autovalores, a proporção da variância explicada por cada componente principal e a proporção da variância total acumulada explicada até cada componente principal, como expresso na Tabela 4.

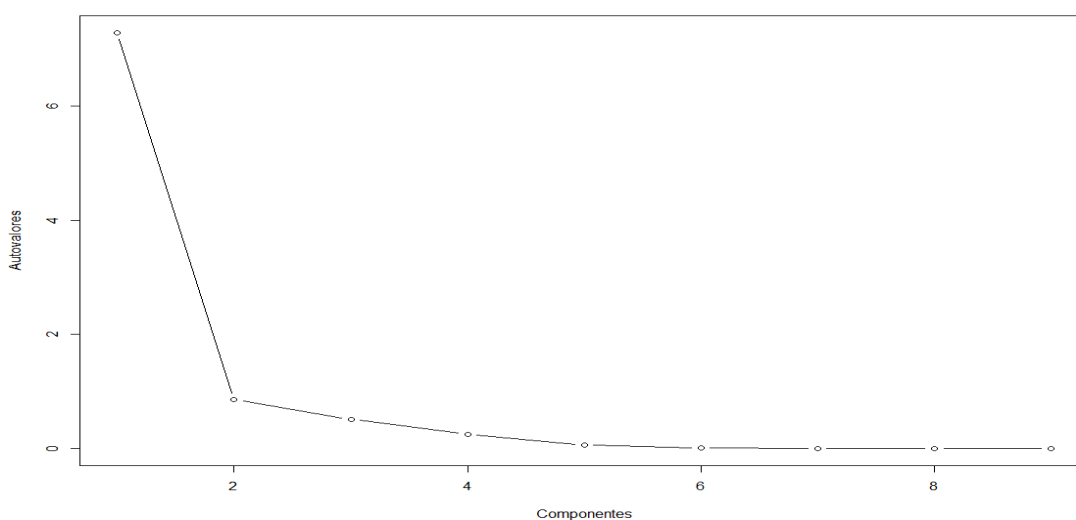
Tabela 4 - Informações importantes da análise de componentes principais

Componente principal	Autovalor λ_i	% da variância Y_i	% Acumulado da variância de Y_i
1	7,290	81,00	81,00
2	0,853	9,48	90,48
3	0,515	5,73	96,21
4	0,253	2,82	99,03
5	0,066	0,74	99,77
6	0,011	0,13	99,90
7	0,007	0,08	99,98
8	0,001	0,01	99,99
9	< 0,001	< 0,01	100

Observando as proporções da variância explicada (Tabela 4), foi possível observar que o primeiro componente principal explica a maior parte da variância (81%), seguido pelo segundo componente principal (9,48%). Os componentes principais subsequentes explicam proporções decrescentes da variância total.

Zoz (2015), em estudo de produtividade de cártamo com base em seis características, verificou que apenas três componentes principais eram suficientes para explicar 85.66% da variância total.

Figura 1 – Gráfico *scree-plot* para determinação do número de componentes principais a serem retidos com base nos autovalores obtidos.



Com base na discussão feita na revisão de literatura, em que os autores Regazzi e Cruz (2020) sugere reter os componentes que explicam 80% ou mais da variância total acumulada, e os autores Matos *et al.* (2019) sugere reter aqueles componentes a partir dos quais o *scree-plot* apresente uma queda acentuada, foram retidos os dois primeiros componentes principais, a redução de dimensionalidade é de nove para dois.

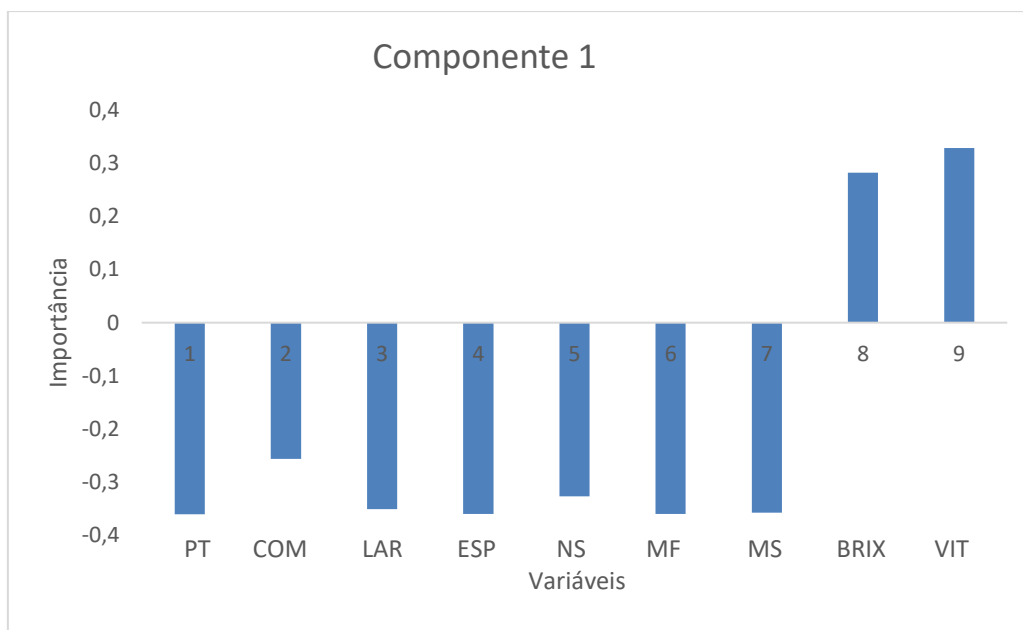
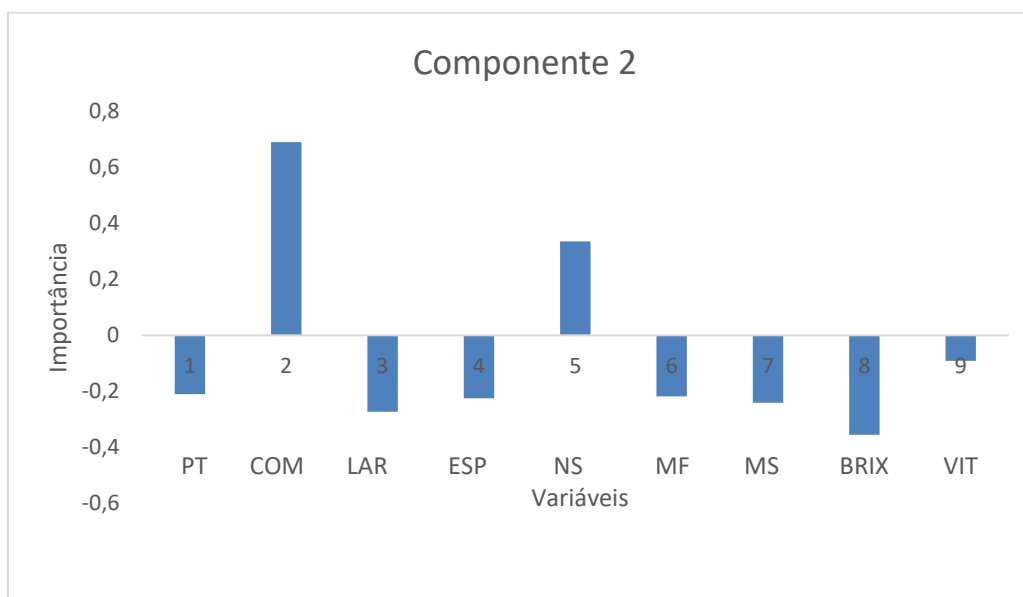
Os dois componentes principais, mantêm a maior parte da variância explicada, acumulando mais de 90% da variância total contida nos dados e, portanto, conseguem explicar a base de dados com menos de 10% de perda de informação.

A fim de conhecer o significado de cada componente retido, foi verificado o grau de influência das variáveis na formação dos dois primeiros componentes principais, foi calculado os coeficientes associados às variáveis e o coeficiente de correlação que quantifica a associação entre cada variável e os respectivos componentes, apresentados na Tabela 5. Nas figuras 2 e 3, está apresentado de forma gráfica, a importância das variáveis nos componentes CP1 e CP2, respectivamente.

Tabela 5 - Coeficiente associados às características e seus coeficientes de correlação com os dois primeiros componentes principais.

	Coeficientes associados às variáveis		Correlação entre as variáveis e componentes	
	CP1	CP2	CP1	CP2
PT	-0,3603	-0,2101	-0,9729	-0,6923
COM	-0,2564	0,6907	-0,1942	0,6382
LAR	-0,3507	-0,2735	0,0835	0,2786
ESP	-0,3598	-0,2251	0,0489	0,1662
NS	-0,327	0,3355	0,0598	0,0858
MF	-0,3599	-0,2177	0,0462	-0,0283
MS	-0,3576	-0,2413	-0,0247	-0,0026
BRIX	0,2822	-0,356	-0,0002	-0,0049
VIT	0,3287	-0,0912	-0,0816	-0,0371

PT: Peso total do fruto, COM: Comprimento do fruto maduro (COM), LAR: Largura do fruto maduro, ESP: Espessura do pericarpo, NS: Número de sementes por fruto, MF: Massa da matéria total do fruto maduro fresco, MS: Massa da matéria total do fruto maduro seco, BRIX: Teor de sólidos solúveis, VIT: Teor de Vitamina C.

Figura 2 - Importância das variáveis para o primeiro componente principal**Figura 3** - Importância das variáveis para o segundo componente principal

Uma vez conhecido a importância das variáveis para cada componente, e considerando a equação geral, as equações para os dois componentes principais foram definidas por:

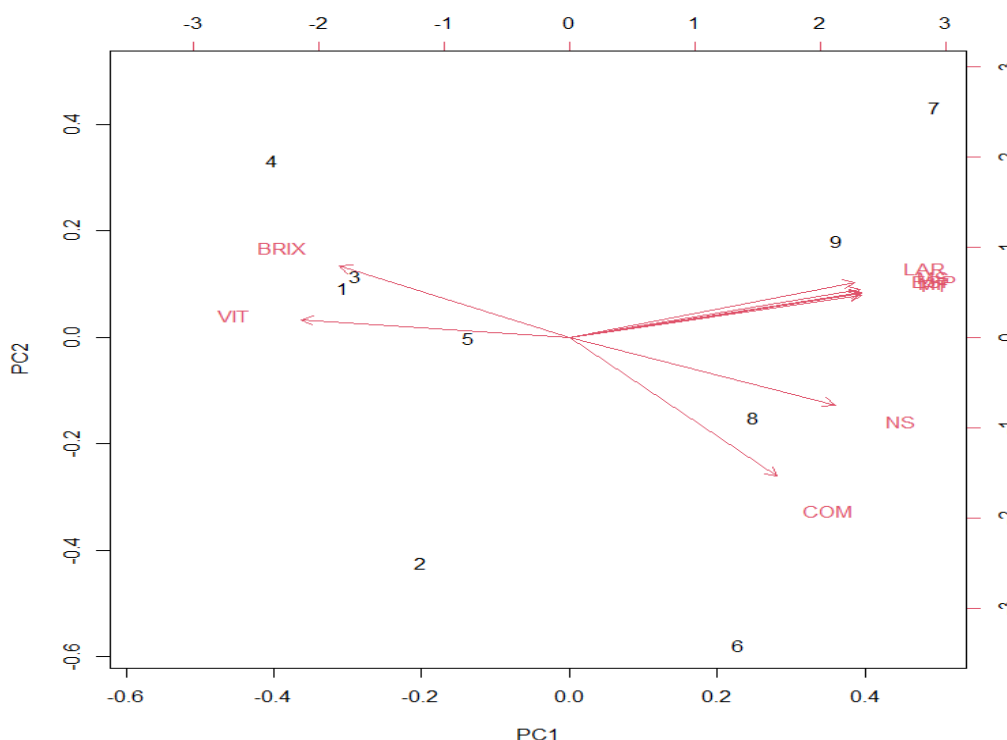
$$CP1 = - 0,3603PT - 0,2564COM - 0,3507LAR - 0,3598ESP - 0,327NS - 0,3599MF - 0,3576MS + 0,2822BRIX + 0,3287VIT \quad (1)$$

$$CP2 = - 0,2101PT + 0,6907COM - 0,2735LAR - 0,2251ESP + 0,3355NS - 0,2177MF - 0,2413MS - 0,356BRIX - 0,0912VIT \quad (2)$$

Das análises gráficas e equações dos componentes, é possível perceber que a variável PT foi a que mais influenciou, em valor absoluto, na composição do primeiro componente, seguido da variável MF, enquanto as variáveis COM e BRIX foram aquelas que tiveram menor influência neste componente. Já CP2 tem maior influência da variável COM e BRIX, enquanto VIT e PT são as variáveis menos relevantes para este componente, representando apenas - 0,0912 e -0,2101 respectivamente.

Analisando a figura 4, que mostra a contribuição das variáveis para os componentes CP1 e CP2, com os números que referem aos tratamentos, verifica-se que há correlação alta entre as variáveis PT, LAR, ESP, MF e MS pois formam ângulos agudos entre elas.

Figura 4 – Gráfico de correlação entre os dois primeiros componentes e as variáveis analisadas.



PT: Peso total do fruto, COM: Comprimento do fruto maduro (COM), LAR: Largura do fruto maduro, ESP: Espessura do pericarpo, NS: Número de sementes por fruto, MF: Massa da matéria total do fruto maduro fresco, MS: Massa da matéria total do fruto maduro seco, BRIX: Teor de sólidos solúveis, VIT: Teor de Vitamina C.

Ainda do gráfico acima, constata-se que o tratamento 2 e o tratamento 7 estão afastados de forma oposta. Enquanto o tratamento 2 tem reduzidas concentrações

das características avaliadas, o tratamento 7 tem altas concentrações da maioria das características.

4.3 Análise de Fatores

Com base na revisão de literatura, foi conduzida uma análise de fatores para exploração das correlações presente entre as variáveis dos nove tratamentos de pimentas advindos do experimento. Espera-se identificar um conjunto reduzido de fatores que possam explicar essas interdependências.

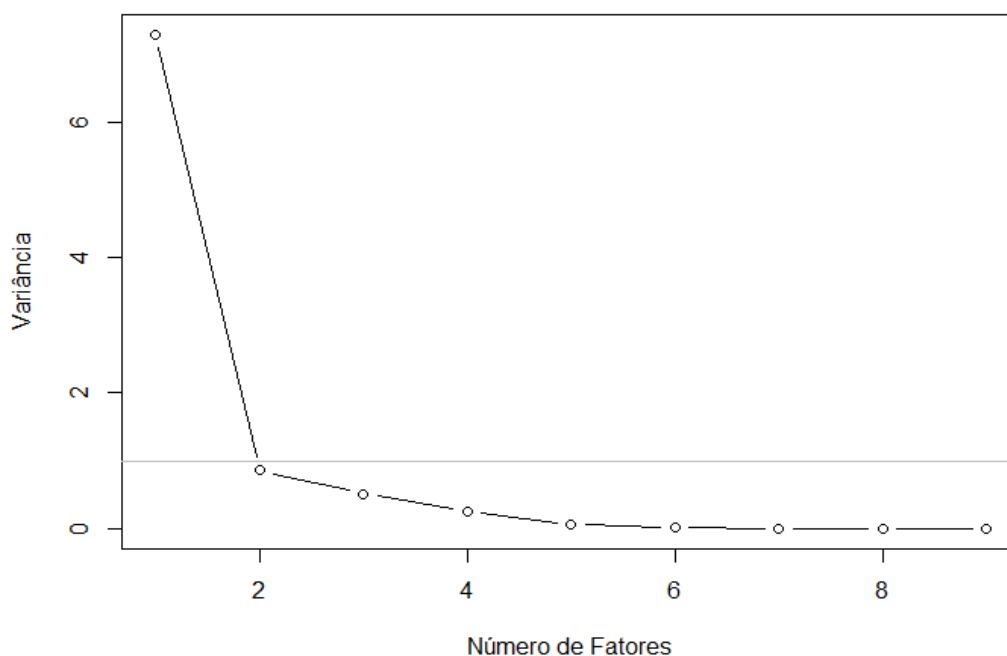
Na análise de fatores, os autovalores indicam a quantidade de variância explicada por cada fator. Quanto maior o autovalor, mais relevante é o fator correspondente na explicação da variância dos dados originais.

Tabela 6 – Informações importantes da análise de fatores

Fator	Autovalor	Variância Explicada	Variância Acumulada
1	7,2872	0,8097	0,8097
2	0,8573	0,0953	0,9049
3	0,5147	0,0572	0,9621
4	0,2541	0,0282	0,9904
5	0,0665	0,0074	0,9978
6	0,0111	0,0012	0,9990
7	0,0077	0,0009	0,9998
8	0,0014	0,0001	0,9999
9	< 0,0001	< 0,0001	1,0000

A Tabela 6 apresenta os autovalores, a variância explicada e a variância explicada acumulada, informações que fornecem bases iniciais sobre a quantidade de fatores a serem retidos na análise. De acordo com o critério de Kaiser, todos os autovalores maiores que um devem ser mantidos na Análise de fatores. Seguindo esse critério, seria retido um fator. Por outro lado, ao adotar o método da variância explicada, se buscarmos capturar 80% ou mais da variância total, então reteríamos um ou dois.

Figura 5 - Gráfico *scree-plot* para determinação do número de fatores a serem retidos, com linha de referência em 1 (Critério de Kaiser)



Na figura 5 tem-se o *scree-plot* que também é um critério que ajuda na definição do número de fatores a reter. É possível perceber uma quebra acentuada a partir do segundo fator e as demais variâncias tendem a uma constante. Portanto, com base nos critérios apresentados devem ser retidos os dois primeiros fatores, juntos eles explicam aproximadamente 90,4% da variabilidade presente nos dados originais.

Esse resultado corrobora os estudos com pimentas *Capsicum annuum* para avaliação do potencial ornamental de Guimarães *et al.* (2020), onde foi possível resumir 12 características estudadas, em apenas três fatores, tornando mais fácil a seleção de acessos de maior potencial.

Depois de retidos os fatores a serem trabalhados, fez-se a rotação dos vetores pelo método ortogonal varimax e os resultados estão apresentados na Tabela 7.

Tabela 7 – Cargas fatoriais antes e depois da rotação ortogonal

Variável	Sem rotação		Com rotação	
	Fator 1	Fator 2	Fator 1	Fator 2
PT	0.98	0.17	0.90	0.42
COM	0.67	-0.56	0.23	0.84
LAR	0.95	0.24	0.92	0.34
ESP	0.98	0.18	0.91	0.41
NS	0.87	-0.36	0.51	0.80
MF	0.98	0.17	0.90	0.41
MS	0.97	0.19	0.91	0.40
BRIX	-0.72	0.21	-0.47	-0.58
VIT	-0.86	0.09	-0.66	-0.57

PT: Peso total do fruto, COM: Comprimento do fruto maduro (COM), LAR: Largura do fruto maduro, ESP: Espessura do pericarpo, NS: Número de sementes por fruto, MF: Massa da matéria total do fruto maduro fresco, MS: Massa da matéria total do fruto maduro seco, BRIX: Teor de sólidos solúveis, VIT: Teor de Vitamina C.

Os valores de cargas fatoriais mais próximos de um indicam uma associação mais forte com o fator correspondente, então com base nos resultados da análise de fatores, pode ser observado que o Fator 1 apresenta altas cargas para as variáveis PT, ESP, MF, MS e LAR e pode ser denominado como fator de qualidade do fruto. Enquanto o Fator 2, apresentou carregamentos mais altos nas variáveis COM e NS e neste caso pode ser chamado de componente de características físicas e reprodutivas. A proporção de variância para os dois fatores é aproximadamente 88% (Tabela 8).

Na abordagem com rotação Varimax, o Fator 1 apresentou carregamentos mais altos nas variáveis LAR, ESP, MS e MF e PT enquanto o Fator 2 continuou com carregamentos mais altos nas variáveis COM, e NS. A variância explicada depois da rotação Varimax, manteve-se similar à abordagem sem rotação, totalizando aproximadamente 87% para ambos os fatores como apresentado na Tabela 8.

A espessura do pericarpo foi uma das características que mais colaborou para a diversidade, estando associada ao fator 1 em ambas abordagens, resultados condizentes com Büttow *et al* (2010), no estudo da divergência genética entre 20 acessos de *C. annuum* com base em 34 descritores.

A importância da espessura do pericarpo do fruto de *C. annuum* está relacionada a susceptibilidade das cultivares a acidentes fisiológicos como o fendilhamento dos frutos e também é uma característica de agregação de valor

econômico, uma vez que correlaciona com o rendimento da polpa, a qual geralmente é utilizada tanto *in natura* como industrializada na forma de condimentos, molhos, páprica, entre outros (ABUD *et al.*, 2018).

Em relação ao Fator 2, tem-se a variável COM e NS apresentando significativa importância para o estudo da divergência genética o que corrobora resultados encontrados por Bento *et al* (2007) que verificaram variação de 10,24 mm a 143 mm no comprimento de frutos do gênero *Capsicum* e afirmam que, em função desta variável, pode se determinar a forma de uso mais adequada para o acesso. Os frutos menores demonstraram um potencial maior para serem empregadas em molhos, enquanto os frutos de tamanho maior foram mais propensos a serem comercializados *in natura* ou desidratados. Já o número de sementes por frutos variou de 21 a 299 unidades, indicando sua importância para a variabilidade genética.

Tabela 8 – Variância explicada pelos fatores

Tipo de rotação	Proporção de variância explicada	
	Fator 1	Fator 2
Sem rotação	0,8	0,08
Varimax	0,56	0,31

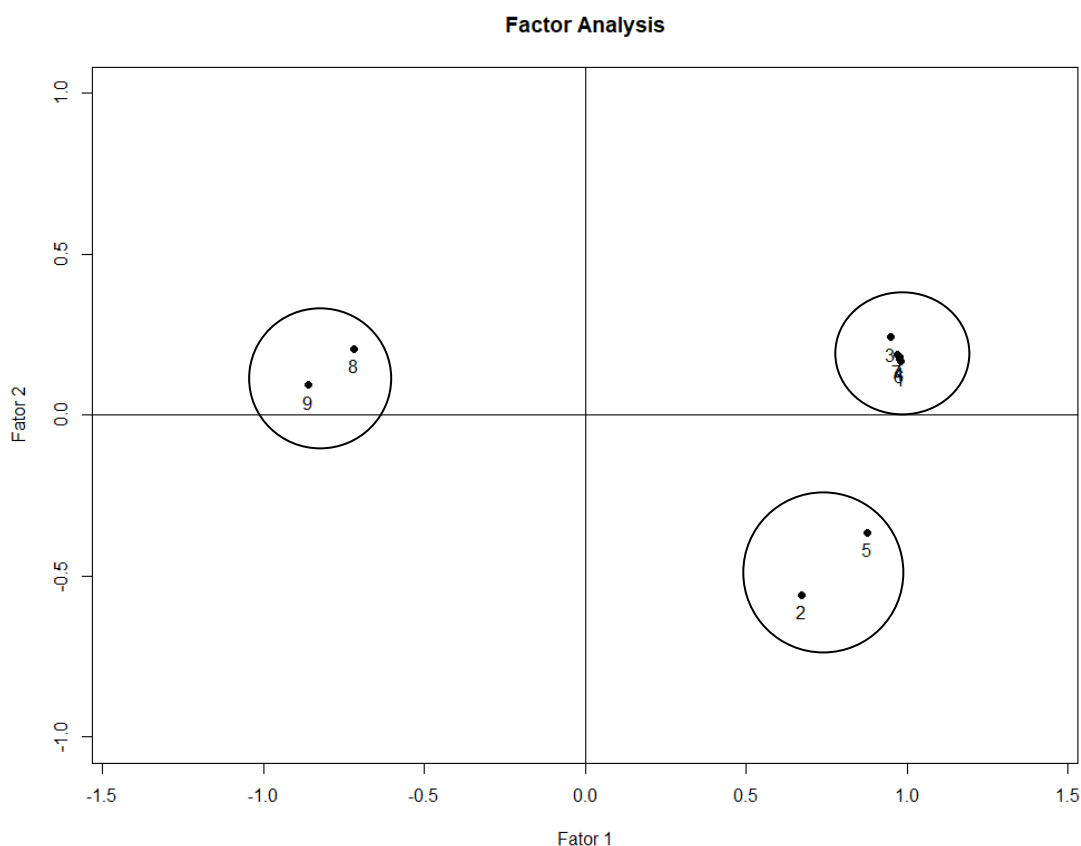
A aplicação da rotação ortogonal varimax, resultou num ajuste semelhante à abordagem sem rotação, devido a interpretação dos fatores já serem relativamente clara inicialmente.

A dispersão gráfica dos escores de nove acessos de pimenta em relação ao Fator 1 e Fator 2, está apresentada na figura 6. Pode-se verificar a formação de três grupos distintos, o primeiro formado pelos acessos 1, 3, 4, 6 e 7 que se caracterizaram por possuírem frutos maduros com comprimento entre 17,87 mm a 121,57 mm, peso total do fruto e número de sementes por frutos altamente variável (0.67g a 109.92g) e (32 a 152 sementes) respectivamente.

O segundo grupo é composto pelos acessos 2 e 5, cujas características têm pouca divergência, peso do fruto maduro relativamente baixo, espessura variando entre 0,91 mm a 1,27 mm e alto teor de vitamina C (109,85 mg100⁻¹ e 98,96 mg100⁻¹ respectivamente). A ingestão recomendada de vitamina C para suprir as necessidades diárias de um indivíduo adulto é de 60 mg (PINTO *et al.*, 2013).

O grupo 3 constituiu-se dos acessos 8 e 9 e apresentou valores medianos altos em todas as características estudadas. Destacando principalmente o comprimento total do fruto maduro e o número de sementes.

Figura 6 – Dispersão gráfica dos acessos em relação aos dois primeiros fatores.



Quanto menor a distância entre os acessos, mais similaridade entre eles são apresentadas (REGAZZI; CRUZ, 2020). Por outro lado, quando mais distantes se encontram os acessos, mais dissimilaridades são apresentadas entre eles. Os programas de hibridação buscam por acessos dissimilares, de forma que possam resultar em populações com grande variabilidade e, portanto, novos cruzamentos devem ser feitos a partir de acessos do grupo 2 e 3 que se encontram mais distantes.

Também deve-se observar que frutos grandes e longos geralmente são mais atraentes para o mercado de pimenta fresca, logo os acessos dos grupos 1 e 3 podem ser utilizados para a obtenção de híbridos com esta finalidade.

Os objetivos estabelecidos foram alcançados e acredita-se que esses resultados fornecerão bases para o desenvolvimento de estratégias de melhoramento genético de pimentas.

Por outro lado, pode haver limitação desse estudo pela limitação dos dados, ou seja, os dados fornecidos podem ser limitados em termos de tamanho da amostra, variedades representadas e informações adicionais sobre as características das pimentas, o que restringe a generalização dos resultados.

O processo de melhoramento genético frequentemente envolve a seleção de indivíduos com características específicas para cruzamento, o que pode levar à redução da diversidade genética dentro de uma determinada espécie de planta. Isso pode tornar as plantas mais suscetíveis a mudanças ambientais, dificultando sua adaptação a novos desafios.

A utilização responsável das técnicas de melhoramento genético, juntamente com uma avaliação cuidadosa dos riscos e benefícios envolvidos, é essencial para maximizar os resultados positivos e mitigar os pontos negativos associados a essa prática.

5. CONCLUSÕES

Com base nos resultados obtidos, a aplicação das técnicas de Análise de Componentes Principais (ACP) e Análise de Fatores em acessos de *Capsicum annuum* L. revelou divergência genética entre os acessos de pimentas avaliados.

Através da ACP, foi possível identificar dois componentes principais que explicam mais de 90% da variabilidade total. O componente principal 1 com maior contribuição das variáveis: peso total do fruto e massa da matéria total do fruto maduro fresco. E componente principal 2 que foi influenciado em maior parte pela variável de comprimento do fruto maduro e número de sementes.

A aplicação da Análise Fatorial contribuiu para explicar as correlações entre as variáveis através de um menor número de fatores e permitiu identificar três grupos geneticamente distintos. O primeiro formado pelos acessos 1, 3, 4, 6 e 7; o segundo é composto pelos acessos 2 e 5 e o terceiro contendo os acessos 8 e 9.

6. REFERÊNCIAS BIBLIOGRÁFICAS

ABUD, H. F. et al. **Caraterização morfológica dos frutos de pimentas malagueta e biquinho**. 2018. Disponível em: <https://www.locus.ufv.br/handle/123456789/24033>. Acesso em 04 mai 2023.

BLAT, S. F. **Herança da reação de *Capsicum* spp. ao oídio (*Leveillula taurica* (lev.) arn.)**. 2004. Tese (Doutorado em Fitotecnia) - Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, 2004. doi:10.11606/T.11.2004.tde-15102004-132941. Acesso em: 26 jun 2022.

BENTO, C.S. et al. Descritores qualitativos e multicategóricos na estimativa da variabilidade fenotípica entre acessos de pimenta. **Scientia Agraria**, v. 8, p.149-157, 2007.

BÜTTOW, M. V., et al. **Diversidade genética entre acessos de pimentas e pimentões da Embrapa Clima Temperado**. *Ciência Rural*, 2010, 40: 1264-1269.

CASALI, V. W. D.; Stringheta, P.C. Melhoramento de pimentão e pimenta para fins industriais. **Informe Agropecuário** 113: 23-25. 1984.

COSTA L. V., et al. Caracterização morfológica de pimentas (*Capsicum* spp.) do Amazonas. **Horticultura Brasileira** 29: S3402-S3410

COSTA, F. R. **Diversidade genética entre acessos de *Capsicum* spp. com base em marcadores RAPD**. 2004. Universidade Estadual do Norte Fluminense Darcy Ribeiro; Campos dos Goytacazes- RJ.

FAO. (2021). FAOSTAT. Disponível em: <http://www.fao.org/faostat/en/#data/QC/>. Acesso em: 06 jan. 2023.

FERRÃO, J. E. M. **A aventura das plantas e os descobrimentos portugueses**. 3.a ed. rev. e atualizada. 2005 Lisboa: Inst. de Investigação Científica Tropical: Fundação Berardo Chaves Ferreira. 287 págs. ilustrada.

FERRÃO, L. F. V., et al. Divergência genética entre genótipos de pimenta com base em caracteres morfoagronômicos. **Horticultura Brasileira**, vol. 29, no. 3, 2011. pp. 354-358. <http://dx.doi.org/10.1590/S0102-05362011000300016>.

FERREIRA, D. F. **Análise multivariada**. 3. Ed. Lavras: UFLA, 2018. p. 394.

FIGUEIREDO, A. N, et al. "Divergência genética em acessos de mamona (*Ricinus communis* L.) baseada nas características das sementes. **Revista de Biologia e Ciências da Terra** 4.2, 2004.

GUIMARÃES, M. E. S. et al. Avaliação do potencial ornamental de genótipos de pimenta com base em análise fatorial. **Research, Society and Development**, v. 9, n. 11, e66291110353, 2020.

GODOI, C. R. Análise estatística multidimensional. In: **Simposio de estatística aplicada a experimentação agrônômica**, Piracicaba:USP/ESALQ, 1985, 187p.

GRAND VIEW RESEARCH. Tamanho do mercado de especiarias, relatório de análise de participação e tendências por produto (pimenta, cúrcuma), por forma (pó, inteiro, picado e triturado), por região (América do Norte, Europa, APAC, CSA, MEA) e previsões de segmento, 2020 – 2028. Disponível em: <https://www.grandviewresearch.com/industry-analysis/spices-market>. Acesso em 03/08/2023.

HAIR JR., et al. **Análise Multivariada de Dados**. 6. ed. – Dados eletrônicos. – Porto Alegre : Bookman, 2009.

IVOGLO, M. G. *et al.* **Divergência genética entre progênies de café robusta**. *Bragantia*, Campinas, v. 67, n. 4, p. 823-831, Out./Dez. 2008. Disponível em: <https://www.scielo.br/j/brag/a/83swHNBBLsWwfh9zFjgHwzc/?format=pdf&lang=pt>. Acesso em 14 jun 2022.

KIRCH, J. L. et al. **Aplicação da Análise de Componentes Principais e de Agrupamento para os Indicadores de Desempenho das Universidades Federais do Brasil**. 2019. Disponível em: <https://publicacoes.unifal-mg.edu.br/revistas/index.php/sigmae/article/view/920>. Acesso em 22 dez 2022.

LOIOLA, B. H. M. **Análise multivariada e redes neurais aplicadas na predição de projeção no BOF**. Dissertação (Mestrado) – Universidade Federal de Ouro Preto, Ouro Preto, 2020. Disponível em: https://www.repositorio.ufop.br/bitstream/123456789/13127/1/DISSERTA%C3%87%C3%83O_An%C3%A1liseMultivariadaRede.pdf. Acesso em 28 jan de 2022.

LONGATTI, B. F. **Melhoramento genético de pimenta do tipo Jalapeño (*Capsicum annuum* L)**. 2019. Tese de doutorado – Universidade de São Paulo, Piracicaba.

MINGOTI, S. A. **Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada**. Belo Horizonte: Editora UFMG, 2020.

MATOS, D. A. S.; RODRIGUES, E. C. **Análise fatorial**. Brasília: Enap, 2019. 74 p. : il. ISBN: 978-85-256-0118-6.

Nuez, F.; Ortega, R. G.; costa, J. 1996. **El cultivo de pimientos, chiles y ajies**. Madri: Mundi-Prensa. 607 p.

PINTO, C.M.F.; PINTO, C.L.O.; DONZELES, S.M.L. **Pimenta *Capsicum*: Propriedades químicas, nutricionais, farmacológicas e medicinais e seu potencial para o agronegócio**. *Revista Brasileira de Agropecuária Sustentável (RBAS)*, v.3, n.2, p.108- 120, 2013.

PORTAL SÃO FRANCISCO. **Pimenta**. Disponível em: <https://www.portalsaofrancisco.com.br/alimentos/pimenta>. Acesso em: 27 jan. 2023.

R CORE TEAM. R: **A language and environment for statistical computing**. Vienna: R Foundation for Statistical Computing, 2020.

REGAZZI, A.J; CRUZ, C.D. **Análise Multivariada Aplicada**. Viçosa, MG, 2020. p. 1-401.

RÊGO, E.R.; FINGER, F.L.; RÊGO, M.M. **Produção, Genética e Melhoramento de Pimentas (*Capsicum spp.*)**. Recife: Imprima, 2011, 223p

RÊGO, E. R. *et al.* Influência de substratos alternativos para produção de pimenteira ornamental (*Capsicum annuum* L). **Revista AGROTEC**.v.34. n1, p.21-29, 2013.

REIFSCHNEIDER, F. J. B. (Org.). **Capsicum: pimentas e pimentões no Brasil**. Brasília: Embrapa Comunicação para Transferência de Tecnologia/ **Embrapa Hortaliças**, 2000. 113p.

RIBEIRO, C. S. C.; HENZ, G. P. Pimenta (*Capsicum spp.*) **Embrapa Hortaliças**. Sistemas de Produção, 2 ISSN 1678-880x Versão Eletrônica. Nov./2007. Disponível em:
https://sistemasdeproducao.cnptia.embrapa.br/FontesHTML/Pimenta/Pimenta_capsicum_spp/botanica.html. Acesso em 28 set 2022.

RIBEIRO, C. S. C. et al. **Embrapa Socioeconomia** - Pré-produção de pimenta. Fev./2022. Disponível em: <https://www.embrapa.br/agencia-de-informacao-tecnologica/cultivos/pimenta/pre-producao/socioeconomia>. Acesso em: 03 mar 2023.

SANTOS, V. S. F. **Caracterização morfológica e determinação da pungência em pimentos picantes**. PhD Thesis. Universidade Tecnica de Lisboa (Portugal). 2009

SANTOS, C.A.P. **Diversidade genética acessada por RAPD e indução de variabilidade fenotípica em pimenteira ornamental mediada por giberelina (*Capsicum annuum* L.)**. Dissertação (Mestrado em Agronomia) - Centro de Ciências Agrárias, Universidade Federal da Paraíba, Areia, 2018.

SUDRÉ, C.P., *et al.* Divergência genética entre acessos de pimenta e pimentão utilizando técnicas multivariadas. **Horticultura brasileira**, 2005, 23: 22-27.

SUDRÉ, C.P; et al. Genetic variability in domesticated *Capsicum spp.* as assessed by morphological and agronomic data in mixed statistical analysis. **Genetics and Molecular Research**, 9: 283-294.

VARELLA, C. A. A. **Análise de componentes principais**. Departamento de Engenharia da Universidade Federal Rural do Rio de Janeiro, Seropédica – RJ, 2008. Disponível em:
<http://www.ufrj.br/institutos/it/deng/varella/Downloads/multivariada%20aplicada%20a%20s%20ciencias%20agrarias/Aulas/analise%20de%20componentes%20principais.pdf>. Acesso em 17 jan 2023.

VIALI, L. **Série: Estatística Multivariada**. Porto Alegre, 2005.

VICINI, Lorena. **Análise multivariada: da teoria à prática**. Santa Maria: UFSM, CCNE, 2005. 215 p.: il.

ZENI, A.L.; BOSIO, F. **O uso de plantas medicinais em uma comunidade rural de Mata Atlântica** – Nova Rússia, SC. *Neotropical Biology and Conservation*, São Leopoldo, v.6, n.1, p.55-63, 2011. Disponível em: https://pdfs.semanticscholar.org/2d86/a86aa14273f5d008ec5ef429c4dfce83a665.pdf?_ga=2.18994168.1937913529.1640552721-344555880.1640552721. Acesso em: 01 out. 2021. doi: 10.4013/nbc.2011.61.07.

ZOZ, T. **Avaliação de genótipos de cártamo quanto ao desempenho agrônômico, divergência genética e produtividade da água**. Tese - Universidade Estadual Paulista, Faculdade de Ciências Agrônômicas, Botucatu, 2015.