

## ANÁLISE DE DADOS ESTATÍSTICOS REFERENTES À DESTINAÇÃO DOS SOLOS PARA PRODUÇÃO DE CAFÉ NOS ESTADOS BRASILEIROS

LUCAS MARUJO<sup>1</sup>, CARLA ADRIANA PIZARRO SCHMIDT<sup>2</sup>, NEUSA IDICK SCHERPINSKI<sup>3</sup>, MARCELO ANDERSON CARLET<sup>4</sup>, ANDRÉ SANDMANN<sup>5</sup>, JOSE AIRTON AZEVEDO DOS SANTOS<sup>6</sup>

<sup>1</sup>Mestrando em Sistema Computacionais para o Agronegócio, UTFPR, lucasmarujo@alunos.utfpr.edu.br

<sup>2</sup>Professora Agrônoma Doutora em Agronomia, UTFPR, Medianeira-PR, carlaschmidt@utfpr.edu.br

<sup>3</sup>Professora de Cálculo e Estatística; Mestre em Engenharia Agrícola, UTFPR, Medianeira-PR, neusascherpinski@gmail.com

<sup>4</sup>Mestrando em Sistema Computacionais para o Agronegócio, UTFPR, (45) 9 9974-7810 marcelocarlet@yahoo.com

<sup>5</sup>Professor de Cálculo e Estatística; Doutor em Engenharia Agrícola, UTFPR, Medianeira-PR, sandmann\_andre@hotmail.com

<sup>6</sup>Engenheiro Eletricista, Dr. Engenharia Elétrica, Professor da UTFPR, Medianeira-PR, airton@utfpr.edu.br

Apresentado no  
XLIX Congresso Brasileiro de Engenharia Agrícola - CONBEA 2020  
23 a 25 de novembro de 2020 - Congresso On-line

**RESUMO:** As áreas destinadas ao cultivo de café em geral seguem sistemas sustentáveis e ajudam a conservar o solo e a água, sendo dessa forma importantes para manutenção da sustentabilidade ambiental das áreas agrícolas do Brasil. Este estudo teve o intuito de analisar a distribuição espacial do cultivo de café no Brasil por meio de técnicas estatísticas de agrupamento. Dessa forma, aplicou-se o algoritmo k-means e uma análise de componentes principais (ACP) à dados de área cultivada das duas espécies do gênero *Coffea* em todos os estados. Optou-se pela divisão em cinco grupos aos quais destacaram-se os estados principalmente devido as espécies cultivadas e quantidade de área destinada para cultivo. Os estados de MG, SP, ES e BA se destacaram em relação a destinação de áreas para essa cultura o que confirma o raciocínio de que estados que possuem elevadas áreas de preservação ambiental possuem também grandes áreas ocupadas com café.

**PALAVRAS-CHAVE:** *Coffea* sp.; Gestão Ambiental; Uso e Ocupação de Solos.

### ANALYSIS OF STATISTICAL DATA RELATING TO THE DESTINATION OF SOILS FOR COFFEE PRODUCTION IN BRAZILIAN STATES

**ABSTRACT:** The areas destined for the cultivation of coffee in general follow sustainable systems and help to conserve soil and water, thus being important for maintaining the environmental sustainability of agricultural areas in Brazil. This study aimed to analyze the spatial distribution of coffee cultivation in Brazil using statistical clustering techniques. Thus, the k-means algorithm and a principal component analysis (PCA) were applied to the cultivated area data of the two species of the genus *Coffea* in all states. It was decided to divide it into five groups to which the states stood out mainly due to the cultivated species and the amount of area destined for cultivation. The states of MG, SP, ES and BA stood out in relation to the allocation of areas for this culture, which confirms the reasoning that states that have high areas of environmental preservation also have large areas occupied with coffee.

**KEYWORDS:** *Coffea* sp.; Environmental management; Land Use and Occupation.

**INTRODUÇÃO:** Apenas duas espécies do gênero *Coffea* são amplamente cultivadas e são denominadas de *Coffea arabica* a qual é nativa das terras altas do sudoeste da Etiópia e o *Coffea canéfora* nativo da África central e ocidental, sendo que de forma comum são denominadas respectivamente de café Árábica e café Robusta. O primeiro representa 65% da produção mundial e é mais aromático e o segundo representa 35% sendo mais produtivo e

resistente, porém menos aromático (SIMON-GRUITA et al., 2019). O café tem uma grande relevância histórica e econômica para o Brasil, segundo dados da International Coffee Organization (ICO, 2020) o país é o maior produtor e exportador do mundo, além de ser também um dos maiores consumidores do produto. Dessa forma, tem-se clara a importância de estudos que colaborem para o entendimento dessa commodity. De acordo com Evaristo de Miranda, chefe geral da Embrapa territorial (EMBRAPA, 2018): “as principais regiões produtoras de café no País estão localizadas em estados onde os imóveis rurais possuem, na média, porcentagem de área dedicada à preservação da vegetação nativa acima do valor exigido pelo Código Florestal”. O agrupamento de dados multivariados é uma das tarefas mais importantes no campo de mineração de dados e a análise de componentes principais (ACP), é frequentemente usada para redução de dimensionalidade (LI, 2019). Nesse estudo objetivou-se aplicar essas técnicas na avaliação da distribuição do cultivo de cafés das espécies Arábica e Robusta nos estados brasileiros e verificar a formação de grupos de estados com comportamento semelhante em relação ao cultivo.

**MATERIAL E MÉTODOS:** Para os procedimentos executados neste trabalho, inicialmente efetuou-se a coleta de dados diretamente na plataforma do IBGE no banco de dados SIDRA na tabela de dados número 1618 para o ano de 2019 (IBGE, 2020). Os dados coletados foram tabulados e agrupados por estados de acordo com os números referentes as áreas de cultivo (plantio e colheita) e produção do ano de 2019 por meio da técnica de agrupamento, reduziu-se posteriormente a dimensionalidade dos dados por meio de uma Análise de Componentes Principais (ACP). Os dados tabulados foram introduzidos no software livre de mineração de dados Tanagra (2020) e as técnicas de análise estatística multivariada de agrupamento por meio de K-means e de a ACP foram aplicadas aos dados completos e aos dados apenas dos estados que produzem café de qualquer uma das duas espécies, café Arábica ou café Robusta.

**RESULTADOS E DISCUSSÃO:** Percebeu-se que o cultivo de café no Brasil ocupa mais de 2,5% das áreas cultivadas, cabe destacar que essas lavouras podem ser conduzidas de maneira sustentável bastando para isso seguir as Boas Práticas Agrícolas descritas por Alixandre et al. (2020), o que possibilita produzir cafés de qualidade e ainda ajudar na preservação do meio ambiente. Em uma breve análise estatística foi possível entender a proporção da distribuição de solo nos estados, em relação à área plantada, colhida e a produção, das duas espécies de café abordadas, dessa forma, identificou-se características dos estados em termos nacionais. Essa relação está ilustrada na Figura 1.

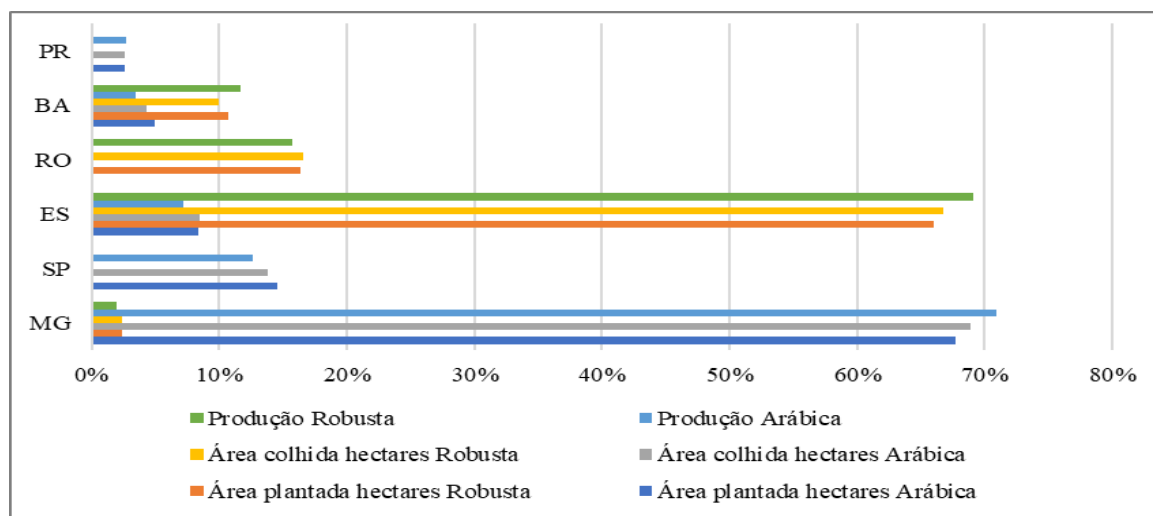


Figura 1. Relação de área plantada, colhida e produção entre os estados mais significativos.

Os estados que apresentaram resultados de maior destaque foram MG, SP e PR, em relação ao café Arábica enquanto que os estados de ES, RO, e BA, dividem a representatividade do café Robusta. Nota-se a desproporcional dimensão de MG e ES em suas respectivas principais espécies de cultivo. Após análise e aplicação de agrupamento por meio do algoritmo K-means, decidiu-se que o melhor resultado foi observado quando se separou os estados em 5 grupos. O grupo 1 ficou apenas com o estado de SP, o principal fator que foi relevante nesse grupo foi a área destinada ao cultivo do café Arábica que é bastante representativa nesse estado, apenas o estado de MG apresenta área superior destinada ao cultivo dessa espécie. Exatamente por esse motivo, o grupo 4 ficou também com apenas um estado que foi exatamente MG, os fatores que foram relevantes para separação desse grupo foram a produção e área destinada ao cultivo de café Arábica bem superior a todos os outros estados do Brasil, cabe destacar que esse estado também destina áreas ao cultivo ao Robusta bem significativas, diferente do estado de SP que possui muito pouco cultivo dessa espécie. O Grupo 2 ficou apenas com o estado de ES, sendo que o fator primordial para formação desse grupo foi a produção de café Robusta, bem como as áreas destinadas ao cultivo dessa espécie que são as maiores do Brasil, em relação ao cultivo de Arábica esse estado fica em terceiro lugar. O grupo 4 ficou com 2 estados BA e RO, devido principalmente a produção de café Arábica diferindo do grupo 3 apenas pelo fato das áreas destinadas ao cultivo serem maiores que as dos demais estados analisados. Dessa forma, o grupo 3 ficou com os 22 demais estados pois são os que não apresentam grande representatividade no cultivo de café, sendo que alguns não apresentam nenhuma área destinada ao cultivo dessa planta, mas a separação em 5 grupos não foi suficiente para separar em um único grupo os estados que não possuem a atividade. Testou-se ainda o agrupamento sem a inclusão desses estados no banco de dados, mas com ou sem os estados que não cultivam café presentes na análise a separação em grupos não se alterou, portanto optou-se por apresentar os resultados que contem no banco de dados os 27 estados brasileiros. Após aplicação de uma análise de componentes principais, pode-se observar claramente a separação dos estados citados dentro dos grupos. As duas componentes foram capazes de explicar 99,96% da variabilidade dos dados e obteve-se índice para o critério de Kaiser-Meyer-Olkin (KMO) acima de 70%, o que caracterizaria uma boa correlação entre os dados utilizados na análise de componentes principais. A Figura 2 demonstra o posicionamento dos estados nas duas dimensões fornecidas pela análise de componentes principais.

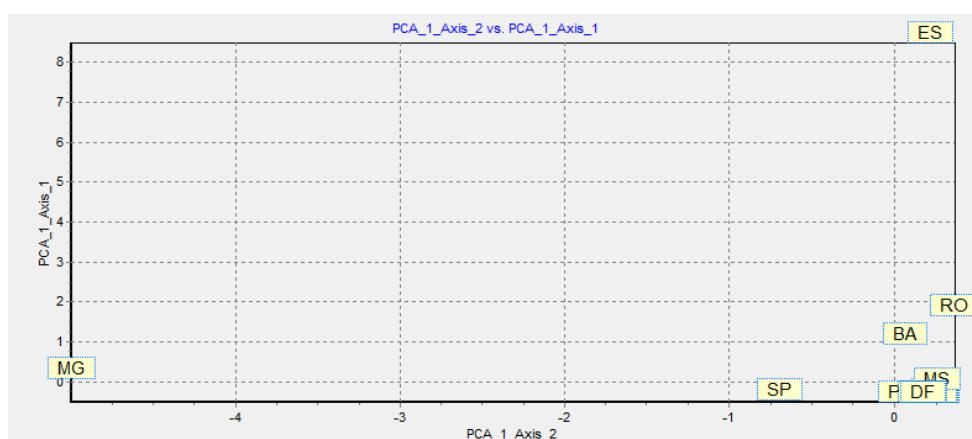


Figura 2. Gráfico da ACP com a posição dos estados nas duas componentes principais.

Os resultados encontrados pela análises estatísticas realizadas por meio de algoritmos de clusterização e análise de componentes principais aplicadas aos dados do ano de 2019 disponibilizados pelo site do IBGE foram bem condizentes com as informações fornecidas pelo ministério da Agricultura (2017), o que nos leva a crer que não houveram grandes

mudanças no cenário de destinação de área para o cultivo do café nesses últimos anos. De acordo com o Ministério da Agricultura (2017): atualmente a produção da espécie Arábica está concentrada nos Estados de MG, como maior produtor, seguido por SP, ES e BA. Esses quatro estados concentram 85% da produção nacional dessa espécie, segundo a Companhia Nacional de Abastecimento – Conab. O Conilon ou Robusta é cultivado principalmente no ES, BA e RO, concentrando 95% da produção nacional. Cabe ainda destacar que, de acordo com Evaristo de Miranda, chefe geral da Embrapa territorial (EMBRAPA, 2018), a maioria desses estados possuem elevadas áreas de preservação sendo que em MG, o valor é de 34%; no ES 33%; SP 22%, e na BA 45% das áreas dedicadas à preservação, o que confirma que os maiores estados produtores de café também possuem extensas áreas destinadas à preservação ambiental.

**CONCLUSÕES:** Observou-se, com a aplicação de técnicas de estatística multivariada aos dados da análise da distribuição do cultivo de café no solo brasileiro, que as análises foram condizentes com as características específicas dos estados produtores, na qual a escolha por um agrupamento em 5 grupos, possibilitou dividir de forma precisa em relação as principais diferenças. Com a ACP, pode-se ainda identificar dentre esses grupos os critérios mais relevantes adotados pelo *software* no agrupamento. Denota-se ainda, que devido a grande diferença entre os estados de MG e ES em relação aos demais, era esperado que esses fossem representados de forma destacada, o que de fato aconteceu, isso corrobora para justificar que os objetivos primordiais desse estudo foram alcançados. Por fim, destaca-se a importância de estudos relacionados a café no Brasil, devido a relevância histórica e econômica para o país. Apresenta-se ainda nesse contexto, a possibilidade de trabalhos futuros que se aprofundem na investigação da destinação do solo cafeeiro, com o intuito de levantar características mais específicas desse cenário e sua correlação com a preservação ambiental.

**AGRADECIMENTOS:** O primeiro autor agradece ao apoio com a bolsa de mestrado concedida pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES).

## REFERÊNCIAS:

- ALIXANDRE, Fabiano Tristão et al. **Cafeicultura sustentável:** boas práticas agrícolas para o café arábica – Vitória, ES: Incaper, 2020. 48 p.: il. Color. – (Incaper, Documentos, 269).
- EMBRAPA. **Papel da cafeicultura na preservação ambiental é tema de palestra.** Gestão Ambiental e Territorial, 2018. Disponível em: <https://www.embrapa.br/busca-de-noticias/-/noticia/34621789/papel-da-cafeicultura-na-preservacao-ambiental-e-tema-de-palestra>. Acesso em: 02 mar. 2020.
- IBGE. **Levantamento Sistemático da Produção Agrícola.** Tabela 1618 - Área plantada, área colhida e produção, por ano da safra e produto das lavouras. Disponível em: <https://sidra.ibge.gov.br/Tabela/1618>. Acesso em: 20 fev. 2020.
- ICO. **Historical Data on the Global Coffee Trade.** 2020. Disponível em: [http://www.ico.org/new\\_historical.asp](http://www.ico.org/new_historical.asp). Acesso em: 03 mar. 2020.
- LI, H. Multivariate time series clustering based on common principal component analysis. **Neurocomputing**, [s.l.], v. 349, p.239-247, jul. 2019. Elsevier BV. <http://dx.doi.org/10.1016/j.neucom.2019.03.060>.
- MINISTÉRIO DA AGRICULTURA. **Café no Brasil.** 2020. Disponível em: <http://www.agricultura.gov.br/assuntos/politica-agricola/cafe/cafeicultura-brasileira>. Acesso em: 27 de jan. 2020.
- SIMON-GRUITA, A., POJOGA, M. D., CONSTANTIN, N., & DUTA-CORNESCU, G. **Genetic Engineering in Coffee.** Caffeinated and Cocoa Based Beverages, 447–488. 2019. doi:10.1016/b978-0-12-815864-7.00014-3
- TANAGRA. **Tanagra Project.** Disponível em: <http://eric.univ-lyon2.fr/~ricco/tanagra/en/tanagra.html>. Acesso em: 20 fev. 2020.