

Universidade Federal de Alfenas – UNIFAL-MG

Instituto de Ciências Sociais Aplicadas – ICSA

Isabella Lucinda da Ponte Floriano

**As principais culturas agrícolas de Minas Gerais:  
uma análise multivariada**

Varginha

2019

Universidade Federal de Alfenas – UNIFAL- MG

Instituto de Ciências Sociais Aplicadas – ICSA

Isabella Lucinda da Ponte Floriano

**As principais culturas agrícolas de Minas Gerais:  
uma análise multivariada**

Trabalho de conclusão de Piepex apresentado ao Instituto de Ciências Sociais Aplicadas da Universidade Federal de Alfenas como requisito parcial para a obtenção do título de Bacharel em Ciência e Economia.

Orientadora: Patrícia de Siqueira Ramos.

Varginha

2019

Isabella Lucinda da Ponte Floriano

**As principais culturas agrícolas de Minas Gerais:  
uma análise multivariada**

A banca examinadora, abaixo assinada, aprova o Trabalho de conclusão de Piepex apresentado ao Instituto de Ciências Sociais Aplicadas da Universidade Federal de Alfenas como requisito parcial para a obtenção do título de Bacharel em Ciência e Economia.

Prof. Patrícia Siqueira Ramos

Universidade Federal de Alfenas – campus Varginha Assinatura: \_\_\_\_\_

Prof. Danilo Machado Pires

Universidade Federal de Alfenas – campus Varginha Assinatura: \_\_\_\_\_

Prof. Lincoln Thadeu Gouvêa de Frias

Universidade Federal de Alfenas – campus Varginha Assinatura: \_\_\_\_\_

Varginha - MG  
2019

## Resumo

Este trabalho tem como objetivo analisar as culturas agrícolas mais influentes para o estado de Minas Gerais no ano de 2017. Para tal estudo, foram utilizados dados fornecidos pela pesquisa Produção Agrícola Municipal (PAM) do IBGE. Foram selecionadas duas variáveis, rendimento (kg/ha) e valor (R\$), de oito das principais culturas agrícolas do estado: abacaxi, alho, batata inglesa, café arábica, feijão, milho, laranja e tomate. Sendo assim, foram observadas dezesseis variáveis, cujas análises foram efetuadas utilizando a linguagem de programação *Python*. A primeira análise aplicada foi o método multivariado dos componentes principais, seguido dos métodos de agrupamento hierárquico (Ward) e não hierárquico ( $k$ -médias). Os resultados sugerem a separação dos municípios em três grupos, que nos diferentes métodos tendem a separar grupos entre frutas, grãos e um terceiro grupo com relação aos dois tipos de variáveis. Em relação aos componentes principais, os resultados destacam os municípios de Frutal, Patrocínio e Uberlândia como grandes produtores de frutas e grãos.

**Palavras-chave:** agricultura; Minas Gerais; análise de agrupamento; componentes principais.

## Sumário

1. Introdução .....	6
2. Referencial Teórico .....	7
2.1. Produção Agrícola em Minas Gerais .....	7
2.2. Análise Multivariada .....	8
3. Metodologia .....	10
4. Resultados e discussão .....	13
4.1. Análise Descritiva .....	13
4.2. Análise de Componentes Principais (ACP) .....	15
4.3. Análise de Agrupamento (AA) .....	17
4.3.1. Método de Ward .....	17
4.3.2. Método <i>k-means</i> .....	20
5. Considerações Finais .....	23
Referências .....	25

## 1. Introdução

A agricultura sempre foi importante para economia brasileira e, observando os principais produtores agropecuários por estado, é possível verificar que Minas Gerais disputa o posto de segundo maior produtor nacional com Paraná (SILVA; SOUZA; MARTINS, 2012). Segundo Bastos (2010), a agricultura desempenha um papel fundamental no Estado de Minas Gerais, proporcionando a ele destaque entre os demais produtores nacionais. Em 2017, segundo dados da Secretaria do Estado de Agricultura, Pecuária e Abastecimento de Minas Gerais (SEAPA, 2019), o estado apresentou uma alta participação na produção nacional dos cultivos das culturas de abacaxi, alho, café, feijão, milho, dentre outras. Tendo um maior destaque na produção de café, representando quase 52% de toda a produção do Brasil.

Quando diversas variáveis são analisadas ao mesmo tempo, é essencial a utilização de uma técnica multivariada. Uma dessas técnicas é a análise de componentes principais (ACP) que, segundo Lattin et al. (2003), permite uma reorientação dos dados de tal forma que as primeiras dimensões consigam explicar a maior parte das informações. Outro método muito utilizado é a análise de agrupamento, que consiste na formação de grupos com observações que apresentem elevada homogeneidade interna e elevada heterogeneidade externa (HAIR et al., 2009).

Nesse contexto, o objetivo deste artigo é relacionar as produções agrícolas mais representativas em Minas Gerais no ano de 2017 com os municípios, e buscar compreender quais são as cidades que mais marcaram as produções do estado. Para obter esses resultados, as análises foram feitas utilizando a linguagem de programação *Python* e foram utilizados os métodos multivariados mencionados acima (análise de componentes principais e análise de agrupamento).

O texto está organizado da seguinte maneira: a seção 2 apresenta o referencial teórico, incluindo a produção agrícola do estado de MG e os métodos multivariados. Na seção 3 é apresentada a metodologia utilizada, explicitando os métodos, conceitos e dados utilizados. Na terceira seção estão apresentados os resultados obtidos com a análise e a discussão sobre eles. E a última seção apresenta as considerações finais sobre a pesquisa.

## 2. Referencial Teórico

### 2.1 Produção agrícola em Minas Gerais

Segundo Scolari (2006), o Brasil é um dos principais produtores mundiais de alimentos e fibras, participando com mais de 4% do valor total das exportações mundiais no agronegócio. Em 2004, Minas Gerais, Espírito Santo e São Paulo foram os maiores estados produtores agrícolas do país, com Minas Gerais representando 49% da produção total do país. Com isso, o estado demonstra ter uma forte influência e grande importância para a agricultura nacional.

A Tabela 1 contém um resumo das principais culturas cultivadas em MG em 2017 (SEAPA, 2019), qual a porcentagem que cada uma representa e sua posição no ranking nacional. Ela foi construída a partir dos dados do SEAPA (Secretaria do Estado de Agricultura, Pecuária e Abastecimento de Minas Gerais) da produção de 2017.

Tabela 1 - Representatividade de Minas Gerais na produção de algumas culturas do Brasil em 2017

Cultura	Porcentagem	Ranking
abacaxi	13,90%	2°
alho	39,10%	1°
café	51,90%	1°
feijão	16,60%	2°
laranja	5,60%	2°
milho	7,10%	5°
tomate	13,60%	3°
sorgo	38,30%	2°
trigo	3,20%	4°
cana de açúcar	11,30%	2°

Fonte dos dados: SEAPA. Elaboração própria.

Considerando os dados do SEAPA (2017), os municípios de Minas Gerais que tiveram maior destaque foram: Monte Alegre de Minas, Frutal, Comendador Gomes e

Prata, na produção de frutas, como abacaxi e laranja; Unaí, Patrocínio, Uberaba, Sacramento, Uberlândia e Perdizes no cultivo de grãos (milho, café e feijão).

De acordo com a pesquisa de Projeções do IBGE (2017), para a safra de 2016/2017, o Brasil produziu 238,8 milhões de toneladas de grãos. Minas Gerais possui uma produção muito diversificada de grãos, produzindo, principalmente, milho, feijão, soja, algodão, trigo, sorgo, amendoim, girassol e mamona (IBGE, 2017). Segundo dados do Ipeadata (2017), Patrocínio é um dos municípios de Minas Gerais onde a produção de café é muito comum, representando aproximadamente 57% de toda a área plantada da cidade, a produção de milho representa 22% e a de feijão cerca de 2%.

Como este trabalho utiliza várias variáveis como forma de analisar a produção agrícola do estado de Minas Gerais, a próxima seção apresenta algumas definições utilizadas.

## 2.2 Análise multivariada

De acordo com Hair et al. (2006), a análise multivariada pode ser definida como todas as técnicas estatísticas que analisam simultaneamente múltiplos dados sobre o objeto analisado. Essas técnicas são muito utilizadas pois permitem uma melhor organização das informações, resultando numa melhor tomada de decisão. Para as variáveis serem consideradas realmente multivariadas, todas elas devem ser aleatórias e inter-relacionadas de forma que não possam ser tão bem interpretadas se separadas.

A representação de dados multivariados se dá de forma parecida com planilhas eletrônicas. Considerando uma amostra aleatória de tamanho  $n$  e, para cada unidade amostral, há valores de  $p$  variáveis, cria-se uma matriz de dados  $\mathbf{X}$  com dimensão  $n$  (linhas) por  $p$  (colunas):

$$\mathbf{X}_{n \times p} = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1p} \\ X_{21} & X_{22} & \dots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{p1} & X_{p2} & \dots & X_{pp} \end{bmatrix},$$

em que cada observação é uma linha da matriz de  $\mathbf{X}$  e cada variável é uma coluna de  $\mathbf{X}$ .



Um exemplo de método multivariado é a análise de componentes principais (ACP). Segundo Lattin et al. (2011), a análise de componentes principais permite que os dados sejam reorientados de modo que as primeiras dimensões consigam explicar o maior número de informações possível. Para fazer uso dos componentes principais amostrais é necessário calcular seus escores  $e$ , de acordo com Mingoti (2005), esses valores são obtidos por meio dos autovetores da matriz de correlações. Dessa forma, os componentes serão os valores dos autovetores multiplicados por cada variável observada.

Assim, ainda segundo Mingoti (2005), para  $p$  variáveis, sejam  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$  os autovalores da matriz de correlação  $\mathbf{P}_{p \times p}$ , e  $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_p$  seus autovetores. Então, o  $j$ -ésimo componente principal é definido por:

$$Y_j = \mathbf{e}_j^T \mathbf{Z} = e_{j1}Z_1 + e_{j2}Z_2 + \dots + e_{jp}Z_p, \quad j = 1, 2, \dots, p,$$

em que  $\mathbf{Z}$  é o vetor que contém as  $p$  variáveis após uma normalização e T representa que tal vetor (ou matriz) é transposto.

Depois de calculados os componentes principais, observa-se o quanto da variância total cada componente explica. A variância de cada componente principal é representada pelo seu autovalor. Portanto, a proporção da variância total explicada pelo  $j$ -ésimo componente principal é igual a  $\frac{\lambda_j}{p}$ ,  $j = 1, 2, \dots, p$  (MINGOTI, 2005).

Assim como a análise de componentes principais, outro método multivariado amplamente utilizado é a análise de agrupamento (AA). Segundo Hair et al. (2006), a AA classifica as variáveis de modo que cada objeto é semelhante a outros dentro do grupo com determinada característica, e diferentes em relação a outros grupos. A análise de agrupamento consiste na divisão de uma grande grupo em grupos menores que sejam similares entre si. (LATTIN, 2003). Essa análise pode ser realizada por métodos hierárquicos e não hierárquicos. Existem muitos tipos de métodos hierárquicos, porém os 5 métodos mais comuns são: o vizinho mais próximo, que consiste em encontrar todas as distâncias entre as observações e unir as que forem mais próximas; vizinho mais distante, se calculam as distâncias, e reúnem em grupos as que possuem valores máximos entre as observações; distância média, consiste no cálculo da média da distância entre duas observações; centróide, método onde se calcula um valor de

centróide das observações e obtém a distância entre elas e o centróide, unindo em grupos mais próximos; e Ward, método que consiste na soma dos quadrados dentro dos agrupamentos feito sobre todas as variáveis. (HAIR et al., 2009).

O método hierárquico de Ward, segundo Hair et al. (2009), tende a combinar observações pequenas e grupos com aproximadamente o mesmo número de observações. Segundo Mingoti (2005), o procedimento de Ward, também conhecido como “mínima variância” se baseia nos princípios de que cada elemento é considerado um único grupo e, em cada passo, o algoritmo de agrupamento calcula a soma dos quadrados de cada conglomerado, essa soma é o quadrado da distância euclidiana, construída de forma matricial, ou seja:

$$SQ_i = \sum_{j=1}^{n_j} (X_{ij} - \bar{X}_i)^T (X_{ij} - \bar{X}_i),$$

em que  $X_{ij}$  representa cada observação pertencente ao grupo  $i$  e  $\bar{X}_i$  representa a média do grupo  $i$  em questão.

Segundo Lattin (2003), o método não hierárquico das  $k$ -médias consiste na divisão em  $k$  grupos, onde se obtém o centróide de cada agrupamento e é calculada a soma das distâncias ao quadrado de cada objeto do seu centróide do grupo, ou seja, o quadrado da soma dos erros de partição, representado por:

$$SQD = \sum_{i=1}^n (X_i - \bar{X}_{C(i)})^T (X_i - \bar{X}_{C(i)}),$$

em que  $C(i)$  é o agrupamento para cada objeto  $i$ . Como o SQD reflete as distâncias entre os objetos internos do grupo, esse é o termo que deve se tornar tão pequeno quanto possível (LATTIN, 2003).

Os métodos não hierárquicos diferem dos hierárquicos em vários aspectos, mas o mais importante se refere à necessidade de que o usuário tenha se especificado previamente o número de grupos  $k$  desejado, ao contrário das técnicas hierárquicas aglomerativas (MINGOTI, 2005).

### 3. Metodologia

Neste trabalho foram analisadas as produções agrícolas em todas as 853 cidades do estado de Minas Gerais em 2017. A partir das informações levantadas sobre a produção agrícola do estado, as culturas que tiveram maior desempenho no ano foram abacaxi, alho, batata inglesa, café arábica, feijão, laranja, milho e tomate, portanto essas foram as oito variáveis selecionadas para a análise. Algumas culturas, como trigo, soja e sorgo, também apresentam destaque no estado, porém, como apresentam alta correlação com algumas das oito selecionadas, não foram consideradas para o presente estudo. As oito culturas foram observadas com base no seu rendimento, representado pela produção (kg) por área plantada (ha), e seu valor de mercado (R\$).

Desse modo, as variáveis utilizadas são:

1. *Rendimento do abacaxi* : razão entre a quantidade produzida de abacaxi por área plantada;
2. *Rendimento do alho*: razão entre a quantidade produzida de alho por área plantada;
3. *Rendimento da batata inglesa*: razão entre a quantidade produzida de batata inglesa por área plantada;
4. *Rendimento do café*: razão entre a quantidade produzida de café por área plantada;
5. *Rendimento do feijão*: razão entre a quantidade produzida de feijão por área plantada;
6. *Rendimento da laranja*: razão entre a quantidade produzida de laranja por área plantada;
7. *Rendimento do milho*: razão entre a quantidade produzida de milho por área plantada;
8. *Rendimento do tomate*: razão entre a quantidade produzida de tomate por área plantada;
9. *Valor do abacaxi*: valor de mercado do abacaxi;
10. *Valor do alho*: valor de mercado do alho;
11. *Valor da batata inglesa*: valor de mercado da batata inglesa;

12. *Valor do café*: valor de mercado do café;
13. *Valor do feijão*: valor de mercado do feijão
14. *Valor da laranja*: valor de mercado da laranja;
15. *Valor do milho*: valor de mercado do milho;
16. *Valor do tomate*: valor de mercado do tomate.

Os dados foram coletados da pesquisa Produção Agrícola Municipal - PAM do IBGE de 2017 (IBGE, 2017). Para as análises foram adotados dois métodos de análise multivariada: análise de componentes principais (ACP) por meio da matriz de correlação e análise de agrupamento (AA). Para tais análises foi utilizada a linguagem de programação *Python* (PYTHON, 2019) e os pacotes *pandas* (para lidar com o conjunto de dados); *numpy* e *scipy* (para lidar com os cálculos vetoriais e matriciais); *matplotlib* e *seaborn* (para gerar os gráficos); *sklearn* (para a análise de componentes principais); *scipy.cluster.hierarchy*, *scipy.spatial.distance*, *scipy.cluster.hierarchy* e *sklearn.cluster* (especificamente para a análise de agrupamento).

Os dados lidos pelo pacote *pandas* tinham a seguinte estrutura:  $n = 853$  linhas (cada uma correspondente a um município de MG) e  $p = 17$  variáveis (as 16 variáveis já descritas e o código de cada município).

Primeiro, foi aplicado o método de análise de componentes principais. Segundo Lattin et al. (2011), a ACP permite que os dados sejam reorganizados de forma que as primeiras dimensões consigam explicar a maior parte das informações disponíveis possível. Seu objetivo é encontrar um meio de condensar a informação contida em várias variáveis originais em um conjunto menor de variáveis com uma perda mínima de informação (HAIR et al., 2006).

Por fim, também foi efetuada a análise de agrupamentos (AA) com base na matriz de correlações. Segundo Hair et al. (2006), a AA classifica as variáveis de modo que cada objeto é semelhante a outros dentro do grupo com determinada característica, e diferentes em relação a outros grupos. A análise de agrupamento consiste na divisão de uma grande grupo em grupos menores que sejam similares entre si. (LATTIN, 2003). Tal método foi aplicado de duas formas, pelo método hierárquico e o não-hierárquico,

em ambos os casos foram consideradas as distâncias euclidianas para a mensuração da proximidade dos grupos.

Nessa pesquisa foram empregados 5 métodos não hierárquicos: o vizinho mais próximo, vizinho mais distante, distância média, centróide e Ward. Porém, após realizada a análise com os métodos, foi observado que Ward trouxe informações mais compactadas, sem a presença de nenhum *outlier*, o que facilita a análise feita durante a pesquisa, portanto, foi o método escolhido para o estudo em questão.

Para a escolha do número de grupos, foi considerado o critério do corte no dendrograma, que consiste em localizar onde há a maior altura livre entre os pontos de agrupamentos. Dessa forma, foi definida a divisão dos municípios em três grupos. Logo depois, foram calculados os valores médios das variáveis analisadas em cada grupo, assim como suas medianas.

Em seguida, foi construído um gráfico onde os grupos obtidos pelo método de Ward são relacionados com os componentes principais já obtidos anteriormente. Em suma, a ideia é analisar como as variáveis se comportam quando são relacionados os grupos com os componentes principais.

Da mesma forma como foi analisado o método hierárquico de Ward, o método não-hierárquico aplicado foi o  $k$ -médias. Segundo Lattin (2003), esse método consiste na divisão em  $k$  grupos, onde se obtém o centróide de cada agrupamento e é calculada a soma das distâncias ao quadrado de cada objeto do seu centróide do grupo. Foi utilizado o método de Ward, já calculado, para definir o número de agrupamentos a ser utilizado no método  $k$ -médias. Após obtidos os grupos, assim como feito no método de Ward, foram obtidas as médias e medianas de cada grupo. Em seguida, foi obtido o gráfico que relaciona os agrupamentos com os componentes principais e os objetos observados em relação às variáveis de estudo.

Sendo assim, após obtidos os resultados das análises de componentes principais e análise de agrupamento, foi possível observar com maior clareza os dados para que fosse possível a comparação das produções agrícolas nas cidades de Minas Gerais, e analisada qual é a representatividade dos municípios em relação aos cultivos estudados.

## **4. Resultados e discussão**

### **4.1 Análise descritiva**

A Tabela 2 apresenta um resumo estatístico das culturas estudadas dentro do estado de Minas Gerais. Trata-se da análise do rendimento e do valor das culturas e nela se pode observar, em se tratando de rendimento, em 2017, o tomate apresentou a maior média dentre as culturas analisadas nos municípios do estado. Já em se tratando do valor das produções, nota-se que o café possui a maior média entre todas as outras produções.

A partir da observação da Tabela 2 nota-se que todas as culturas analisadas possuem pelo menos de 0% até 25% de cidades que não atuam em nenhuma produção. A partir de 50% dos municípios do estado atuam somente na produção de feijão, milho e café.

O café possuir maior média e desvio padrão pode ser explicado pelos quartis, que mostram que pelo menos metade do estado atua na produção de café. Porém, a cultura mais popular no estado é a de feijão, que por outro lado, não possui maior média nem desvio padrão. Pelo menos 75% dos municípios de Minas Gerais produzem somente café, feijão, laranja, milho e tomate; 50% produzem somente café, feijão e milho; e 25% produzem somente feijão e milho. Ou seja, a maior parte das cidades não produz grande parte das culturas analisadas. Como a análise é feita sobre todas as 853 cidades do estado, esse resultado é esperado, dado o fato de que não são todos os municípios que trabalham com agricultura, em especial das oito produções analisadas.

Tabela 2 – Resumo estatístico do rendimento e da produção das oito culturas estudadas,  
Minas Gerais, 2017

	abacaxi	alho	batata inglesa	café arábica	feijão	laranja	milho	tomate
Rendimento								
média	2.309,7	662,0	4.284,3	830,0	930,0	4.725,6	4.006,1	18.623,0
desvio	6.743,6	2.492,1	10.644,6	889,8	668,8	8.373,5	2.468,8	29.194,7
min	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
25%	0,0	0,0	0,0	0,0	438,0	0,0	1.905,0	0,0
50%	0,0	0,0	0,0	699,0	760,0	0,0	3.760,0	0,0
75%	0,0	0,0	0,0	1.482,0	1.250,0	7.000,0	6.000,0	47.667,0
max	40.000,0	18.700,0	47.300,0	4.767,0	3.455,0	49.146,0	11.846,0	130.000,0
Valor								
média	236,1	393,8	970,3	12455,6	1325,1	554,7	3735,7	649,6
desvio	2799,8	5817,1	8554,2	33364,5	7550,4	4802,1	12817,2	3868,6
min	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
25%	0,0	0,0	0,0	0,0	27,0	0,0	72,0	0,0
50%	0,0	0,0	0,0	33,0	128,0	0,0	311,0	0,0
75%	0,0	0,0	0,0	5751,0	461,0	12,0	1572,0	98,0
max	56400,0	144233,0	180702,0	459021,0	147075,0	74888,0	161220,0	73440,0

Fonte dos dados: SIDRA-IBGE(Tabela 5457). Elaboração própria.

Foram obtidos os valores dos coeficientes de correlação de Pearson entre todas as culturas analisadas tanto em rendimento quanto em valor (valores não apresentados). As culturas que apresentaram maior correlação positiva foram o feijão e o milho, em torno de 0,67, o que significa dizer que, segundo os dados, nas regiões de Minas onde se concentra maior produção de feijão também se concentram as de milho. Isso acontece porque é comum plantar os dois juntos na mesma lavoura, fato conhecido como plantio consorciado. Duas culturas que também apresentam correlação alta foram abacaxi e laranja, com valor da correlação em torno de 0,42.

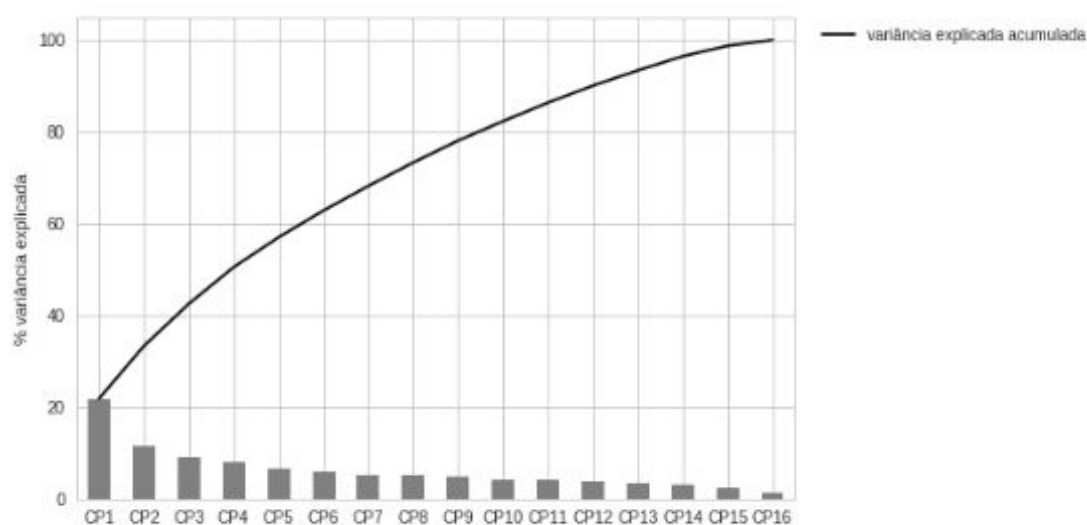
Apesar de muitas culturas possuírem correlação positiva baixa, as menores apresentadas foram abacaxi e café, em torno de 0,003. Isso significa que grande parte das culturas em estudo não possuem relação entre si no estado. Existem também as

culturas que apresentam correlação negativa, mas seus valores são muito baixos, variando de -0,005 a -0,025.

#### 4.2 Análise de componentes principais (ACP)

A variância de um componente principal revela o quanto ele explica da variância total das variáveis. No Gráfico da Figura 1 é apresentado o *scree plot*, cujo intuito é expressar essa explicação de cada componente. Percebe-se que o componente principal 1 (CP1) explica sozinho 21% da variância total. Para a análise, seria ideal utilizar oito componentes, que juntos explicam 73%, porém, oito componentes seriam difíceis de serem analisados e interpretados. Portanto, foram considerados os três primeiros CPs, que explicam juntos 42% da variância total.

Figura 1 - Variância explicada dos componentes principais (CPs)



Fonte dos dados: SIDRA-IBGE (Tabela 5457). Elaboração própria.

A Tabela 3 apresenta os coeficientes dos 3 primeiros CPs e é possível observar que o primeiro componente principal (CP1) possui coeficientes positivos para todas as variáveis, isso significa dizer que o CP1 é influenciado positivamente pelas oito culturas analisadas, tanto pelo rendimento quanto pelo valor. Porém, os maiores coeficientes positivos correspondem ao café, ao feijão e ao milho, o que significa dizer



que, quanto maior o escore do CP1 para um município, maior a importância dessas culturas.

Já o segundo componente principal (CP2) possui coeficientes positivos altos para as produções de abacaxi e laranja, ou seja, quanto maior o escore do CP2, maior a produção de frutas no município em questão. O segundo componente possui coeficientes mínimos ou negativos para todas as outras variáveis, o que indica que os municípios mais relacionados com o CP2, não apresentam escores de produções altas nas outras culturas estudadas, somente nas frutas.

O terceiro componente principal (CP3) é fortemente influenciado, de forma positiva, pela produção de alho, e negativamente, pela produção de tomate. Isso significa que, nos municípios onde o escore do tomate estiver alto, o do alho estará baixo, e vice versa. Ou seja, as cidades do estado com forte relação com o CP3 ou apresentará forte produção de alho e nenhuma de tomate, ou uma alta produção de tomate e nenhuma de alho. Também significa que, quando ambos estão próximos de zero, se anulam.

Tabela 3 – Componentes principais das variáveis selecionadas

	abacaxi	alho	batata inglesa	café arábica	feijão	laranja	milho	tomate
Rendimento								
CP1	0,07	0,22	0,32	0,26	0,36	0,19	0,22	0,21
CP2	0,47	-0,15	-0,17	-0,02	0,02	0,32	-0,01	-0,08
CP3	0,16	0,39	0,08	-0,3	0,02	-0,19	-0,11	-0,37
Valor								
CP1	0,06	0,16	0,24	0,27	0,27	0,11	0,41	0,17
CP2	0,53	-0,18	-0,12	-0,07	-0,01	0,54	0,02	-0,03
CP3	0,14	0,36	0,34	-0,26	0,11	0,04	0,14	-0,41

Fonte dos dados: SIDRA-IBGE(Tabela 5457). Elaboração própria.

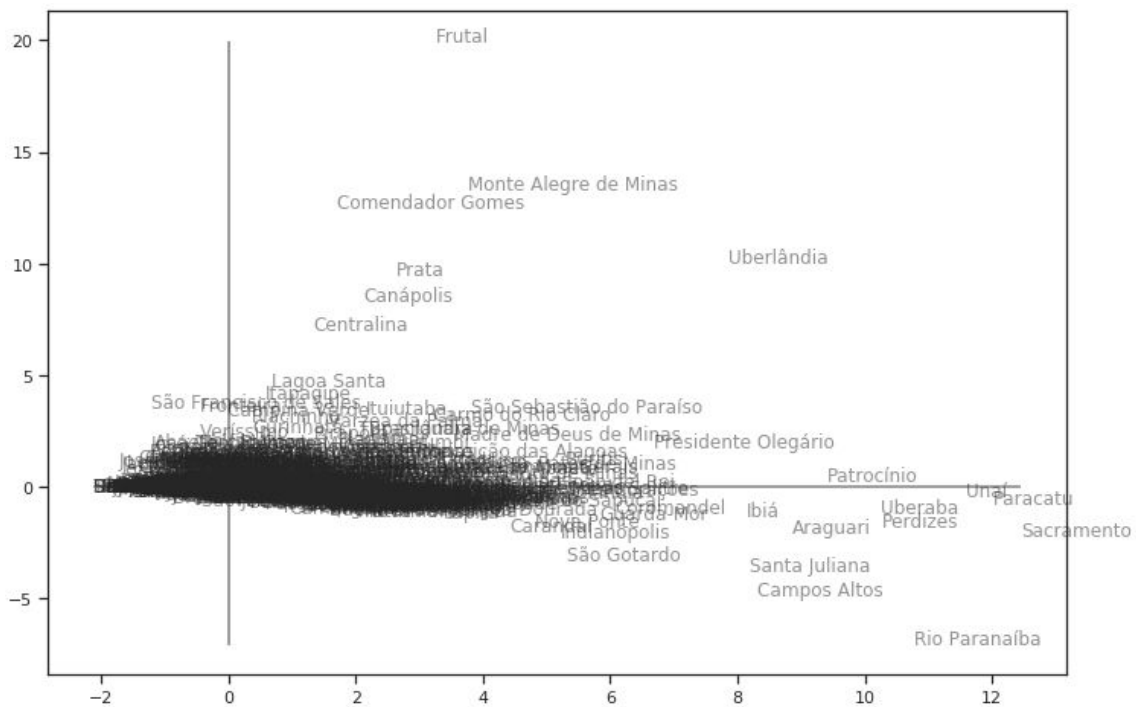
O gráfico da Figura 2 é um diagrama de dispersão dos dois primeiros CPs, mostrando a relação de todas as cidades de Minas Gerais com o CP1, a linha horizontal, e o CP2, linha vertical. Observa-se então, cidades como Patrocínio, Unaí, Presidente Olegário, dentre outras, que possuem uma relação muito alta com o CP1, ou seja, são

municípios que apresentam produções em todas as variáveis, porém, com maiores cultivos de feijão, café e milho.

Cidades como Frutal, Monte Alegre de Minas e Comendador Gomes, apresentam relação alta com o CP2 e baixa relação com o CP1. Isso significa que são municípios com maiores dados de produções de frutas, como abacaxi e laranja, e poucos cultivos das outras variáveis.

Contudo, a cidade de Uberlândia, por exemplo, apresenta tanto alta relação com o CP1, quanto com o CP2, ou seja, é um município que apresenta forte participação na cultura e feijão, café, milho, abacaxi e laranja.

Figura 2 - Diagrama de dispersão dos municípios de MG com base nos componentes principais



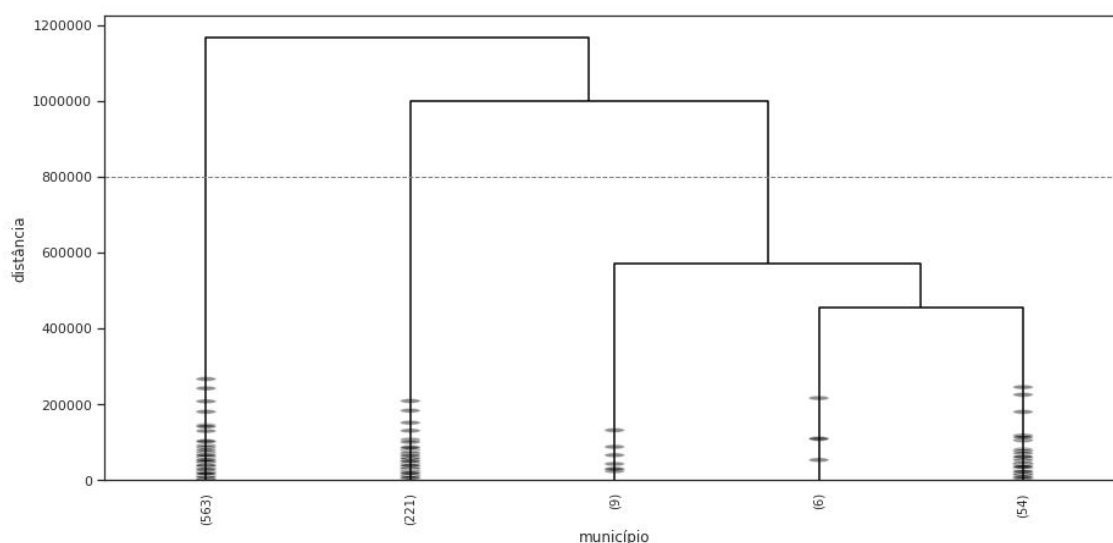
Fonte dos dados: SIDRA-IBGE (Tabela 5457). Elaboração própria.

### 4.3. Análise de agrupamento (AA)

#### 4.3.1. Método de Ward

A Figura 3 mostra o dendrograma resultante do agrupamento pelo método de Ward. Nele, o corte apresentado tende a separar os municípios de Minas Gerais em 3 grupos. O primeiro grupo contém 563 municípios, o segundo 221, e o terceiro 69.

Figura 3 - Dendrograma dos grupos pelo método de Ward



Fonte dos dados: SIDRA-IBGE (Tabela 5457). Elaboração própria.

A Tabela 4 mostra a média e a mediana das variáveis em cada um dos três grupos apresentados. O primeiro grupo (Grupo 0) apresenta maiores médias no rendimento do tomate e no valor do abacaxi. Ele apresenta mediana positiva no rendimento e valor do café, feijão, milho e tomate, vale ressaltar que a mediana da produção de tomate é a maior. Ou seja, os municípios que fazem parte do grupo 0, são os mais influenciados na produção de tomate. Além disso, 50% dos municípios desse grupo não produzem abacaxi, alho, batata inglesa e laranja.

Faz parte do segundo grupo (Grupo 1) as cidades com médias altas na produção de abacaxi, laranja e milho. Apesar de apresentarem médias altas para essas culturas, suas medianas são zero, o que significa que 50% dos municípios que fazem parte do grupo 1 não atuam na produção dessas culturas, mas por possuírem cidades que produzem muito, isso faz com que a média do grupo seja tão representativa para essas variáveis.

Esse grupo pode ser associado ao CP2, que apresentava maior relação com os municípios produtores das frutas abacaxi e laranja.

O último grupo (grupo 2), é composto pelas cidades com maior rendimento médio na produção de batata inglesa e tomate. Já no valor médio, as produções que se destacam são café, milho e batata inglesa. Apesar de se destacar na produção de batata, o grupo 2 apresenta metade de municípios não produtores desse cultivo. Esse grupo apresenta mediana igual a zero para as culturas de abacaxi, alho e batata inglesa. Ele contém o menor número de municípios, porém, é o grupo que apresenta maior diversidade produtiva das variáveis selecionadas do estado de Minas Gerais.

Tabela 4 – Médias e medianas dos grupos pelo método de Ward

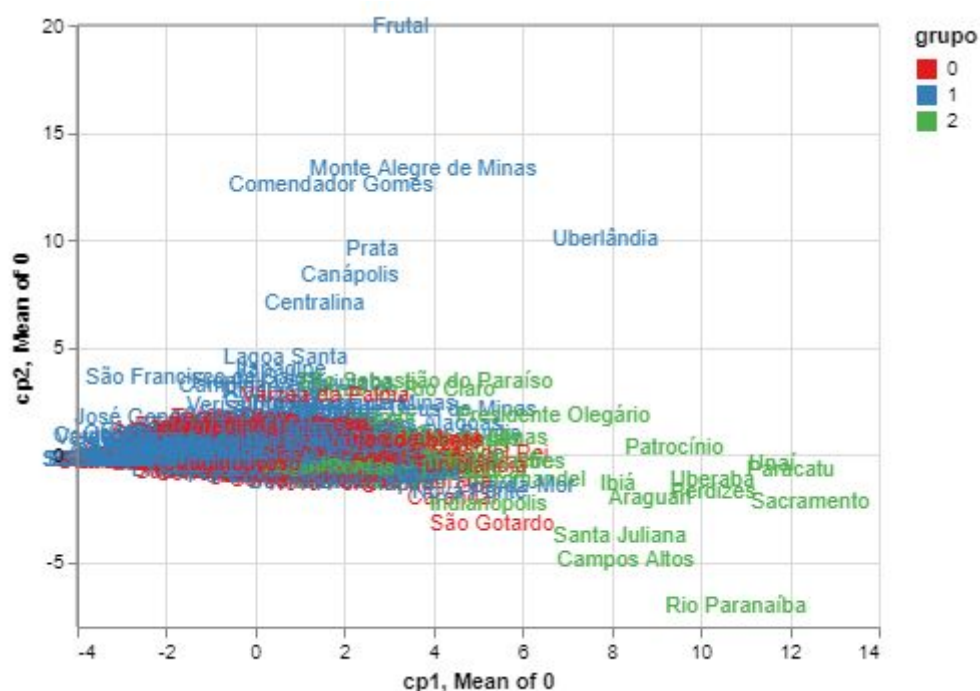
		abacaxi	alho	batata inglesa	café arábica	feijão	laranja	milho	tomate
Rendimento									
Grupo	Média	1.964,7	855,0	5.103,9	927,0	946,7	5.979,0	4.261,6	57.994,7
0	Mediana	0,0	0,0	0,0	933,0	800,0	0,0	4.000,0	60.000,0
Grupo	Média	2.457,3	434,7	2.811,6	687,3	871,7	3.775,5	3.638,8	181,2
1	Mediana	0,0	0,0	0,0	0,0	680,0	0,0	3.111,0	0,0
Grupo	Média	2.210,1	1.898,6	13.676,0	1.683,9	1.352,3	8.462,8	6.185,3	42.993,6
2	Mediana	0,0	0,0	0,0	1.680,0	1.289,0	5.000,0	6.519,0	51.000,0
Valor									
Grupo	Média	11.891,4	251,8	389,3	5.732,0	859,6	147,8	2.248,2	1.701,6
0	Mediana	0,0	0,0	0,0	168,0	189,0	0,0	449,0	346,0
Grupo	Média	346,0	45,7	283,3	4.488,8	849,5	687,8	2.164,9	5,1
1	Mediana	0,0	0,0	0,0	0,0	90,0	0,0	180,0	0,0
Grupo	Média	57,0	3.689,1	8.436,8	98.995,2	6.696,4	772,3	21.316,9	2.539,4
2	Mediana	0,0	0,0	0,0	81.972,0	613,0	5,0	5.544,0	249,0

Fonte dos dados: SIDRA-IBGE (Tabela 5457). Elaboração própria.

O gráfico da Figura 4 mostra a separação das cidades de Minas Gerais com base nos grupos obtidos pelo método de Ward. Nele, pode-se observar a forte relação do grupo 1 com o CP2. Ambos apresentaram maior similaridade com as variáveis abacaxi e laranja. As cidades que mais se destacam, então, na produção de frutas no estado são Frutal, Monte Alegre de Minas e Comendador Gomes.

Também é possível observar a forte relação entre o grupo 2 e o CP1. Esse grupo apresentou médias e medianas altas para boa parte das culturas analisadas, isso é notável no Gráfico 2 pelo fato de as cidades desse grupo estarem bem dispersas. Porém, por se relacionar tão fortemente com o CP1, se pode entender que nas produções de feijão, milho e café, esse grupo se destaca, apesar de não possuir as maiores médias. Isso acontece pelo fato de que, para outras variáveis, existem municípios nesse grupo com valores muito altos, que fazem com que a média suba, mas não significa necessariamente que essa é a variável mais representativa do grupo, e ao observar essa relação com o CP1, isso fica muito claro.

Figura 4 - Separação das cidades de MG em grupos pelo método de Ward relacionadas aos componentes principais



Fonte dos dados: SIDRA-IBGE (Tabela 5457). Elaboração própria.

#### 4.3.1. Método das $k$ -médias

Como definido pelo método de Ward, o número de grupos escolhido para o método das  $k$ -médias é  $k = 3$ . Dessa forma, a separação ficou: o grupo 0 contendo 589 municípios, o grupo 1 com 223; e o grupo 2 com 41.

A Tabela 5 mostra as médias e medianas de cada grupo obtido pelo método das  $k$ -médias. Nele se observa que, em se tratando tanto de rendimento quanto de valor, no grupo 0 se destaca a produção de milho, que possui médias e medianas altas. Apesar de o maior rendimento médio ser da produção de laranja, a mediana da variável é zero, ou seja, pelo menos 50% dos municípios presentes no grupo 0 não atua no cultivo de laranja, porém, por fazer parte dele algumas cidades que tem a produção muito forte, isso faz com que a média seja tão relevante.

Em relação ao grupo 1, nota-se que fazem parte dele os municípios onde a produção de tomate é forte, tanto em rendimento quanto em valor. Apesar de não ser a variável com maiores médias, a produção de tomate apresenta altos valores de medianas, o que indica que muitos municípios presentes nesse grupo atuam na produção de tomate. Desse grupo também fazem parte cidades que atuam na produção de milho, feijão, café e laranja, porém não de forma tão representativa.

No grupo 2 estão concentradas as cidades onde de fato a produção do tomate, tanto em valor quanto em rendimento, é mais intensa. Também fazem parte desse grupo municípios com altas produções de laranja e abacaxi. A maior mediana apresentada por esse grupo também foi na produção de tomate, e apesar de o valor médio da produção de abacaxi ser alta, sua mediana é zero.

Tabela 5 – Médias e medianas dos grupos pelo método das *k*-médias

		abacaxi	alho	batata inglesa	café arábica	feijão	laranja	milho	tomate
Rendimento									
Grupo 0	Média	2.430,3	469,9	2.808,8	692,1	868,4	3.771,3	3.606,1	935,5
	Mediana	0,0	0,0	0,0	333,0	680,0	0,0	3.033,0	0,0
Grupo 1	Média	3.109,8	1.585,4	13.343,4	1.738,4	1.410,8	10.415,1	6.507,7	40.721,0
	Mediana	0,0	0,0	0,0	1.735,0	1.374,0	6.000,0	6.805,0	50.000,0
Grupo 2	Média	1.843,9	999,8	6.515,9	1.027,2	1.004,5	6.200,0	4.602,6	61.277,2
	Mediana	0,0	0,0	0,0	1.164,0	900,0	0,0	4.400,0	60.000,0
Valor									
Grupo 0	Média	330,4	67,3	519,8	5.048,6	1.005,6	536,1	2.083,2	2,4
	Mediana	0,0	0,0	0,0	7,0	93,0	0,0	182,0	0,0
Grupo 1	Média	28,9	5.608,4	7.252,5	129.820,7	6.179,8	1.111,0	20.295,2	3.043,8
	Mediana	0,0	0,0	0,0	111.376,0	936,0	5,0	6.709,0	225,0
Grupo 2	Média	25,0	297,6	1.005,2	10.441,2	1.276,5	501,8	5.055,9	1.918,8
	Mediana	0,0	0,0	0,0	712,0	205,0	0,0	597,0	400,0

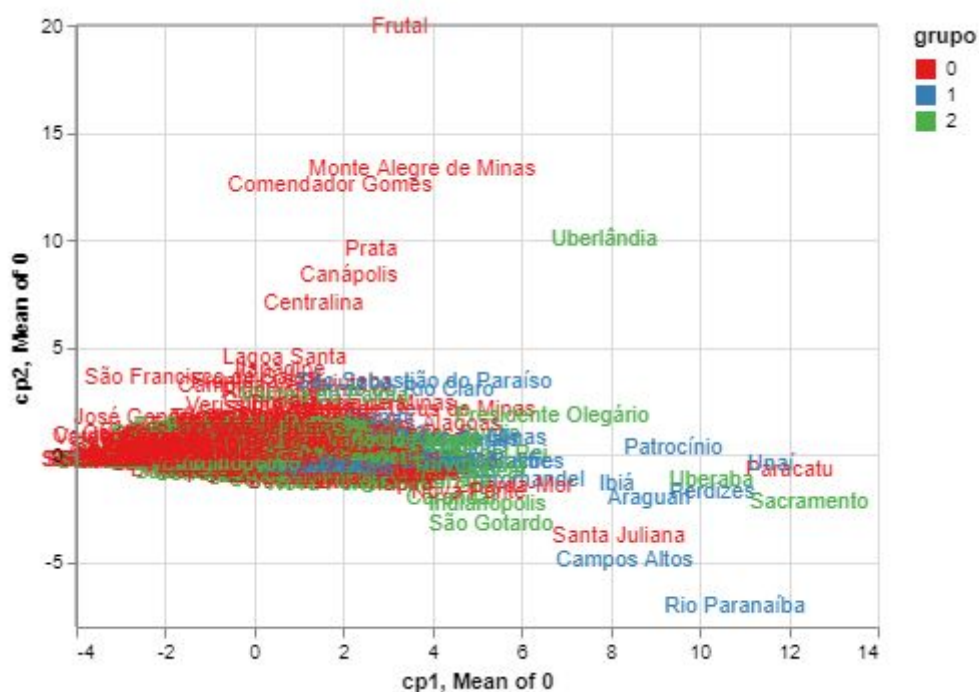
Fonte dos dados: SIDRA-IBGE (Tabela 5457). Elaboração própria.

O gráfico da Figura 5 mostra como as cidades de Minas se comportam na separação dos grupos quando relacionadas aos componentes principais. Se observa que o método das *k*-médias tendeu a misturar um pouco mais os municípios do que o observado no método de Ward. Apesar disso, é possível observar a relação do grupo 0 com o CP2, ambos apresentaram uma forte produção de laranja. Ou seja, a análise indica que os municípios como Frutal, Monte Alegre de Minas e Comendador Gomes, tiveram uma alta produção de laranja no ano observado.

Também é possível observar a relação dos grupos 1 e 2 com o CP1. Isso acontece porque ambos os grupos apresentaram, por menor que fosse, um valor relevante nas produções de milho, feijão e café. Como observado, os grupos 1 e 2, ficaram com informações parecidas, e no gráfico é possível enxergar essa similaridade,

pois, mesmo que apresentem uma relação mais forte com o CP1, ambos são muito relacionados com os dois componentes.

Figura 5 - Separação das cidades de MG em grupos pelo método das  $k$ -médias relacionadas aos componentes principais



Fonte dos dados: SIDRA-IBGE (Tabela 5457). Elaboração própria.

Entre os dois métodos foi possível observar algumas diferenças. O método de Ward tendeu a separar três grupos onde um apresentou maior produção de tomate, outro maior produção de abacaxi e laranja, e o último com destaque nas produções de milho, café e batata inglesa. Já o método das  $k$ -médias, tendeu a criar três grupos mais mistos, ou seja, grupos que possuem médias parecidas em suas variáveis mais similares. Por esse método, foram obtidos três grupos, onde o primeiro se destacavam as culturas de milho e laranja; no segundo a produção com maiores médias foi a de tomate, mas também apresenta valores significativos na produção de milho, feijão, café e laranja; e no terceiro as produções mais intensas são de tomate, laranja e abacaxi. Como se pode observar, a variável laranja apareceu em todos os 3 grupos do método das  $k$ -médias, assim como o milho apresentou valores relevantes em dois grupos e o tomate também.



Apesar das diferenças entre os métodos, eles também apresentaram a maior semelhança entre o grupo 1 (Ward) e o grupo 0 ( $k$ -médias), pois são neles onde se concentram as maiores produções de frutas, e fazem parte deles um grupo de cidades similares, como Frutal, Monte Alegre de Minas e Comendador Gomes. Também apresentou semelhança entre o grupo 2 (Ward) e o grupo 1 ( $k$ -médias), onde as médias eram relevantes nas produções de milho, feijão e café, e fazem parte deles cidades como Patrocínio, Unaí, e Rio Paranaíba. Isso dentre muitas outras cidades em comum entre os grupos.

## 5. Considerações finais

Para a elaboração deste trabalho foram utilizadas informações das produções agrícolas mais influentes em Minas Gerais em 2017. Com essas variáveis selecionadas foi possível fazer uma análise multivariada dos dados para que se pudesse enxergar com mais clareza os resultados. Foram utilizados dois tipos de análise: componentes principais e agrupamento - sendo o método hierárquico de Ward e o não hierárquico das  $k$ -médias.

A partir dessa análise, é possível fazer algumas observações. A primeira é que em Minas Gerais, mesmo que o rendimento do café arábica não seja tão alto quanto das outras culturas analisadas, em questão de valor ele se destaca muito. Como foi possível notar, o valor do café se destacou muito em todas as análises envolvidas no artigo. Isso acontece pelo fato de ele ter um valor de mercado muito alto, especialmente quando comparado aos outros cultivos estudados.

Os métodos de agrupamento utilizados na análise apresentaram algumas semelhanças, principalmente na separação das cidades onde existe maior influência na produção de frutas. As cidades que se destacaram foram Frutal, Monte Alegre de Minas, Prata, Canápolis, Centralina, Lagoa Santa, dentre outras.

Também houve destaque da cidade de Patrocínio, que aparece em ambas análises como um município forte na produção de milho, feijão e café, mais especialmente na de café, assim como Unaí, Perdizes, Ibiá, Rio Paranaíba, Campos Altos, dentre outros. Como apresentado anteriormente, Patrocínio é um dos municípios

de Minas Gerais onde a produção de café é muito intensa, seguida pelo cultivo de milho e feijão.

Como o ramo da agricultura é muito amplo e muito importante para a economia do Brasil, vale ressaltar a extrema necessidade de mais pesquisas e estudos futuros focados na abordagem desse tema e na análise de produtividade dos estados para o país.

## 6. Referências bibliográficas

BASTOS, S.; GOMES, J. E. Dinâmica da agricultura no Estado de Minas Gerais: análise diferencial-estrutural para o período 1994 a 2008. **RURIS - Revista do Centro de Estudos Rurais - UNICAMP**, v. 5, n. 2, 2011.

EVERITT, B. S. et al. **Cluster analysis**. 5. ed. Reino Unido: John Wiley & Sons, 2011.

FERREIRA J. R. S. F.; BAPTISTA, A. J. M. S.; LIMA, J. E. A modernização agropecuária nas regiões do Estado de Minas Gerais. **Revista de Economia e Sociologia Rural**, v. 42, n. 1, p. 73-89, 2004.

FREITAS, R. E.. Produtividade agrícola no Brasil. Produtividade no Brasil: desempenho e determinantes. **IPEA - Instituto de Pesquisa Econômica Aplicada** v. 1, 2014.

GASQUES; J. G. et al. **Desempenho do crescimento do agronegócio no Brasil**. Brasília: Texto para discussão do IPEA, 2004.

HAIR, J. F. et al. **Análise multivariada de dados**. 6. ed. Porto Alegre: Bookman, 2009.

**IPEADATA**. Disponível em: <http://www.ipeadata.gov.br>. Acesso em 10 de julho de 2019.

LATTIN, J.; CARROLL, J. D.; GREEN, P. E. **Análise de dados multivariados**. São Paulo: Cengage Learning Edições LTDA, 2003.

MINGOTI, S. A. **Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada**. Belo Horizonte: Editora UFMG, 2005.

PEREIRA, F.; SOARES, A.. Avaliação dos riscos envolvidos na produção de milho e soja nos municípios do Mato Grosso. **Revista iPecege**. v. 3, n. 2, p. 38-51, 2017.

PYTHON. **The Python programming language**. Disponível em:  
<https://docs.python.org/3/reference>. Acesso em 17 de junho de 2019.

SCOLARI D. D. **Produção agrícola mundial: o potencial do Brasil**. Área de Informação da Sede-Capítulo em livro científico (ALICE). 2006.

SECRETARIA DE ESTADO DE AGRICULTURA, PECUÁRIA E ABASTECIMENTO DE MINAS GERAIS. **Projeções do agronegócio em Minas Gerais: projeções de longo prazo - 3 ed.** Disponível em:  
[http://www.agricultura.mg.gov.br/images/Arq\\_Relatorios/Publicacoes/projecoes\\_2017\\_a\\_2027.pdf](http://www.agricultura.mg.gov.br/images/Arq_Relatorios/Publicacoes/projecoes_2017_a_2027.pdf). Acesso em 10 de julho de 2019.

SECRETARIA DO ESTADO DE AGRICULTURA, PECUÁRIA E ABASTECIMENTO DE MINAS GERAIS (SEAPA). **Relatórios da Agricultura**. 2019. Disponível em:  
<http://www.agricultura.mg.gov.br/index.php/2014-09-23-01-07-23/relatorios>. Acesso em: 20 de junho de 2019.

SEIDEL, E. J. et al. Comparação entre o método Ward e o método K-médias no agrupamento de produtores de leite. **Ciência e Natura - UFSM**, v.1, n. 30, p.7-15, 2008.

SILVA, G. J. C.; SOUZA, E. C.; MARTINS, H. E. P. Produção agropecuária em municípios de Minas Gerais (1996-2006): padrões de distribuição, especialização e associação espacial. **Revista de Economia e Sociologia Rural - RESR**, v. 50, n. 2, p. 333-350, 2012.

SISTEMA IBGE DE RECUPERAÇÃO AUTOMÁTICA (SIDRA). Produção Agrícola Municipal (PAM). Disponível em: <https://sidra.ibge.gov.br/pesquisa/pam/tabelas>. Acesso em 23 de janeiro de 2019.