



Simpósio de Integração Acadêmica

“Bicentenário da Independência: 200 anos de ciência, tecnologia e inovação no Brasil e 96 anos de contribuição da UFV”

SIA UFV 2022



A IMPORTÂNCIA DA PADRONIZAÇÃO DE DADOS EM ANÁLISE DE AGRUPAMENTO

Thaynara Aparecida de Souza Neto^{1a}, Jhennifer dos Santos Nascimento^{1b}, Alice dos Santos Ribeiro^{1c}, Paulo Roberto Cecon^{1d}
¹Departamento de Estatística, Programa de Pós-graduação em Estatística Aplicada e Biometria, Universidade Federal de Viçosa

^athaynara.neto@ufv.br, ^bjhennifer.nascimento@ufv.br, ^calice.ribeiro@ufv.br, ^dcecon@ufv.br

Ciências Agrárias – Probabilidade e Estatística

Categoria: Pesquisa

Palavras-Chave: Análise de agrupamento, *Capsicum annuum* L., Padronização de dados

Introdução

A padronização de dados faz-se de extrema importância, pois evita que as escalas escolhidas para medir as variáveis afetem a similaridade entre os indivíduos, além de fazer com que as variáveis com grande dispersão dominem a classificação das distâncias. Portanto, ao padronizar os dados e aplicar uma técnica de análise de agrupamento, que requer uma medida de dissimilaridade para informar o grau de semelhança ou de diferença entre os indivíduos, faz com que retorne grupos mais compactos e não-viesados para as variáveis.

Resultados e Discussão

Por meio das estimativas das distâncias euclidianas, os acessos que apresentaram menor dissimilaridade foram o 1 e 5, com $d_{15} = 0,529$; já os acessos 4 e 6 apresentaram maior dissimilaridade com $d_{46} = 5,097$. Ao aplicar o método UPGMA, obtivemos dois grupos, compostos pelo genótipo 6, 8, 7 e 9 no primeiro grupo e os genótipos 1, 2, 3, 4, 5 no grupo 2. Ao calcular o coeficiente de correlação cofenética, obteve-se um valor igual a 0,819 que, segundo Vaz Patto *et al.* (2004), é considerado adequado, pois reflete uma boa consistência de agrupamento.

Objetivos

Os objetivos deste trabalho são:

- Mostrar a importância da padronização de dados na análise de agrupamento;
- Fazer uma aplicação em um conjunto de dados reais.

Conclusões

Pode-se concluir que a padronização dos dados foi significativa, pois permitiu que dados mensurados em unidades de medida diferentes, fossem avaliados de maneira equivalente e também permitiu que os mesmos tivessem uma boa consistência de agrupamento.

Material e Métodos

Para exemplificar a técnica, utilizou-se dados de um experimento com as seguintes especificações:

- DIC, com quatro repetições;
- **Tratamentos:** Nove genótipos de *Capsicum annuum* registrados no Banco de Germoplasma de Hortaliças da UFV;
- **Características avaliadas:** massa da matéria total do fruto maduro fresco, número de sementes por fruto, comprimento do fruto maduro e altura da planta;

Além disso, utilizou-se a Distância Euclidiana como medida de dissimilaridade e o método UPGMA como método de agrupamento.

Bibliografia

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

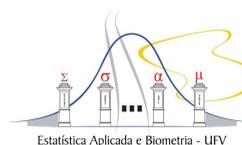
CRUZ, C. D.; FERREIRA, F. M.; PESSONI, L. A. *Biometria aplicada ao estudo da diversidade genética*. Visconde do Rio Branco: Ed. Suprema, 620p., 2011.

VAZ PATTO, M.C; SATOVIC, Z; PÊGO, S; FEVEREIRO, P. Assessing the genetic diversity of Portuguese maize germoplasm using microsatellite markers. *Euphytica*, Wageningen, v. 137, n.1, p. 63 – 72, 2004.

Apoio Financeiro



Agradecimentos



DET

UFV
Universidade Federal
de Viçosa