

UNIVERSIDADE FEDERAL DE ALFENAS

RENAN SERENINI BERNARDES

**ANÁLISE ESPACIAL DA PRODUTIVIDADE DE CAFÉ
EM MINAS GERAIS (2002 - 2017)**

Varginha/MG

2019

RENAN SERENINI BERNARDES

**ANÁLISE ESPACIAL DA PRODUTIVIDADE DE CAFÉ
EM MINAS GERAIS (2002 - 2017)**

Dissertação apresentada como parte dos requisitos para a obtenção do título de Mestre em Estatística Aplicada e Biometria pela Universidade Federal de Alfenas. Área de Concentração: Estatística Aplicada e Biometria.
Orientadora: Profa. Dra. Patrícia de Siqueira Ramos.

Varginha/MG

2019

Dados Internacionais de Catalogação na Publicação (CIP)
Sistema de Bibliotecas da Universidade Federal de Alfenas
Biblioteca *campus* Varginha

Bernardes, Renan Serenini.
B522a Análise espacial da produtividade de café em Minas Gerais (2002-2017) /
Renan Serenini Bernardes. - Varginha, MG, 2019.
67 f. : il. -

Orientadora: Patrícia de Siqueira Ramos.
Dissertação (mestrado em Estatística Aplicada e Biometria) -
Universidade Federal de Alfenas, 2019.
Bibliografia.

1. Análise espacial (Estatística). 2. Produtividade - Café. 3. Estatística
Produtividade. I. Ramos, Patrícia de Siqueira. II. Título.

CDD – 519.537

Ficha Catalográfica elaborada por Natália Maria Leal Santos
Bibliotecária-Documentalista CRB6/2734



MINISTÉRIO DA EDUCAÇÃO
Universidade Federal de Alfenas / UNIFAL-MG
Programa de Pós-graduação em Estatística Aplicada e Biometria

Rua Gabriel Monteiro da Silva, 700. Alfenas - MG CEP 37130-001
Fone: (35) 3701-9600
<https://www.unifal-mg.edu.br/ppgeab/>



RENAN SERENINI BERNARDES

“ANÁLISE ESPACIAL DA PRODUTIVIDADE DE CAFÉ EM MINAS GERAIS (2002 -
2017) ”

A Banca Examinadora, abaixo assinada, aprova a
Dissertação apresentada como parte dos requisitos para
a obtenção do título de Mestre em Estatística Aplicada
e Biometria pela Universidade Federal de Alfenas.
Área de Concentração: Estatística Aplicada e
Biometria

Aprovado em: 25 de junho de 2019.

Profa. Dra. Patrícia de Siqueira Ramos
Instituição: UNIFAL-MG

Assinatura: Patrícia Ramos

Prof. Dr. Leandro Ferreira
Instituição: UNIFAL-MG

Assinatura: Leandro Ferreira

Prof. Dr. Denismar Alves Nogueira
Instituição: UNIFAL-MG

Assinatura: Denismar Nogueira

*Este trabalho é dedicado a quem mais se
esforçou para sua conclusão: minha mãe.*

AGRADECIMENTOS

Este trabalho jamais seria concluído sem a participação de muitas pessoas, as quais deixo meus agradecimentos.

Agradeço primeiramente à toda minha família e à minha noiva, pelo apoio, compreensão e suporte constante durante toda a jornada.

À UNIFAL-MG, pelas oportunidades concedidas durante esses quase dez anos de graduação e pós-graduação.

Aos meus orientadores, prof^a Patrícia de Siqueira Ramos e prof. Lincoln Frias pela enorme dedicação que tiveram comigo.

Aos professores do mestrado e aos colegas de turma com quem convivi ao longo de 4 anos.

Aos professores e colegas da graduação em Economia, que, além da formação, despertaram meu interesse pela pesquisa.

Ao IBGE, pela oportunidade de conciliar os estudos com o trabalho, em especial aos colegas da agência de Varginha, que sempre me incentivaram e me deram suporte para tal.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

RESUMO

O Brasil é o maior produtor mundial de café. Nos últimos 20 anos, foi responsável por aproximadamente um terço da produção mundial. No Brasil, o estado de maior destaque é Minas Gerais, que, segundo estimativas da CONAB - Companhia Nacional de Abastecimento, em 2018 produziu mais de 53% de todo o café produzido no país. Considerando a relevância do café e do estado de Minas Gerais no contexto, este trabalho tem como objetivo analisar a distribuição espacial da produtividade (medida pela produção de café sobre a área colhida) do café em MG no período de 2002 a 2017 e verificar a evolução da produção e da produtividade ao longo do período. Especificamente, foi analisada a distribuição espacial através da divisão do estado em regiões geográficas imediatas, nova divisão regional publicada pelo Instituto Brasileiro de Geografia e Estatística (IBGE), em 2017. Para atingir o objetivo foram analisadas as séries de produção, produtividade e área colhida, verificando a existência de dependência espacial entre as regiões e *clusters* de alta ou baixa produtividade ao longo dos anos analisados. A dependência espacial foi estimada através da estatística *I* de Moran e a indicação de *clusters* ou *outliers* espaciais foi verificada pelo *I* de Moran local. Os resultados indicaram que a produtividade do café em Minas Gerais como um todo vem aumentando nos últimos anos. Algumas regiões que possuíam pequenas áreas de café deixaram de produzir e outras regiões, que possuíam grandes áreas, mas com níveis de produtividade baixos, aumentaram a produtividade, indicando um movimento de concentração da produção. Foi identificado que a dependência espacial da produtividade se mostrou mais expressiva no primeiro período (2002 a 2009) do que no segundo (2010 a 2017). Alguns *clusters* de alta produtividade foram identificados, mostrando-se mais expressivos no primeiro período. Os resultados do trabalho podem servir como subsídio para novas pesquisas, implementando variáveis associadas à produtividade do café, bem como apoio à tomada de decisões para políticas do setor.

Palavras-chave: Estatística Espacial. Dados espaciais. Autocorrelação espacial. *I* de Moran.

Produtividade do café.

ABSTRACT

Brazil is the world's largest coffee producer. For the last 20 years, has been responsible by approximately one third of world production. In Brazil, the most prominent state is Minas Gerais, which, according to estimates by CONAB - National Supply Company, In 2018 it produced over 53% of all coffee produced in the country. Considering the relevance of coffee and the state of Minas Gerais in the context, this paper aims to analyze the spatial distribution of productivity (measured by the ratio of coffee production over the harvested area) of the coffee in MG from 2002 to 2017 and to verify the evolution of production and productivity over the period. Specifically, the spatial distribution was analyzed through the division of the state in immediate geographic regions, the new regional division published by The Brazilian Institute of Geography and Statistics (IBGE), in 2017. To achieve the objective, the series of production, productivity and harvested area were analyzed, verifying the existence of dependence between the regions and clusters of high or low productivity over the years. Spatial dependence was estimated using Moran's I statistics and cluster indication or space outliers was verified by the local Moran's I. The results indicated that the coffee productivity in Minas Gerais as a whole has been increasing in recent years. Some regions that had small areas of coffee stopped producing and other regions, which had large areas, but with low productivity levels, increased its productivity, indicating a concentration movement of production. It was identified that the spatial dependence of productivity was more significant in the first period (2002 to 2009) than in second (2010 to 2017). Some high productivity clusters have been identified, showing most significant in the first period. The results of the work can serve as subsidy for new research, implementing variables associated with coffee productivity, as well as support for policy making in the sector.

Keywords: Spatial Statistics. Spatial data. Spatial autocorrelation. Moran's I. Coffee productivity.

SUMÁRIO

1	INTRODUÇÃO	9
2	REFERENCIAL TEÓRICO	11
2.1	ANÁLISE ESPACIAL	11
2.1.1	Dados espaciais	12
2.1.1.1	Tipos de dados espaciais	12
2.1.1.2	Problemas com dados espaciais	15
2.1.2	Vizinhança e matrizes de ponderação	16
2.1.2.1	Matriz de Contiguidade espacial	17
2.1.2.2	Matriz de Distância geográfica	19
2.1.2.3	Matriz de pesos espaciais gerais de Cliff e Ord	19
2.2	ANÁLISE EXPLORATÓRIA DE DADOS ESPACIAIS EM ÁREA	20
2.2.1	Contagem de junções	21
2.2.2	<i>C</i> de Geary	22
2.2.3	Índice de Moran global	23
2.2.3.1	Diagrama de dispersão de Moran	24
2.2.3.2	<i>I</i> de Moran local	25
2.3	ANÁLISE ESPACIAL DA PRODUTIVIDADE DO CAFÉ - REVISÃO DE LITERATURA	28
3	MATERIAL E MÉTODOS	32
3.1	DADOS	32
3.2	ANÁLISE	33
4	RESULTADOS E DISCUSSÃO	37
4.1	ESTATÍSTICAS DESCRITIVAS	37
4.2	MAPAS TEMÁTICOS	46
4.3	<i>I</i> DE MORAN	52
4.3.1	Diagramas de dispersão de Moran	54
4.4	MAPAS LISA (<i>I</i> DE MORAN LOCAL)	56
5	CONSIDERAÇÕES FINAIS	60
	REFERÊNCIAS	61
	APÊNDICE A	64

1 INTRODUÇÃO

O agronegócio (insumos para a agropecuária, produção agropecuária básica, agroindústria e agrosserviços) possui grande destaque na economia brasileira. Segundo cálculos do Centro de Estudos Avançados em Economia Aplicada (CEPEA), da Esalq/USP, em 2017 o PIB do Agronegócio representou 21,6% do PIB do Brasil, sendo que o ramo agrícola representou 15%, enquanto o ramo pecuário 6,6% (CEPEA, 2018).

Na agricultura, um dos principais produtos é o café. Além de seu papel histórico no desenvolvimento econômico do Brasil, o café possui grande representatividade para o país. De acordo com dados da Organização Internacional do Café (ICO, sigla em inglês - *International Coffee Association*), o Brasil é o maior produtor mundial de café. Nos últimos 20 anos, o Brasil foi responsável, em média, por aproximadamente um terço da produção de café mundial. Em 2017, a produção brasileira foi de aproximadamente 51 milhões de sacas, representando 35% da produção de todo o mundo (ICO, 2018).

No Brasil, a maior parte da produção de café se encontra no estado de Minas Gerais. Segundo relatório da Companhia Nacional de Abastecimento (ACOMPANHAMENTO DA SAFRA BRASILEIRA, 2018), o estado produziu, no último ano, aproximadamente 53% do café do Brasil. Para dimensionamento, esse valor representa quase 20% da produção mundial de café, deixando evidente a importância do estado para este produto.

Portanto, ressaltada a importância e representatividade do café para a economia brasileira, mostram-se relevantes estudos acerca do produto, buscando compreender aspectos que afetem a produtividade e que possam contribuir para a otimização da produção. Dentre os estudos, destaca-se a importância da análise espacial pela possibilidade de identificar fatores locais da produtividade e a distribuição espacial da mesma. Conforme ressaltam Teixeira e Bertella (2015) e Souza e Perobelli (2007), a agricultura é muito suscetível às questões espaciais, já que o desenvolvimento das culturas agrícolas é heterogêneo, por depender de fatores como solo, temperatura e condições climáticas, daí a importância de se investigar a distribuição espacial da produtividade. Nesse sentido, o presente trabalho tem por objetivo avaliar a distribuição espacial da produtividade do café em Minas Gerais entre 2002 e 2017. Busca-se verificar se, no estado, a produtividade do café se distribui aleatoriamente ou se existem padrões de distribuição espacial e se existem regiões de alta ou baixa produtividade. Além disso, objetiva-se verificar também a evolução da produtividade do café no período espacialmente, identificando regiões

que aumentaram ou diminuíram os índices de produtividade.

O trabalho foi estruturado da seguinte forma: a primeira seção apresenta uma revisão sobre análise exploratória de dados espaciais, com breve contexto histórico, principais conceitos e exemplos de aplicação; a segunda seção apresenta e descreve os dados a serem utilizados; a seção seguinte apresenta os resultados e discussões e a última seção traz as considerações finais.

2 REFERENCIAL TEÓRICO

2.1 ANÁLISE ESPACIAL

Os termos “Estatística espacial” e “Análise espacial” muitas vezes são usados como sinônimos na literatura para descrever o conjunto de técnicas utilizadas para estudar entidades a partir de suas propriedades geométricas, topológicas ou geográficas. Para Silva (2010), estatística espacial é o conjunto de técnicas e métodos estatísticos que podem ser utilizados para se analisar dados espaciais. Assunção (2001) explica que estatística espacial é um ramo da Estatística que estuda métodos científicos para a coleta, descrição, visualização e análise de dados que possam ser modelados como processos estocásticos, onde o espaço índice é um conjunto de dimensão maior que um. Haining (2003) explica que o termo análise espacial tem origem na Geografia, na década de 50 e é comumente utilizado na literatura de Sistemas de Informações Geográficas (SIG). Para ele, a definição de análise espacial é uma “coleção de técnicas e modelos que explicitamente usam a referência espacial associada a cada dado ou objeto especificado no sistema em estudo” .

A análise espacial possui três elementos principais. O primeiro inclui a modelagem cartográfica, onde um conjunto de dados representa um mapa e operações lógicas e algébricas sobre o mapa geram novos mapas. O segundo elemento seria a modelagem matemática, onde os modelos dependem da interação entre os objetos no espaço. Por último, a análise espacial inclui o desenvolvimento e a aplicação de técnicas estatísticas para a análise apropriada de dados espaciais e uso de referências espaciais nos dados (HAINING, 2003). Este último elemento é o aspecto que será abordado neste trabalho.

Metodologicamente, a análise espacial se diferencia pela incorporação dos efeitos espaciais sobre os dados e a utilização de dados espaciais, os quais serão abordados na próxima seção.

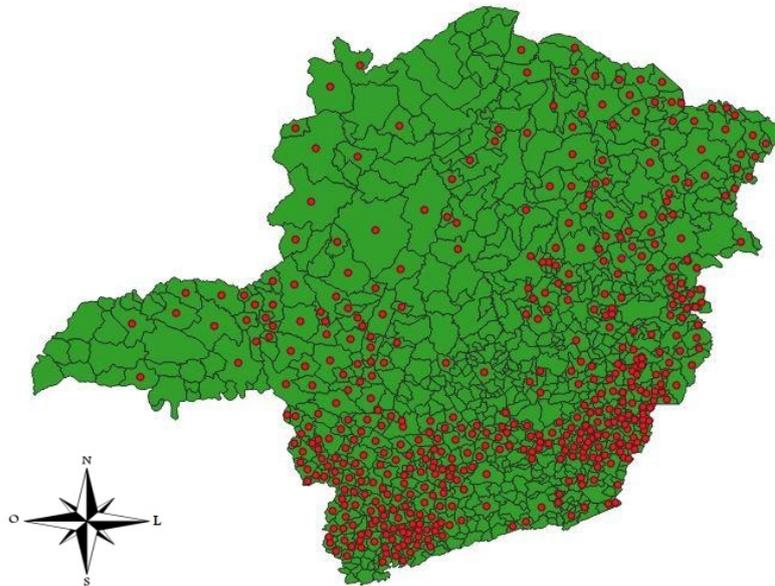
2.1.1 Dados espaciais

Ao se realizar análise espacial, não apenas a metodologia é diferenciada, utilizando os efeitos espaciais, mas também os dados são diferentes, são utilizados os dados espaciais. Em primeiro lugar, a diferença entre dados espaciais e dados não espaciais seria que, segundo Fotheringham, Brundson e Charlton (2000) os dados espaciais compreendem as observações de alguns fenômenos que possuem referência espacial. Ou seja, os dados espaciais denotam quanto varia algum fenômeno, preocupando-se em saber onde ocorre tal variação. Já os dados não espaciais preocupam-se apenas com a variação, sem dar atenção à localização. Assim, as propriedades dos dados espaciais seriam a magnitude da variação do fenômeno observado e a localização geográfica do mesmo. Os dados espaciais podem ser encontrados em diferentes formas, as quais serão apresentadas adiante.

2.1.1.1 Tipos de dados espaciais

Câmara *et al.* (2004) e Cressie (1993) explicam que, geralmente, os problemas de análise espacial consideram três tipos de dados: dados na forma de pontos ou eventos, dados de superfícies contínuas ou geoestatísticos e dados agrupados em áreas com contagens e taxas agregadas. Dados agrupados em pontos representam casos em que cada ponto é uma localidade de um evento discreto, como a existência ou não de determinado tipo de serviço público em determinado município, a ocorrência de doenças ou a localização de espécies vegetais. Nesses casos, o objeto de interesse é a localização espacial do evento. Como exemplo, a Figura 1 mostra o mapa de Minas Gerais com pontos nos municípios que produziram café no ano de 2016.

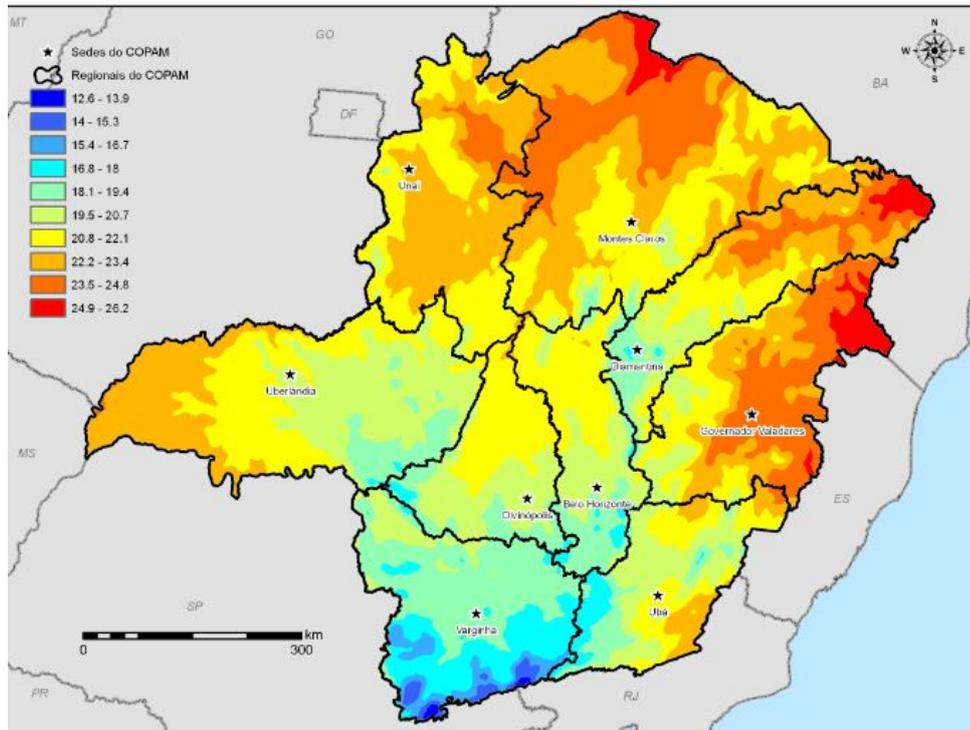
Figura 1 – Municípios de Minas Gerais que produziram café em 2016



Fonte: Elaboração própria

Os dados geoestatísticos, ou dados de superfície, são de uso mais frequente em estudos de ciências físicas e naturais e são dados em que a variação é contínua e pode ser avaliada em qualquer ponto do mapa, como exemplo, a medida de temperatura em uma superfície. Nesses casos, segundo Câmara *et al.* (2004), o objetivo é reconstruir a área da qual se retiraram e mediram as amostras. Como exemplo, a Figura 2 mostra a temperatura média em Minas Gerais.

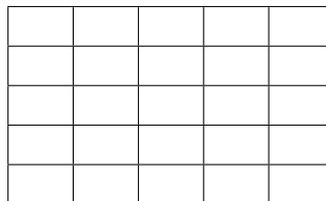
Figura 2 – Temperatura média em MG



Fonte: Scolforo *et al.* (2008)

Por último, os dados agrupados em área, que são o foco deste trabalho, são dados espaciais descontínuos agrupados em polígonos regulares ou irregulares (na maioria dos casos são irregulares) em forma e tamanho. Um exemplo frequentemente demonstrado para exemplificar os dados agrupados em áreas regulares é o gride apresentado na Figura 3.

Figura 3 – Gride 5 por 5

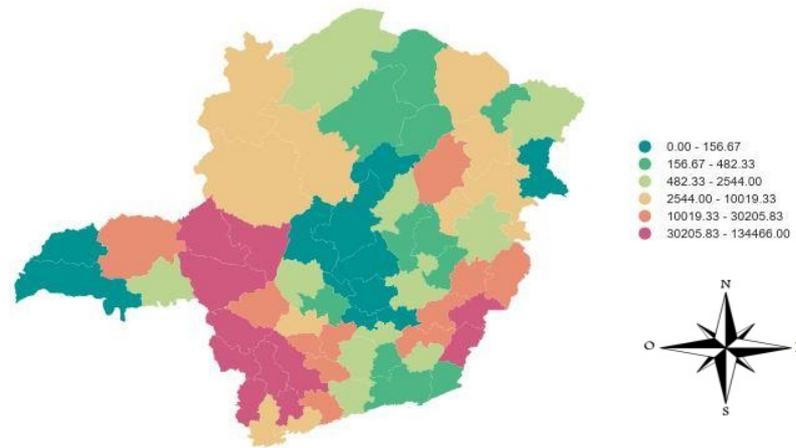


Fonte: Elaboração própria

Na prática, na maioria das vezes, os dados estão agrupados em polígonos não regulares, como municípios, bairros, setores censitários. Assim, os valores estão agregados nos polígonos e variam apenas quando se passa de um polígono para outro. Nessa situação, conforme Câmara *et al.* (2004), a variação do fenômeno é considerada constante dentro de cada região. Como exemplo, a Figura 4 mostra a área total, em hectares, de café plantada em cada microrregião de

Minas Gerais em 2016.

Figura 4 – Produção de café por microrregião de MG em 2016



Fonte: Elaboração própria

2.1.1.2 Problemas com dados espaciais

Os principais problemas encontrados na literatura ao se trabalhar com dados espaciais são: a falácia ecológica, relacionada à inferência ecológica; o problema da unidade modificável, denominado MAUP (sigla em inglês, *modifiable areal unit problem*); o efeito beirada e a influência de *outliers* espaciais. Esses termos são explicados a seguir.

A falácia ecológica se refere ao erro de realizar inferência e deduzir conclusões para comportamentos individuais baseando-se em dados agregados que mostram o comportamento coletivo. Em geral, apenas dados individuais devem ser utilizados para conclusões acerca de comportamento individual. O fato de que muitas vezes não há disponibilidade de dados espaciais individuais é uma das fontes desse problema. Um exemplo clássico, que talvez tenha sido o primeiro a identificar a questão, foi demonstrado por Robinson (1950) com dados dos estados americanos do censo de 1930. Foi medida a correlação entre a porcentagem da população que era estrangeira e a porcentagem que era alfabetizada e o resultado foi de uma correlação positiva. Por outro lado, com dados individuais a correlação foi negativa. Ou seja, o fato de que estados (dados em nível agregado) que tinham mais estrangeiros tinham maior taxa de alfabetização não significava que os estrangeiros eram mais alfabetizados.

O problema da unidade de área modificável é frequentemente abordado em estudos de análise espacial. Segundo Haining (2003 *apud* ÁVILA; MONASTÉRIO, 2008), a natureza do MAUP é que uma mesma estatística calculada sobre duas unidades espaciais geralmente gera diferenças de acordo com o recorte utilizado, mesmo que os dados originais sejam os mesmos. Ou seja, o resultado da análise dos dados irá depender da forma como os dados foram agregados. Este problema foi particularmente abordado no trabalho de Ávila e Monasterio (2008). Os autores realizaram um estudo no Rio Grande do Sul, a partir da renda per capita, utilizando diferentes agregações, municípios e COREDEs (Conselhos Regionais de Desenvolvimento), e mostraram que houve interferência do MAUP. Isso mostrou que a situação de um município poderia ser mascarada dependendo da agregação utilizada. A recomendação dos autores foi de que, sempre que possível, o pesquisador utilize mais de uma forma de agregação dos dados e compare os resultados.

O problema denominado efeito de beirada se refere ao fato de que, muitas vezes, a região na área observada não capta na totalidade a dependência espacial do fenômeno sob investigação. Uma região pode sofrer dependência espacial de outras regiões que não estão no espaço estudado. Sobre este problema, Anselin (1988) ressalta que, “se o valor de uma variável mensurada para uma unidade espacial depende dos valores nas outras unidades espaciais do conjunto de dados, pode depender também de valores de unidades espaciais que não estão inclusas” (ANSELIN, 1988, p. 172). Almeida (2012) aponta que existem algumas tentativas de se lidar com o problema em questão, como “dobrar o mapa”, estender a área de estudo para as observações além da fronteira, ou estimar o modelo por mínimos quadrados generalizados, dando um peso menor às observações de fronteira.

Outro problema espacial são os *outliers* espaciais. Para Almeida (2012), os *outliers* espaciais são observações que possuem uma dependência espacial diferente das observações vizinhas. O autor explica ainda que a origem de *outliers* pode estar ligada ao processo de obtenção e armazenamento dos dados, ocasionando erros de medida.

2.1.2 Vizinhaça e matrizes de ponderação

A autocorrelação espacial implica que o valor observado de determinada variável em determinada região está associado ao valor dessa mesma variável em regiões vizinhas. A ques-

tão inicial é determinar o conceito de vizinhança no espaço. A partir de então será possível estimar um ou mais parâmetros que forneçam o grau de interação espacial. Para isso, utiliza-se a chamada matriz de ponderação espacial (ou matriz de pesos espaciais ou ainda matriz de proximidade espacial) comumente denotada por W . Na matriz, cada conexão entre duas regiões é representada por uma célula. Cabe ressaltar que, conforme Tyszler (2006) a proximidade ou vizinhança não necessariamente irá seguir um critério geográfico, podendo adotar-se critérios socioeconômicos, culturais ou outros. O autor explica que a vizinhança pode ser baseada em similaridade ou dissimilaridade. Alguns exemplos seriam matrizes de vizinhança baseadas no Índice de Desenvolvimento Humano (IDH), correntes de comércio (importações e exportações), ou fluxos migratórios. A seguir, serão apresentadas as matrizes de pesos espaciais mais utilizadas.

2.1.2.1 Matriz de Contiguidade espacial

A matriz de pesos espaciais construída a partir do critério de contiguidade irá considerar que duas regiões são vizinhas se partilharem uma fronteira física comum. Anselin (1988) explica que a medida original para dependência espacial ou autocorrelação espacial, apresentada nos primeiros trabalhos relativos ao tema foram baseadas na noção de contiguidade binária entre regiões. Ou seja, se duas regiões (i e j) são vizinhas (partilham de uma fronteira comum) é atribuído o valor 1, caso contrário, 0. Formalmente, para a construção da matriz W :

$$w_{ij} = \begin{cases} 1, & \text{se } i \text{ e } j \text{ são contíguos} \\ 0, & \text{se } i \text{ e } j \text{ não são contíguos} \end{cases} \quad (2.1)$$

Por convenção $w_{ii} = 0$, ou seja, uma região não é vizinha de si mesma. Obviamente, a contiguidade parte da existência de um mapa, onde as fronteiras possam ser verificadas. Em áreas não regulares, as fronteiras são facilmente identificáveis, contudo, tratando-se de áreas regulares com o gride apresentado na Figura 3, o conceito de contiguidade não é único. Utilizando o gride apresentado, serão exemplificadas as três principais convenções (os nomes das convenções são alusões aos movimentos das peças no jogo de xadrez) de contiguidade, segundo Anselin (1988). Nas Figuras 5, 6 e 7 os vizinhos da região A são as células que contêm um X.

Figura 5 – Convenção rainha

X	X	X
X	A	X
X	X	X

Fonte: Elaboração própria

Figura 6 – Convenção torre

	X	
X	A	X
	X	

Fonte: Elaboração própria

Figura 7 – Convenção bispo

X		X
	A	
X		X

Fonte: Elaboração própria

Assim, a convenção de contiguidade denominada bispo considera como vizinhas apenas as regiões que possuem um vértice em comum. A convenção torre considera como vizinhas as regiões que partilham de uma fronteira em comum, mas não considera as regiões que possuem apenas vértice em comum. Já a convenção rainha, assim como no jogo de xadrez, é a junção dos critérios adotados nas convenções bispo e torre e considera como vizinhas regiões que partilhem de uma fronteira ou um vértice em comum.

Em alguns casos, a matriz de vizinhança baseada em contiguidade pode subestimar a dependência espacial entre duas regiões, principalmente se possuem uma distância geográfica pequena. Para esse e vários outros casos pode ser adotado o critério de vizinhança baseado na distância geográfica, que será explicado a seguir.

2.1.2.2 Matriz de Distância geográfica

A matriz de vizinhança pode considerar o critério de distância geográfica. A ideia é que duas regiões sejam consideradas vizinhas se estiverem em uma distância menor que o limite pré-estabelecido. Em geral, Almeida (2012) explica que esse critério costuma ser utilizado calculando uma distância de modo que todas as regiões tenham um mesmo número de vizinhos. Formalmente:

$$w_{ij}(k) = \begin{cases} 1, & \text{se } d_i < d_j(k) \\ 0, & \text{se } d_i > d_j(k) \end{cases} \quad (2.2)$$

em que $d_i(k)$ é a distância de corte específica para que a região i possua k vizinhos. Esse critério garante que a matriz esteja equilibrada, eliminando a possibilidade de uma região sem vizinhos.

2.1.2.3 Matriz de pesos espaciais gerais de Cliff e Ord

Segundo Almeida (2012), a matriz de pesos espaciais de Cliff e Ord considera os dois critérios apresentados anteriormente, a fronteira e a distância geográfica. A matriz é construída a partir de pesos espaciais que consistem no comprimento relativo da fronteira comum entre as regiões ajustado pela distância inversa entre elas. Formalmente, os pesos podem ser expressos como:

$$w_{ij} = \frac{f_{ij}^{\phi}}{d_{ij}^{\delta}}, \quad (2.3)$$

em que f_{ij}^{ϕ} é a proporção da fronteira entre as regiões i e j no perímetro de i e ϕ e δ são parâmetros a serem definidos.

Selecionado o critério de vizinhança, tem-se a matriz de proximidade espacial de primeira ordem $\mathbf{W}^{(1)}$. A partir de então pode-se generalizar a ideia para vizinhos de maior ordem, ou seja, vizinhos dos vizinhos, encontrando-se a matriz de proximidade $\mathbf{W}^{(k)}$ em que k representa a ordem desejada. Câmara *et al.* (2004) sugerem que, para facilitar muitos cálculos de índices de autocorrelação espacial, é útil padronizar as linhas a matriz de proximidade espacial, dividindo cada elemento pela soma total dos elementos da linha a que pertence. O processo gera

uma nova matriz, denominada matriz de proximidade espacial padronizada, em que a soma de cada linha da nova matriz é igual a um.

Por fim, cabe ressaltar que não há consenso na literatura, ou testes formais, para se definir a matriz mais adequada para cada situação ou variável de estudo. É aconselhável que o pesquisador utilize critérios para a escolha da matriz e faça testes com outras matrizes utilizando os mesmos dados. Após a apresentação dos principais conceitos, serão aprofundadas, a seguir, as principais técnicas de análise de dados espaciais.

2.2 ANÁLISE EXPLORATÓRIA DE DADOS ESPACIAIS EM ÁREA

Segundo Anselin (1999) análise exploratória de dados espaciais (AEDE) é o conjunto de técnicas utilizadas para descrever e analisar dados espaciais. A primeira etapa num estudo de AEDE é verificar se os dados são distribuídos aleatoriamente no espaço. Assim, a análise espacial parte da hipótese nula de aleatoriedade espacial dos dados. Aleatoriedade espacial significa afirmar que o valor de uma variável em determinado local não depende do valor dessa variável nos outros locais. Em termos estatísticos, têm-se as hipóteses:

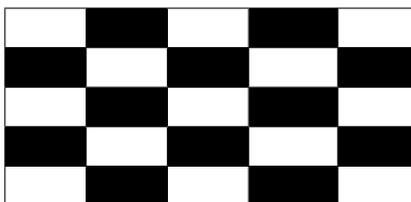
H_0 : presença de aleatoriedade espacial

H_1 : ausência de aleatoriedade espacial

Ao se rejeitar H_0 há evidências de uma estrutura espacial. Rejeitando-se a hipótese, há duas possibilidades: autocorrelação espacial positiva, em que valores parecidos próximos ocorrem mais frequentemente do que sob H_0 ; autocorrelação espacial negativa, em que valores diferentes próximos ocorrem mais frequentemente do que sob H_0 .

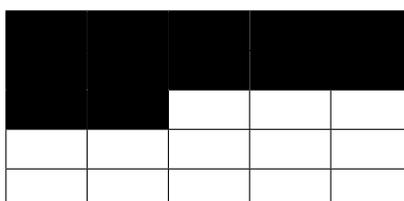
Os três exemplos nas Figuras 8, 9 e 10 mostram, a partir de dados dispostos em um gride, respectivamente, situações de: autocorrelação espacial negativa, autocorrelação espacial positiva e aleatoriedade espacial.

Figura 8 – Autocorrelação espacial negativa



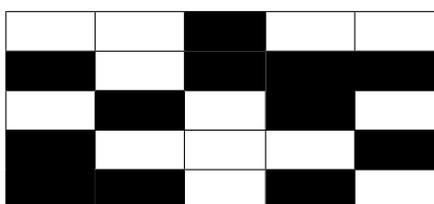
Fonte: Elaboração própria

Figura 9 – Autocorrelação espacial positiva



Fonte: Elaboração própria

Figura 10 – Aleatoriedade espacial



Fonte: Elaboração própria

Os exemplos apresentados são apenas ilustrativos e a definição de existência de autocorrelação espacial deve seguir testes estatísticos. A seguir, serão apresentados alguns dos principais testes de autocorrelação espacial utilizadas.

2.2.1 Contagem de junções

Fischer e Gettis (2009) explicam que a contagem de junções é uma estatística que testa a presença de autocorrelação espacial para variáveis binárias. Atribui-se o valor 1 à presença de determinada característica do fenômeno em estudo na região e 0 à ausência. A região com a característica presente é denominada Negra (N) e a ausente Branca (B). Assim, é feita a

contagem de junções entre as regiões, ou seja, fronteira entre duas regiões negras (NN), entre duas brancas (BB) e entre negra e branca (NB). Como exemplo, na Figura 10 apresentada anteriormente, os valores seriam: NN= 5, BB=12 e NB=23. A interpretação é de que, se houver uma concentração de junções haverá um número alto de NN e BB e um número baixo de NB. Na Figura 9, classificada como um exemplo de autocorrelação espacial positiva os valores seriam: NN=13, BB=22 e NB=5. Formalmente, em um conjunto com n células, existindo n_N células negras, o número de células brancas é $n_B = n - n_N$. Assim, a probabilidade de uma certa região ser negra é:

$$p_N = \frac{n_N}{n} \quad (2.4)$$

e a probabilidade de ser branca é:

$$p_B = \frac{n_B}{n}. \quad (2.5)$$

Portanto, existindo L junções em um mapa, o número esperado de junções de cada tipo é:

$$E(NN) = p_N^2 L \quad (2.6a)$$

$$E(BB) = p_B^2 L \quad (2.6b)$$

$$E(NB) = 2p_B p_N L \quad (2.6c)$$

Em resumo, existindo um número de junções NN significativamente maior que o esperado, considera-se que há evidências para rejeitar a hipótese nula de aleatoriedade espacial.

2.2.2 C de Geary

A estatística C de Geary parte da hipótese de aleatoriedade espacial, utilizando a matriz de pesos W , descrita na seção 2.1.2:

$$C = \frac{n-1}{2} \frac{\sum_i \sum_j w_{ij} (y_i - y_j)^2}{\sum_i (y_i - \bar{y})^2}, \quad (2.7)$$

em que n é o número de observações, y_i é o valor observado na região i e w_{ij} é um elemento da matriz de proximidade espacial. O numerador é a medida de covariância, utilizando o quadrado da diferença das observações. Assim, a interpretação é que se os dados forem concentrados, as regiões contíguas i e j apresentarão um valor de quadrado da diferença baixo, ao passo que se estiverem dispersas o quadrado da diferença será elevado (BAILEY; GATRELL, 1995). O valor de C de Geary é entre 0 e 2, e seu valor esperado é 1. Assim, valores abaixo de 1 indicam autocorrelação espacial positiva, enquanto valores acima de 1 indicam autocorrelação espacial negativa. A significância da estatística é obtida através do pressuposto de permutação aleatória, mecanismo também utilizado na estatística I de Moran, que será detalhado adiante.

2.2.3 Índice de Moran global

Possivelmente, a estatística mais utilizada para se testar a autocorrelação espacial é o I de Moran. Diferentemente da estatística C de Geary, a medida de autocovariância utilizada para o cálculo do coeficiente de autocorrelação é o produto cruzado. Assim, a estatística é descrita como:

$$I = \frac{n}{S_0} \frac{\sum_i \sum_j w_{ij} z_i z_j}{\sum_i z_i^2}, \quad (2.8)$$

em que i e j representam regiões distintas, $z_i = y_i - \bar{y}$ representa o desvio da variável de interesse em relação à média, n é o número de observações, w_{ij} representa um elemento da matriz de pesos espaciais e S_0 é a soma de todos os elementos da matriz de pesos espaciais, ou seja, $S_0 = \sum_i \sum_j w_{ij}$. Adotando-se a matriz de pesos espaciais padronizada, o valor de n será igual ao de S_0 , conseqüentemente a estatística será:

$$I = \frac{\sum_i \sum_j w_{ij} z_i z_j}{\sum_i z_i^2}. \quad (2.9)$$

A estatística I de Moran pode ser interpretada como um coeficiente de autocorrelação (apesar de não ser centrada em zero), assim, à medida que os valores de I excedem o valor esperado, há a indicação de autocorrelação positiva, e valores abaixo do esperado indicam autocorrelação negativa. Conforme Bailey e Gatrell (1995), o valor esperado da estatística I de

Moran é $[-1/(n - 1)]$. Este é o valor que deveria ser obtido na ausência de qualquer padrão espacial nos dados. Conforme pode ser observado, o valor tende a zero à medida que se aumenta o tamanho da amostra.

Em relação à inferência, são duas as formas utilizadas para se verificar a significância estatística do índice. A primeira forma é o pressuposto de normalidade, ou seja, assume-se que a variável padronizada $Z(I)$ tem uma distribuição amostral que segue uma distribuição normal com média zero e variância unitária. Nesse caso, a fórmula para $Z(I)$ é:

$$Z(I) = \frac{I - E(I)}{DP(I)}, \quad (2.10)$$

em que $E(I)$ e $DP(I)$ são o valor esperado e o desvio padrão teórico de I , respectivamente.

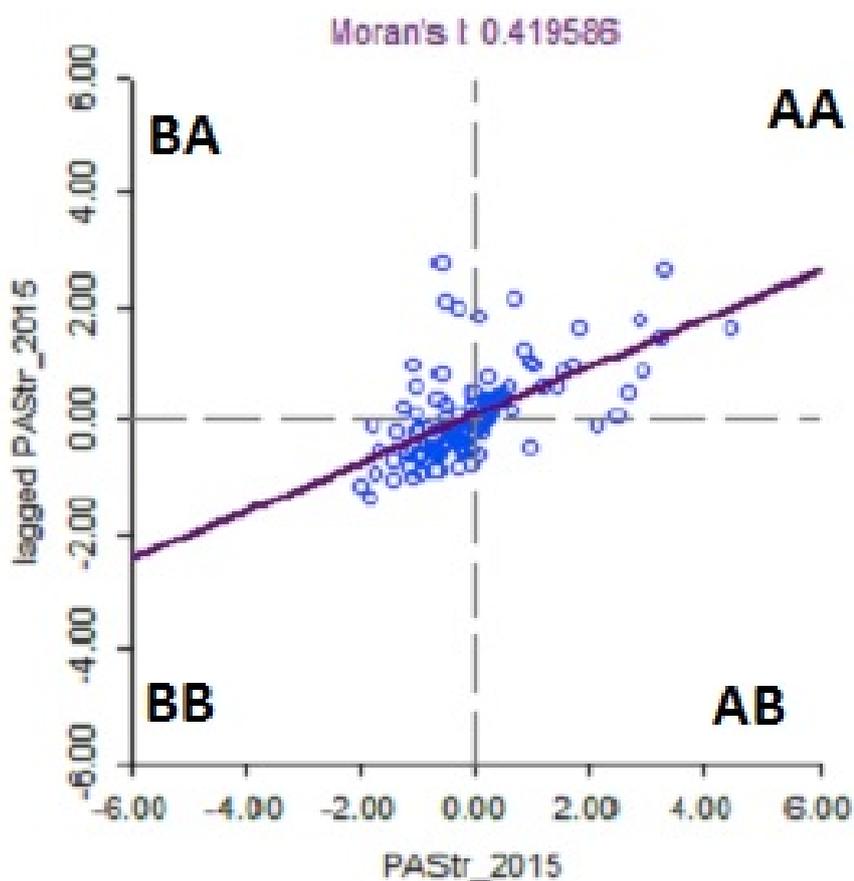
A segunda forma, que é a mais encontrada na literatura, é a permutação aleatória. Assume-se que o padrão observado dos dados é um dos vários possíveis. Assim, todos os valores são permutados aleatoriamente (embaralhados) para todas as regiões e obtém-se uma estatística para cada permutação. Desse modo, é calculada uma distribuição de referência empírica para a região de estudo, a partir da qual é possível definir se o padrão dos dados está dentro ou fora da região crítica definida pelo pesquisador (ANSELIN, 1995).

2.2.3.1 Diagrama de dispersão de Moran

O diagrama de dispersão de Moran é o gráfico de dispersão da nuvem de pontos representando as regiões de estudo, assim, pode ser entendido como uma ferramenta de interpretação gráfica do I de Moran. O diagrama é construído considerando o valor do atributo na região e a média dos seus vizinhos. Assim, o diagrama é dividido em quatro quadrantes. No primeiro quadrante, denominado Alto-Alto (AA, ou, em inglês, como frequentemente encontrado na literatura, *High-High*), estão as regiões que exibem valor da variável de interesse acima da média, com regiões vizinhas que também exibem valores acima da média. O segundo quadrante, Baixo-Alto (BA, ou, em inglês, *Low-High*), engloba as regiões com valores da variável de referência abaixo da média, rodeadas por regiões que possuem valores acima da média. O terceiro quadrante, chamado de Baixo-Baixo (BB, em inglês, *Low-Low*), contempla regiões com valores abaixo da média, cujas regiões vizinhas também possuem valores abaixo da média. Por fim, o quarto quadrante, Alto-Baixo (AB, ou, em inglês, *High-Low*), se refere às regiões com valor

alto, rodeados por vizinhos de valores baixos. A Figura 11 mostra um exemplo do diagrama de dispersão de Moran.

Figura 11 – Diagrama de dispersão de Moran para o percentual de área de café segurada em 2015 na mesorregião Sul e Sudoeste de MG



Fonte: Garcia Filho (2018)

Além do I de Moran e do diagrama de dispersão, outra técnica utilizada para se verificar a presença de autocorrelação espacial é o I de Moran local, adotado neste trabalho.

2.2.3.2 I de Moran local

Em complemento ao I de Moran, Anselin (1995) propôs um indicador com o objetivo de capturar padrões locais de autocorrelação¹. O autor denominou o indicador como *Local Indi-*

¹Existem também outros índices locais, como o C de Geary local, os quais não serão aprofundados aqui, pelo fato de se ter adotado o I de Moran como estatística para a autocorrelação espacial

ator os Spatial Autocorrelation (LISA). O indicador seria qualquer estatística que atendesse às seguintes condições: o valor do indicador para cada observação deve indicar *clusters* espaciais, estatisticamente significativos, ao redor da observação e a soma dos valores do indicador local deve ser proporcional ao indicador global utilizado.

Assim, análogo ao I de Moran, foi proposto pelo autor o I de Moran para o local i :

$$I_i = z_i \sum_j w_{ij} z_j, \quad (2.11)$$

em que os termos w_{ij} e z são os mesmos da equação do indicador global, ou seja, um elemento da matriz de pesos espaciais e o desvio da variável de interesse em relação à média, respectivamente.

O indicador de autocorrelação local permite verificar a existência de *clusters*, bem como a identificação de *outliers* espaciais locais. Localmente, um *outlier* seria uma unidade espacial com um valor alto de determinada variável, em uma região com várias unidades espaciais de valor baixo, ou vice-versa. Dessa forma, o I de Moran Local pode resolver o problema de padrões locais mascarados pelo I de Moran Global. A partir do cálculo do I de Moran Local é possível gerar um mapa LISA (*Local Indicator of Spatial Autocorrelation*), indicando as áreas onde o valor de I_i for significativo. Na Figura 12 é possível visualizar um mapa onde, para cada critério do diagrama de dispersão (Alto-Alto, Alto-Baixo, Baixo-Baixo e Baixo-Alto), é atribuída uma cor, utilizada apenas nos municípios onde o índice foi significativo.

Figura 12 – Porcentagem de área de café segurada em 2015 no Sul e Sudoeste de MG



Fonte:Garcia Filho(2018)

O mapa da Figura 12 mostra a mesorregião do Sul e Sudoeste de MG. Para cada um dos 146 municípios foi calculado o valor do I de Moran Local. Os municípios coloridos de cinza (Não significativo) são municípios cujo valor da variável (porcentagem de área de café segurada em 2015) não é diferente, em termos estatísticos, da média de todas as regiões. Apresentadas as técnicas de análise exploratória de dados espaciais, para exemplificação, cabe mencionar alguns trabalhos que utilizaram as ferramentas em estudos sobre o café. Os municípios coloridos de vermelho são municípios cujo valor da variável é acima da média de todas as regiões e cujos vizinhos também possuem valor acima da média, logo estão no grupo Alto-Alto. E assim sucessivamente. A rigor, quando várias regiões de um mesmo grupo estão juntas, considera-se a existência de um *cluster*. No mapa em questão existem dois *clusters* do tipo Alto-Alto, um com 9 municípios, na área Norte da mesorregião, e outro com 4 municípios na área Noroeste. Além destes, existem também quatro *clusters* do tipo Baixo-Baixo, dois contendo dois municípios cada, um com cinco municípios e outro com seis municípios, na área Leste da mesorregião. A identificação de um *cluster* pode ser útil para caracterizar uma grande região em relação à variável de estudo.

Apresentadas as técnicas de análise exploratória de dados espaciais, para exemplificação, cabe mencionar alguns trabalhos que utilizaram tais ferramentas em estudos sobre o café.

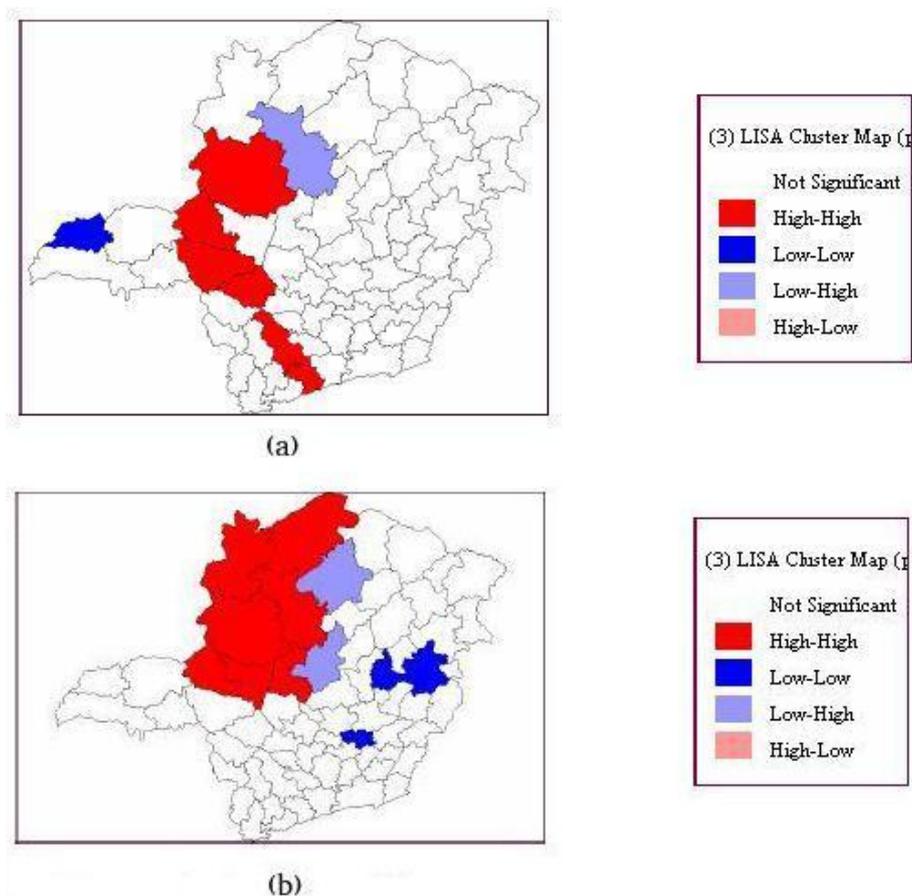
2.3 ANÁLISE ESPACIAL DA PRODUTIVIDADE DO CAFÉ - REVISÃO DE LITERATURA

Dada a relevância do tema, são vários os estudos que utilizam a análise espacial para variáveis relacionadas ao café no Brasil. Neves e Luiz (2006) utilizaram os dados da pesquisa Produção Agrícola Municipal (PAM), do Instituto Brasileiro de Geografia e Estatística (IBGE), para analisar a distribuição espacial da cultura de café no estado de São Paulo. O objetivo dos autores era subsidiar um plano de estratificação dos municípios para fins de previsão de safra. Os métodos utilizados foram o índice de Moran global, o índice local e o gráfico de dispersão de Moran. A variável utilizada foi o volume de produção por município no ano de 2004.

Santos e Nantes (2009) analisaram a distribuição espacial da cafeicultura no estado do Espírito Santo, também a partir de dados da PAM, com a variável volume de produção, para o ano de 2007. Dos 78 municípios capixabas, 76 produziram café e a análise foi feita de modo a verificar a existência de dependência espacial no volume de produção de cada município. Foi utilizado o índice I de Moran global, que apresentou valor de 0,59. Para efeitos de comparação, os autores calcularam o I de Moran para outros produtos, adotando a variável volume de produção no estado, no mesmo ano. Os valores foram: 0,31 para banana, 0,19 para o maracujá e 0,03 para a uva. Ao avaliar o indicador I de Moran local, os autores identificaram agrupamentos produtivos de destaque do tipo Alto-Alto e do tipo Baixo-Baixo ao longo do estado.

Almeida *et al.* (2006) analisaram a distribuição espacial da produtividade média do café em Minas Gerais para os anos de 2000 e 2004. Foram utilizados dados da PAM, considerando a produtividade média de cada microrregião de MG para os dois anos. Foram avaliados o I de Moran global e local, considerando uma matriz de pesos espaciais binária na convenção rainha, adotando os vizinhos de primeira ordem. Os autores concluíram que, para os dois anos, existia autocorrelação espacial positiva para os dados de produtividade média de café nas microrregiões. Em complemento, os autores testaram também a autocorrelação espacial bivariada, cruzando os dados de produtividade com outras duas variáveis: área plantada e quantidade produzida. Eles concluíram que não havia autocorrelação espacial entre a produtividade e as duas outras variáveis. Os agrupamentos encontrados podem ser visualizados na Figura 13.

Figura 13 – Mapas LISA para a produtividade média do café nas microrregiões de MG em 2000 e 2004



Fonte: Almeida et al(2006)

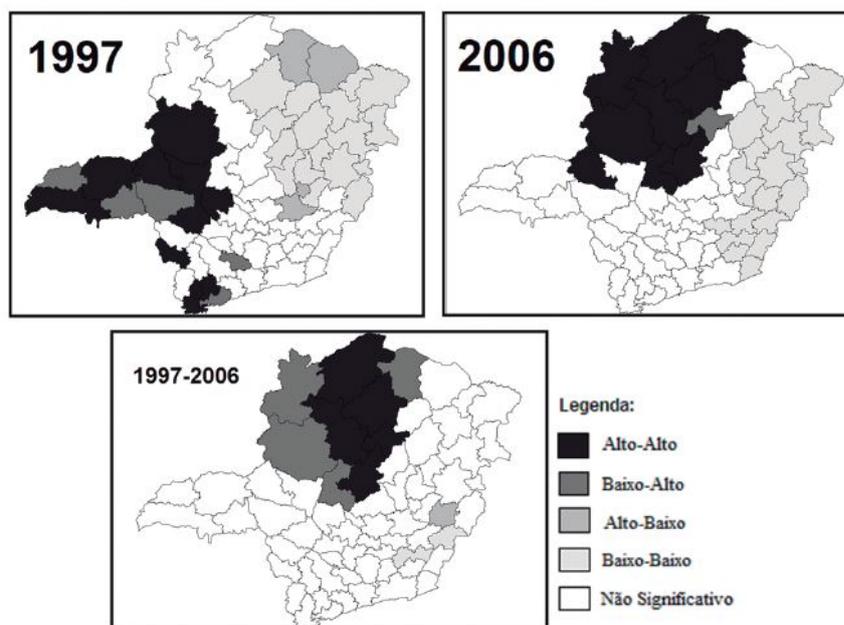
Legenda: (a) Produtividade média do café em 2000
(b) Produtividade média do café em 2004

Pelo mapa de *clusters* é possível visualizar a existência de autocorrelação espacial positiva localmente em ambos os anos, dado que é predominante a classificação de regiões do tipo Alto-Alto entre as que apresentaram valores com maior significância estatística. Ressalta-se que, no ano de 2004, o número de regiões do tipo Alto-Alto é maior do que em 2000. Segundo os autores, este resultado comprova a hipótese de que as regiões de maior produtividade podem estar influenciando o surgimento de regiões de igual desempenho, dada à influência da proximidade espacial.

Da mesma forma, Teixeira e Bertella (2015) usaram a produtividade média das microrregiões de MG para verificar a distribuição espacial nos anos de 1997 e 2006, bem como a distribuição espacial do aumento da produtividade nesse intervalo de tempo, buscando verificar a existência de convergência da produtividade média. A conclusão também foi pela existência

de autocorrelação espacial positiva em ambos os anos e na diferença de produtividade entre os anos. Além disso, os autores propuseram um modelo econométrico para investigar de maneira mais aprofundada a convergência espacial. Os mapas LISA encontrados são apresentados na Figura 14.

Figura 14 – Mapas LISA para a produtividade média do café nas microrregiões de MG em 1997 e 2006



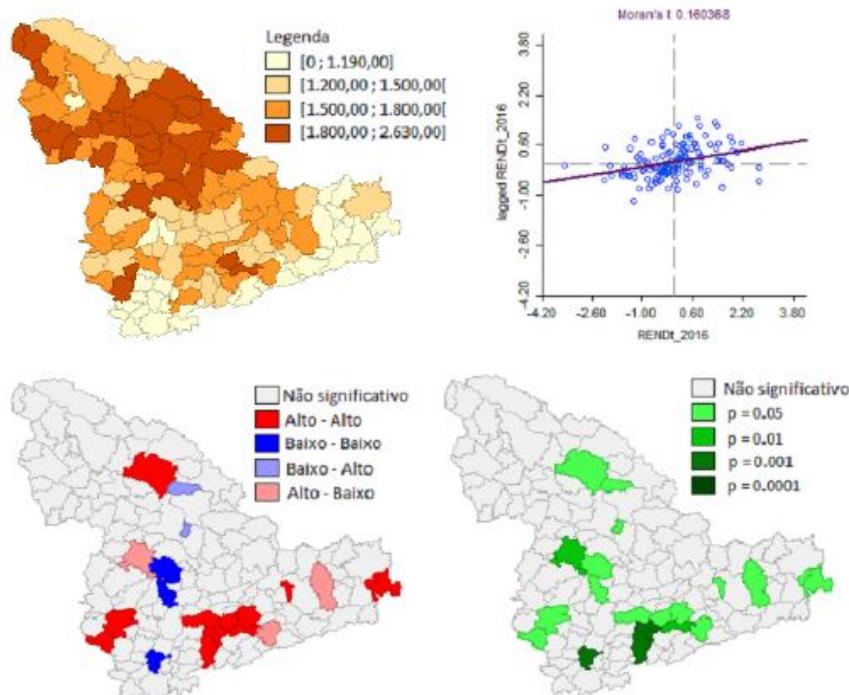
Fonte: Teixeira e Bertella (2015)

Os autores analisaram três diferentes mapas LISA: referente ao ano de 1997, a 2006 e referente à diferença da produtividade entre 1997 e 2006. O mapa de 1997 revelou os 4 tipos de padrões (Alto-Alto, Alto-Baixo, Baixo-Alto e Baixo-Baixo) espalhados pelo estado. Já em 2006 o mapa apresenta dois grandes agrupamentos, um do tipo Alto-Alto, na região Noroeste de Mg e outro do tipo Baixo-Baixo abrangendo as regiões Nordeste e Sudeste do estado. Por último, o mapa da diferença 1997-2006 revela, segundo os autores, a tendência espaço-temporal da taxa de crescimento da produtividade média do café. O comportamento da produtividade média ao longo indica uma mudança na dinâmica espacial da estrutura produtiva do café, sugerindo a ocorrência de um processo de convergência da produtividade entre as microrregiões, indicando que a produtividade estaria se tornando mais homogênea em todo o estado.

Garcia Filho (2018), também utilizando dados da PAM, estudou a distribuição espacial da produtividade média de café por município, considerando a mesorregião do Sul/Sudoeste de MG de 2012 a 2016. O objetivo era verificar a existência de autocorrelação espacial e

correlacionar as variáveis de percentual de área de café segurada no município e valor segurado por hectare. O autor identificou a existência de autocorrelação espacial para a produtividade de café nos anos de 2012, 2014 e 2016. A Figura 15 apresenta os resultados encontrados para o ano de 2016.

Figura 15 – Mapas LISA para a produtividade média do café nos municípios do Sul/Sudoeste de MG em 2016



Fonte: Garcia Filho (2018)

A Figura 15 mostra a distribuição da produtividade do café em 2016, o diagrama de dispersão de Moran, o mapa de *clusters* e o mapa com as respectivas significâncias dos municípios que se mostraram estatisticamente significantes no mapa LISA. No ano de 2016 foram identificados agrupamentos de municípios do grupo Alto-Alto e do grupo Baixo-Baixo, fato que também se repetiu em 2014 e 2012.

Baseando-se nestes e em outros trabalhos estudados, foram definidos os materiais e métodos para o trabalho em questão, os quais serão apresentados adiante.

3 MATERIAL E MÉTODOS

3.1 DADOS

Os dados utilizados neste trabalho são provenientes da pesquisa Produção Agrícola Municipal (PAM) do IBGE (IBGE, 2018). Foram utilizados os dados a partir do ano de 2002, quando houve alteração da unidade de medida do café para café beneficiado ou em grão. Até 2001 a pesquisa considerava o café em coco. Assim, os dados utilizados são de 2002 a 2017. Eles foram obtidos a partir do Sistema IBGE de Recuperação Automática (SIDRA), por meio da Tabela 1613 (área destinada à colheita, área colhida, quantidade produzida, rendimento médio -produtividade- e valor da produção das lavouras permanentes).

A área delimitada para o estudo é o estado de Minas Gerais. Minas Gerais é o primeiro estado do Brasil em número de municípios (853 municípios), o segundo em população (população estimada de 1^a de julho de 2018 de 21.040.662 pessoas) e o quarto em área territorial (586.521,121 km²). Minas Gerais possui também o terceiro maior PIB entre os estados da federação, estando atrás do Rio de Janeiro e São Paulo. Em 2018 o rendimento nominal mensal domiciliar per capita no estado era de R\$ 1322,00 (IBGE,2019).

Para o estado de MG estão disponíveis no banco de dados do IBGE informações de 852 dos 853 municípios, não constando o município de Santa Cruz de Minas, conhecido por ser o menor município do Brasil em área territorial. Conforme notas da Tabela 1613, os municípios sem informação para pelo menos um produto não aparecem. Foi realizado contato com o IBGE para confirmar a informação de inexistência de produção de café no município no período de estudo. Confirmada a informação, o município foi incluído com o valor 0 para a produção em todos os anos. Foi utilizada para análise da produtividade a variável denominada rendimento médio, razão entre o volume total da produção (em kg) e a área total colhida (em hectares) no ano. Assim, neste trabalho, ao se referir à produtividade da região (Região Geográfica Imediata) trata-se da soma da produção total da região no ano dividida pela área total de café colhida na região no ano.

3.2 ANÁLISE

Os dados foram analisados usando a linguagem de programação Python (PYTHON, 2019) por meio da interface `jupyter notebook`. Os principais pacotes foram: `pandas` (leitura e organização dos dados), `matplotlib` (elaboração de gráficos) e `geopandas` (análise espacial). Os dados e códigos estão disponíveis em github.com/patriciasiqueira.

Como primeira parte da Análise Exploratória de Dados Espaciais (AEDE), os dados agregados por Regiões Geográficas Imediatas serão apresentados por meio de mapas temáticos, de forma a ilustrar o padrão espacial da produtividade ao longo do período. Foi adotado o recorte geográfico de Regiões Geográficas Imediatas para eliminar o problema de dados inflacionados por zeros existente na divisão por municípios. Para contextualização, no ano de 2016, dos 853 municípios de MG, 383 (45%) tinham área de produção igual a zero. Contudo, considerando as 70 regiões geográficas imediatas, apenas 5 não produziram. Além disso, a divisão em questão é a divisão regional mais recente e atualizada adotada pelo IBGE.

A divisão dos estados em mesorregiões e microrregiões foi utilizada pelo IBGE de 1989 a 2017. Nessa divisão, os estados da federação se subdividem em mesorregiões, as quais se subdividem em microrregiões, que são compostas por um conjunto de municípios. Em 2017 foi publicado pelo IBGE um novo modelo de divisão regional do Brasil, a partir do qual os estados se dividem em Regiões Geográficas Intermediárias, as quais se subdividem em Regiões Geográficas Imediatas, que são compostas por um grupo de municípios. Assim, as Regiões Geográficas Intermediárias e as Regiões Geográficas Imediatas substituem as antigas mesorregiões e microrregiões, respectivamente. O novo recorte buscou atualizar a divisão regional do Brasil a partir das mudanças ocorridas desde a última divisão e utilizou em sua metodologia diversas variáveis como a localização dos municípios além de dados de pesquisas do IBGE como o deslocamento da população em busca de serviços em municípios vizinhos, definindo um município polo para cada região, dentre outros (IBGE, 2017). A Figura 16 apresenta o mapa de Minas Gerais com a divisão das Regiões Geográficas Intermediárias:

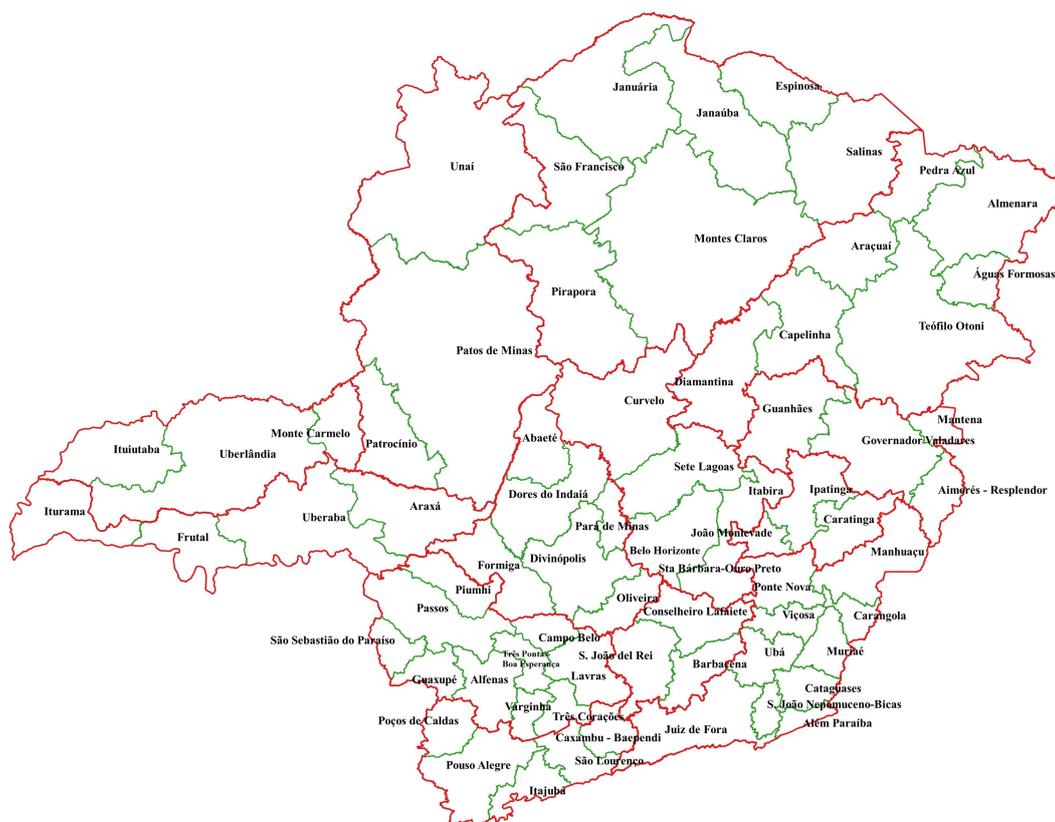
Figura 16 – Regiões Geográficas Intermediárias de Minas Gerais



Fonte: Elaboração própria

Comparando a divisão em questão com a divisão anterior, o Estado de Minas Gerais possui 12 mesorregiões, mas 13 regiões geográficas intermediárias. As regiões são bastante heterogêneas no que diz respeito à quantidade de municípios: a Região Intermediária de Juiz de Fora possui 146 municípios, enquanto a Região de Uberlândia possui 24. Analisando o nível inferior da divisão, Minas Gerais possui 66 microrregiões, mas 70 regiões geográficas imediatas, as quais foram adotadas neste trabalho. A Figura 17 apresenta o mapa de Minas Gerais com a divisão das Regiões Geográficas Imediatas (linhas verdes), destacando as Regiões Geográficas Intermediárias (linhas vermelhas):

Figura 17 – Regiões Geográficas Imediatas de Minas Gerais



Fonte: Elaboração própria

Nas Regiões Geográficas Imediatas também há grandes diferenças entre os números de municípios, como exemplo, a Região Imediata de Pouso Alegre possui 34 municípios, enquanto algumas regiões como Varginha, Campo Belo e Iturama, apenas 5.

No segundo passo da AEDE, o valor do I de Moran global foi calculado e sua significância avaliada por meio do pseudo valor- p , obtido a partir de 999 permutações aleatórias. Neste estudo foi considerado um nível de significância $\alpha = 0,05$. Foi adotada a matriz de pesos espaciais de contiguidade com a convenção rainha (dado que as regiões tem contornos irregulares, a adoção do padrão rainha ou torre tem os mesmos resultados) considerando os vizinhos de primeira ordem. De acordo com Anselin (1988), a escolha adequada da matriz de pesos espaciais é uma das questões metodológicas mais difíceis e controversas na análise exploratória de dados espaciais e na própria econometria espacial. No trabalho em questão poderia ser adotada uma matriz baseada em distância ou contiguidade. As matrizes baseadas em distâncias são especialmente úteis em situações em que existem ilhas entre as regiões de estudo (o que

não é o caso das regiões avaliadas neste trabalho), as quais ficariam sem vizinhos se fosse adotado um critério de contiguidade. Na adoção de uma matriz baseada em distância geográfica o número de vizinhos é fixo para todas as regiões e, dadas as características das Regiões Geográficas Imediatas de MG, tal critério causaria algumas distorções devido à heterogeneidade das áreas das regiões. Algumas regiões (especialmente as menores em área) são limítrofes de um número elevado de regiões: Lavras, Oliveira, Itabira e Campo Belo, por exemplo, são limítrofes de seis outras regiões, cada uma; Ubá faz divisa com sete regiões e Ponte Nova com oito. Sendo assim, adotar uma distância que estabeleça um número de vizinhos menos que seis, por exemplo, poderia estar subestimando a influência de regiões que compartilham fronteiras. Contudo, algumas regiões (especialmente as maiores) tem poucos vizinhos limítrofes e poucas regiões próximas. Ao se adotar um valor de $k = 5$, por exemplo, as regiões de Iturama e Monte Carmelo passam a ser consideradas vizinhas, entretanto, a menor distância entre as duas regiões é de mais de 200 km. Assim, poderia se estar superestimando a capacidade de influência de regiões de elevada distância geográfica. Dado o exposto, aliado ao fato de que a matriz de contiguidade espacial foi a mais encontrada na literatura em trabalhos semelhantes, esta foi a adotada. Todavia, destaque-se que, a título de conferência, também foram testadas as matrizes de distância (k -vizinhos mais próximos) geográfica com $k = 4, 5, 6$ e 7 vizinhos, as quais apresentaram resultados (disponíveis no apêndice do trabalho) semelhantes aos da matriz de contiguidade, tanto em sinal quanto em significância, reforçando a robustez dos resultados.

Definida a matriz, se o valor calculado do I de Moran global for significativo, isso indica que a hipótese de aleatoriedade espacial deve ser rejeitada e há evidências de autocorrelação espacial. Além disso, o diagrama de dispersão de Moran foi obtido em cada caso, de forma a identificar se há predominância de *clusters* espaciais (agrupamentos do tipo AA e BB) ou de *outliers* espaciais, conforme apresentado na subseção 2.2.3.1.

O terceiro passo da análise espacial da produtividade consistiu na obtenção dos mapas LISA, de forma a identificar padrões locais de autocorrelação espacial e *outliers* espaciais. Esses mapas ilustram cada uma das observações e indicam se seus valores foram considerados significativos em relação à estatística do I de Moran local. Nos mapas LISA, os *clusters* espaciais e *outliers* espaciais são identificados por cores diferentes das regiões não significativas.

4 RESULTADOS E DISCUSSÃO

4.1 ESTATÍSTICAS DESCRITIVAS

Inicialmente são apresentadas as estatísticas descritivas para a produtividade do café, em kg/hectare, de 2002 a 2017. A Tabela 1 apresenta as estatísticas da produtividade por regiões geográficas imediatas.

Tabela 1 - Estatísticas descritivas para a produtividade de café por região geográfica imediata de MG em kg/hectare

ano	média	desvio padrão	mín	$q_{0,25}$	$q_{0,50}$	$q_{0,75}$	máx
2002	1.025,35	542,33	0,0	710,69	999,98	1.238,12	2.866,10
2003	916,67	409,19	0,0	660,23	866,48	1.084,82	2.200,00
2004	1.106,10	499,10	0,0	878,36	1.034,40	1.239,48	3.057,69
2005	1.071,16	508,68	0,0	784,71	999,99	1.185,97	3.425,00
2006	1.241,51	616,48	0,0	840,24	1.096,16	1.447,11	4.000,00
2007	1.061,82	487,15	0,0	799,60	969,79	1.130,25	2.638,86
2008	1.243,36	574,50	0,0	1.000,51	1120,11	1.433,52	3.734,17
2009	1.195,08	568,00	0,0	986,13	1.079,42	1.294,82	3.491,24
2010	1.307,11	698,93	0,0	1.008,05	1.122,58	1.575,06	3.871,42
2011	1.284,83	623,75	0,0	994,14	1.151,67	1.410,18	3.832,00
2012	1.426,60	659,20	0,0	1.143,64	1.336,75	1.751,76	3.993,97
2013	1.420,03	634,88	0,0	1.145,30	1.432,11	1.632,95	3.967,54
2014	1.312,20	650,23	0,0	968,64	1.183,35	1.640,84	3.632,93
2015	1.260,02	505,72	0,0	1.100,99	1.232,41	1.422,57	2.710,58
2016	1.501,88	632,61	0,0	1.198,09	1.550,85	1.884,10	2.793,19
2017	1.382,23	637,61	0,0	1.082,86	1.479,77	1.782,05	2.547,43

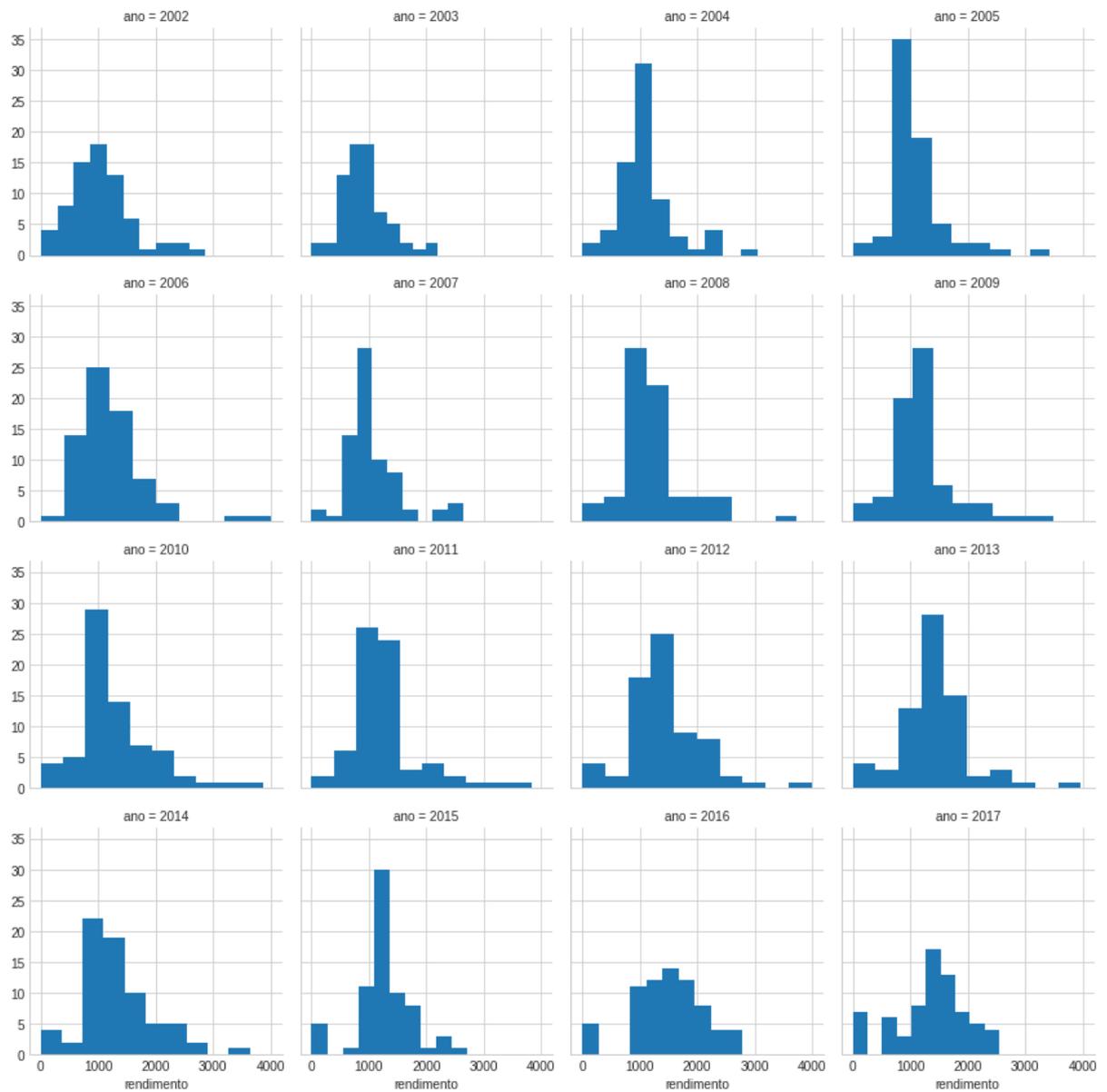
Fonte: Elaboração própria

Nota: $q_{0,25}$: Primeiro quartil. $q_{0,50}$: Segundo quartil. $q_{0,75}$: Terceiro quartil.

Pela Tabela 1 é possível visualizar que a produtividade média vem aumentando ao longo do tempo (mais precisamente, um aumento de 34,8% entre 2002 e 2017), indicando um aumento de eficiência na produção de café, considerando a produção pela área colhida. O valor mínimo se manteve em zero durante todo o período, dado que em todos os anos pelo menos uma região não produziu café no ano. O valor do primeiro quartil (25%) também aumentou, em um valor de 52,4%, indicando uma grande evolução das regiões que apresentavam pior desempenho. A mediana teve aumento ligeiramente menor, de 48%. O terceiro quartil (75%) teve seu valor aumentado em 43,9%. Por fim, o valor máximo foi reduzido de 2866,10 kg/hectare em 2002 para 2547,43 kg/hectare em 2017. O aumento relativamente maior do primeiro quartil em relação

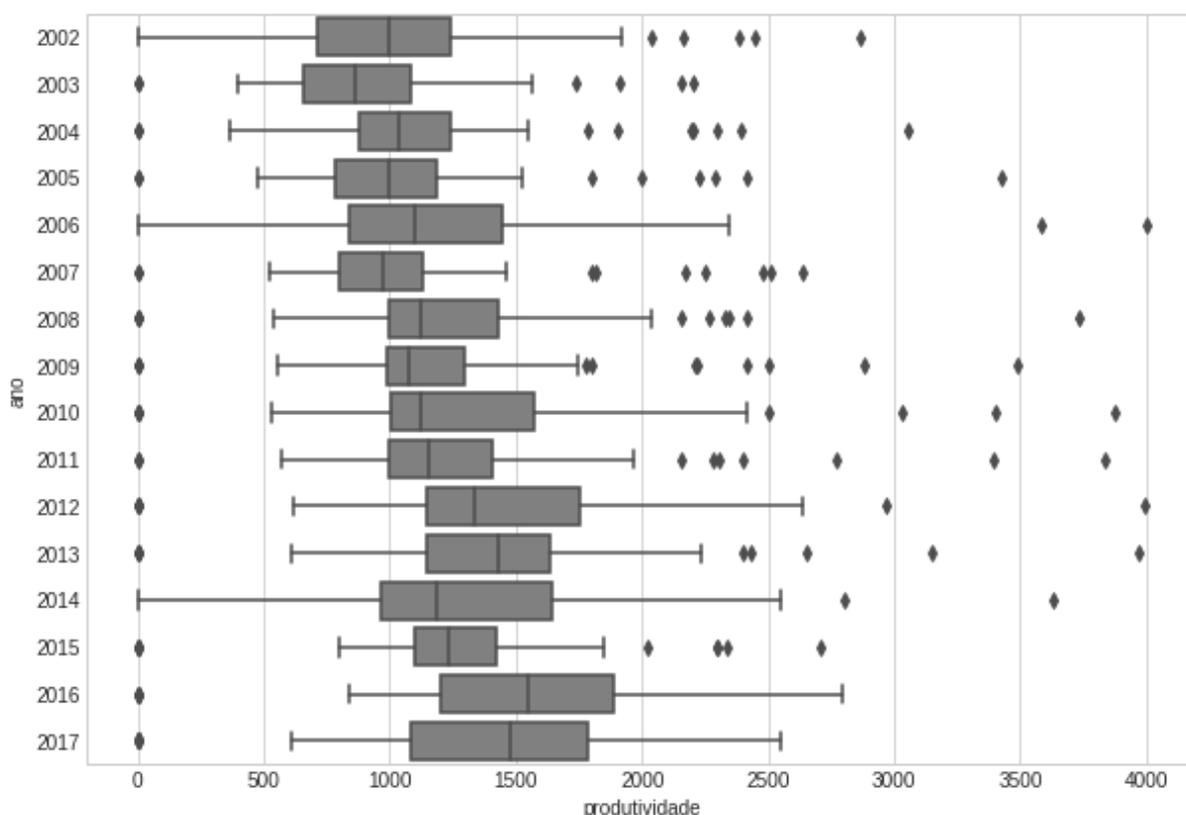
à mediana e da mediana em relação ao terceiro quartil indica que as regiões que apresentavam pior desempenho foram as que mais evoluíram seus níveis de produtividade, sugerindo uma tendência de homogeneização dos valores. Nos Gráficos 1 e 2 estão os histogramas da produtividade e *boxplots* da produtividade, respectivamente.

Gráfico 1 – Histogramas da produtividade de café por regiões geográficas imediatas de MG de 2002 a 2017



Fonte: Elaboração própria

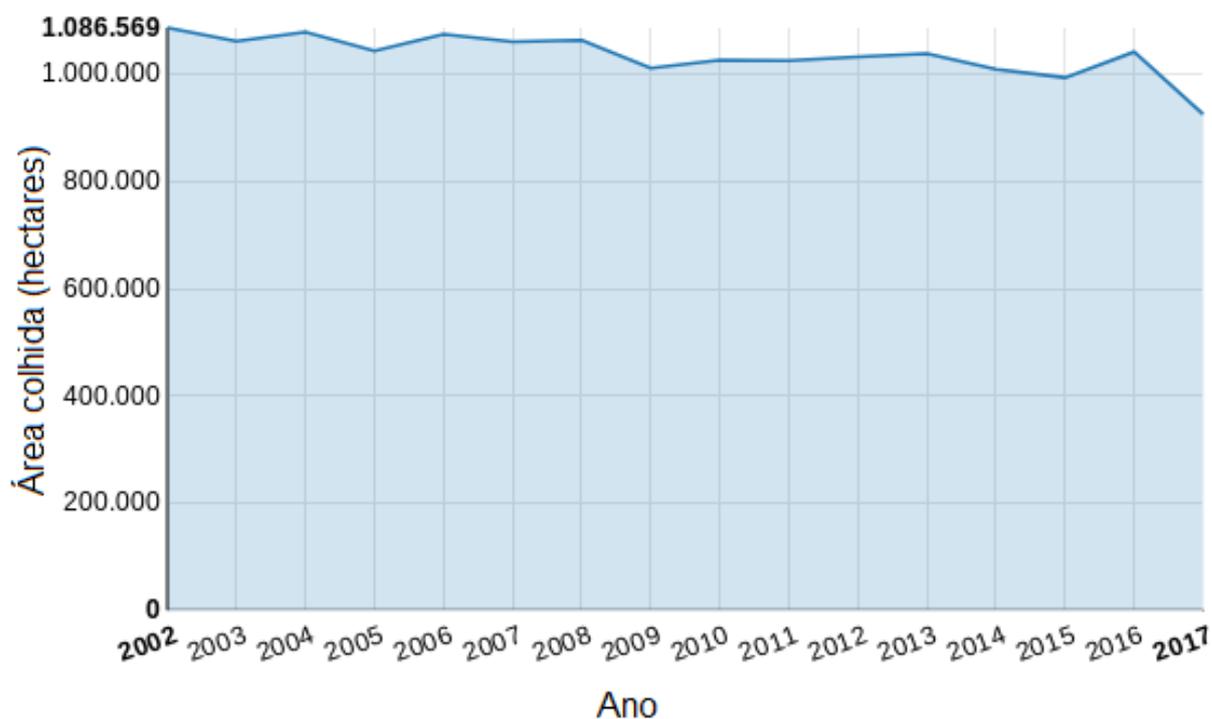
Gráfico 2 – *Boxplots* da produtividade de café, por regiões geográficas imediatas de MG, de 2002 a 2017



Fonte: Elaboração própria

Pelos Gráficos 1 e 2 também é possível visualizar que a produtividade média vem aumentando. No Gráfico 2 identifica-se que em vários anos, os valores máximo se caracterizam como *outliers*, estando isolados do restante dos dados. De fato, considerando os dados do estado de Minas Gerais como um todo, a produtividade média de todo o estado cresceu ao longo do período estudado. Analisando os dados de MG, percebe-se que o aumento da produtividade não foi efeito apenas da redução da área colhida (mantida a produção) ou aumento da produção (mantida a área colhida), mas sim consequência dos dois fatores: aumento da produção e redução da área colhida. Em 2002 foram produzidas 1.301.209 toneladas de café em 1.086.569 hectares, resultando em uma produtividade de 1.197 kg/hectare. Em 2017 foram produzidas 1.454.779 toneladas em 925.108 hectares, resultando em uma produtividade de 1.573 kg/hectare. O Gráfico 3 mostra a evolução da área total colhida de café no estado, em hectares, de 2002 a 2017:

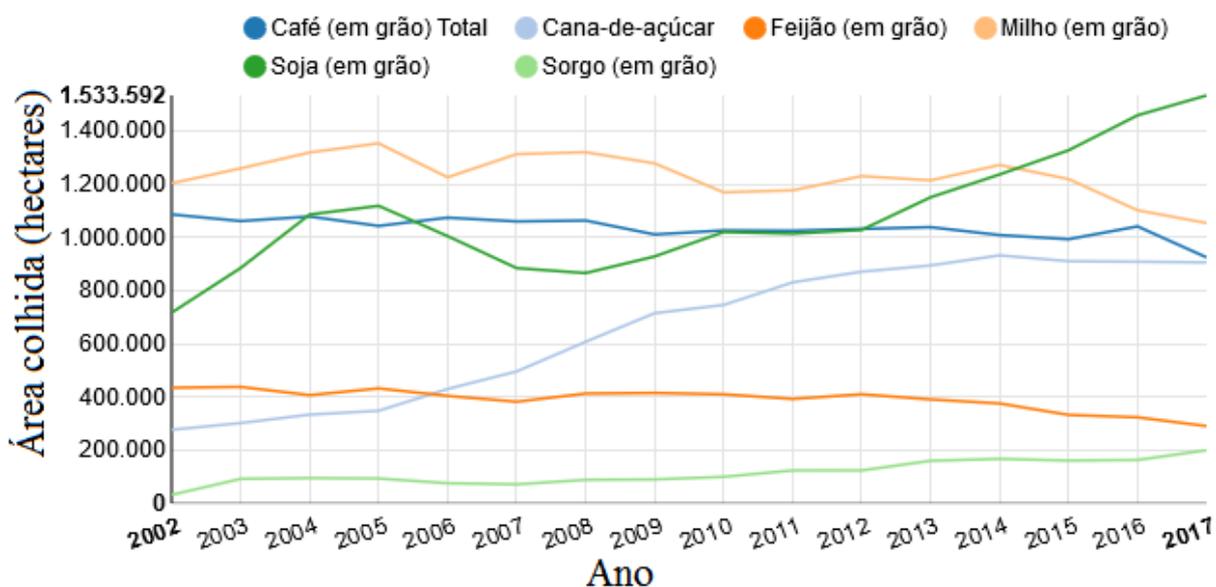
Gráfico 3 – Área colhida de café em MG, em hectares, de 2002 a 2017



Fonte: IBGE; SIDRA, Tabela 1613

Pela Gráfico 3 é possível visualizar a redução da área colhida de café em Minas Gerais ao longo do período. Houve uma redução de mais de 160 mil hectares de área. Este fato talvez possa ser explicado, dentre outros fatores, pela substituição de produtos agrícolas no estado nos últimos anos. O Gráfico 4 mostra a evolução da área colhida das 6 principais (foram selecionados os produtos que apresentaram área colhida acima de 100000 hectares em pelo menos dois anos da série) lavouras de Minas Gerais:

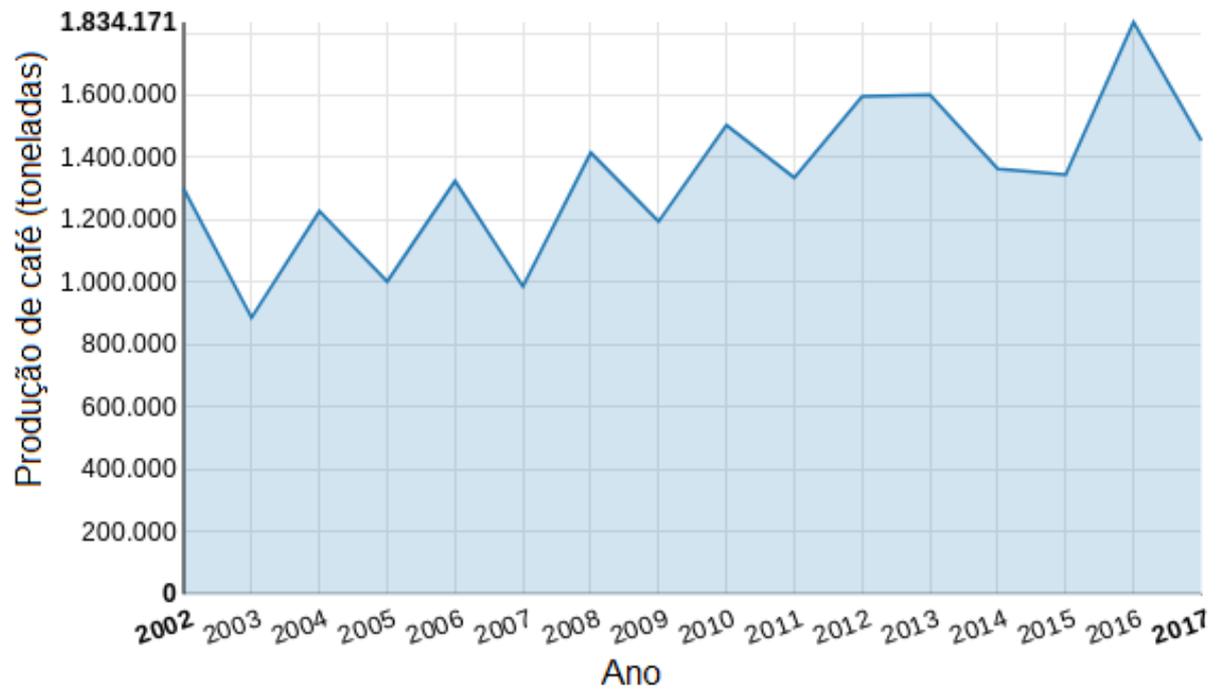
Gráfico 4 – Área colhida em MG, em hectares, de 2002 a 2017



Fonte: IBGE - SIDRA, Tabela 5457

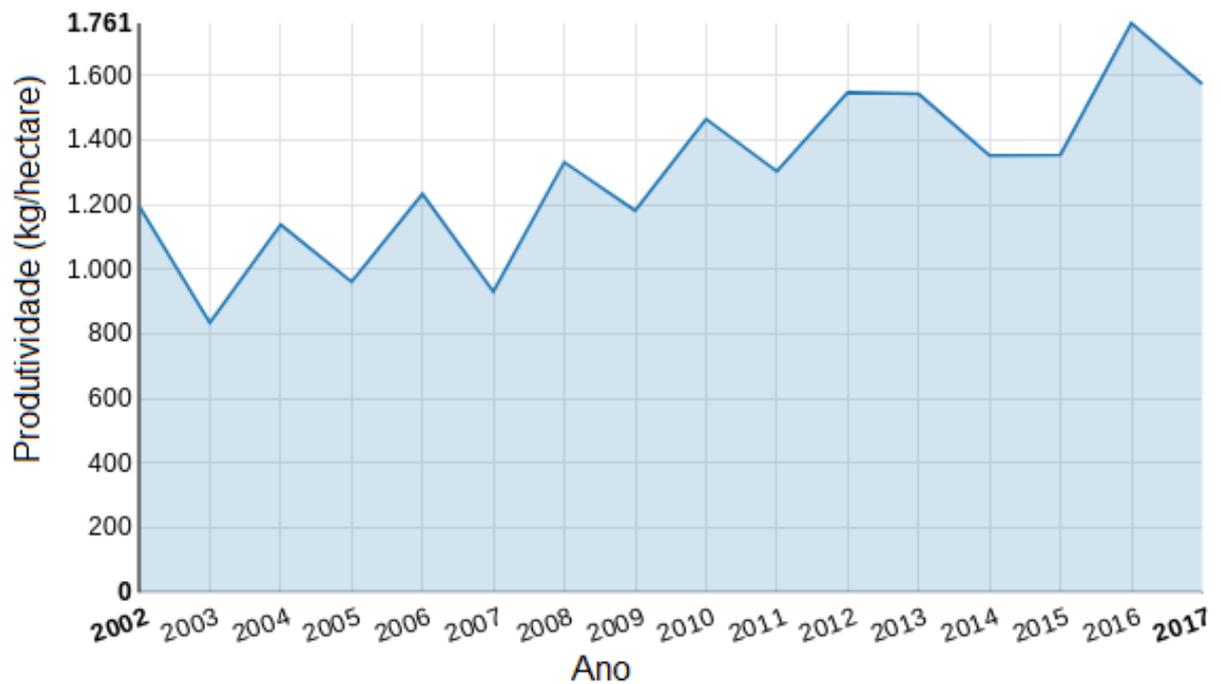
Das seis principais culturas, três aumentaram e três reduziram a área colhida. Contudo, em números absolutos, o aumento foi bem maior que a redução: A área colhida de soja teve aumento de mais de 815 mil hectares, a área de cana-de-açúcar aumentou mais de 628 mil hectares e a área colhida de sorgo, mais de 167 mil hectares. Somadas, as três culturas tiveram aumento de 1 milhão e 600 mil hectares. A soma das áreas reduzidas de café, feijão e milho e de aproximadamente 450 mil hectares. Estes movimentos podem indicar uma tendência de alteração da dinâmica agrícola do estado. Prosseguindo com a análise específica do café, o Gráfico 5 apresenta a evolução da produção de café, em toneladas no estado de MG e o Gráfico 6 apresenta a produtividade do café em MG ao longo do período estudado:

Gráfico 5 – Produção de café em MG, em toneladas, de 2002 a 2017



Fonte: IBGE; SIDRA, Tabela 1613

Gráfico 6 – Produtividade de café em MG, em kg/hectare, de 2002 a 2017



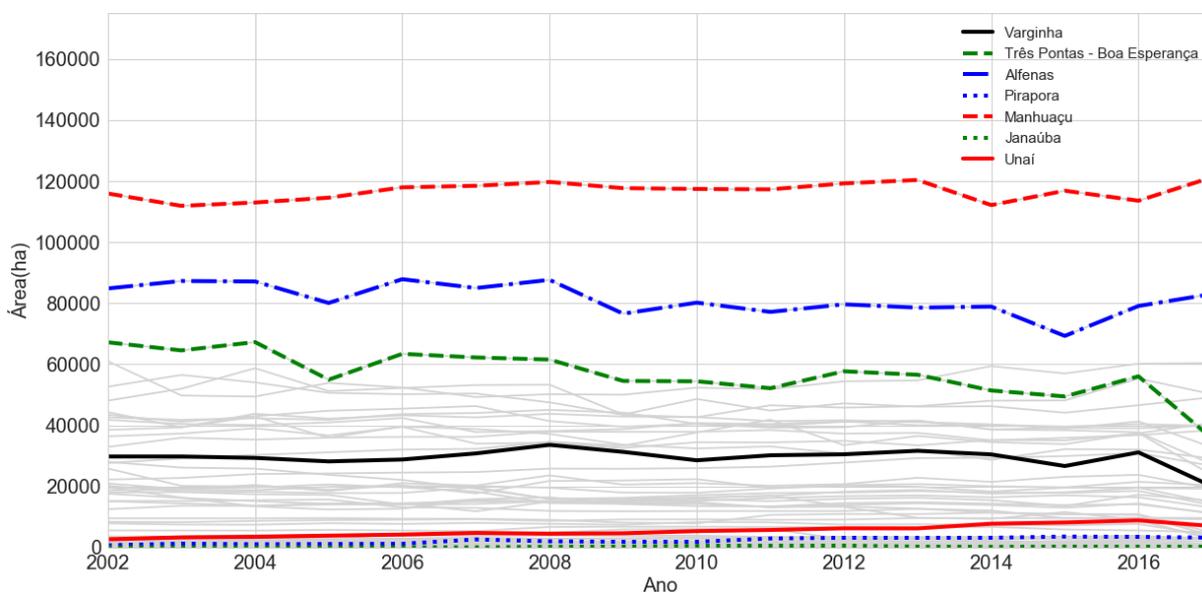
Fonte: IBGE; SIDRA, Tabela 1613

De acordo com os Gráficos 5 e 6 é possível visualizar a tendência de aumento da produtividade e da produção ao longo do período. O aumento da produtividade foi de 31,4% de 2002 para 2017 e pode-se perceber que o aumento na produção (11,8% de 2002 para 2017) foi, em valores relativos, um pouco menos impactante do que a redução da área (14,9% de 2002 para 2017) no aumento da produtividade, considerando o estado de MG como um todo.

A série de produção, mesmo com tendência de crescimento, apresenta comportamento oscilante, alternando entre altos e baixos, dada a característica da bienalidade da produção do café. Destaca-se uma quebra no padrão de alternância entre safras altas e baixas no ano de 2015. O ano de 2014 apresentou uma produção baixa e se esperava uma alta produção em 2015, contudo, devido a uma forte seca, a produção de 2015 também foi baixa, sendo que a produção alta ocorreu apenas em 2016 (SECRETARIA DE ESTADO DE AGRICULTURA, PECUÁRIA E ABASTECIMENTO DE MINAS GERAIS, 2017). O efeito da bienalidade se reflete na série de produtividade, que apresenta o mesmo comportamento durante todo o período. A série de produção também se caracteriza por grande dispersão entre as regiões em todos os anos, dado que algumas regiões não produzem e outras produzem quantidade muito pequena. Em 2002, a média de produção foi de aproximadamente 18,6 mil toneladas por região geográfica, mas 50% das regiões produziram até 3,4 toneladas. Por outro lado, algumas regiões como Manhuaçu e Alfenas, apresentaram produção acima de 120 mil toneladas em quase todos os anos. A série de produção, assim com a de produtividade, teve seu pico em 2016, com produção acima de 1,8 milhão de toneladas (média de mais de 26 mil toneladas por região).

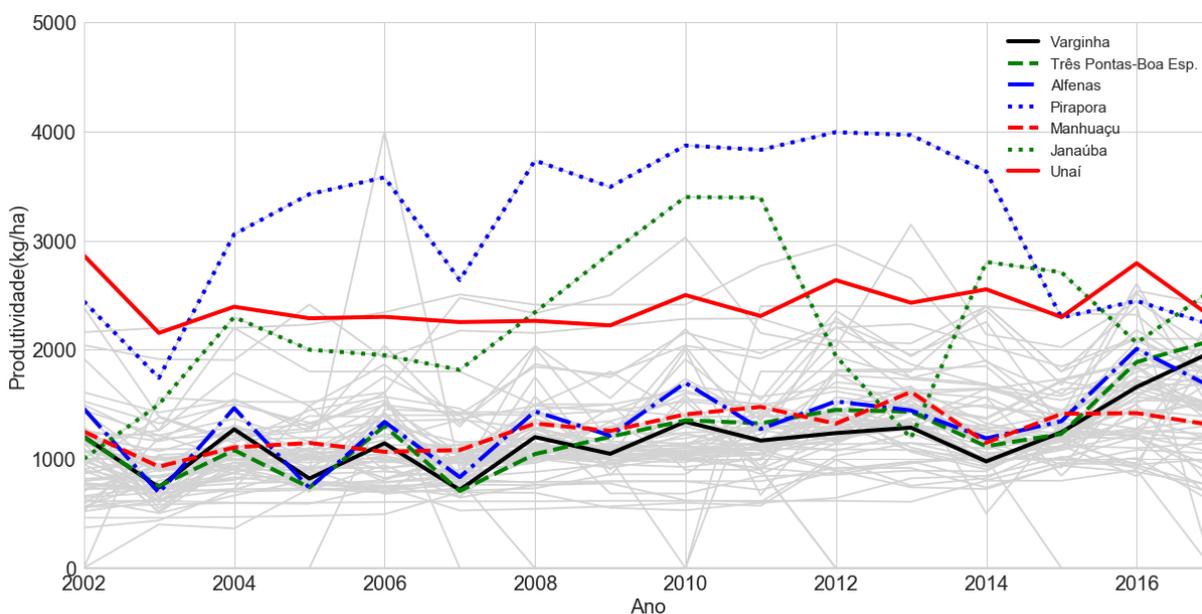
Em complemento, a fim de compreender a evolução das séries de produtividade no estado ao longo do tempo foram analisadas também as séries históricas de algumas regiões, individualmente, com maiores detalhes. São apresentadas nos Gráficos 7 e 8 as séries de área e produtividade de sete regiões de destaque. Três regiões dentre as que apresentaram, em média, as maiores áreas colhidas ao longo do tempo, Manhuaçu, Alfenas e Três Pontas-Boa Esperança, três dentre as regiões que apresentaram maior produtividade ao longo do tempo, Pirapora, Janaúba e Unaí, além da região de Varginha. Os Gráficos 7 e 8 apresentam as séries históricas de área colhida e produtividade, respectivamente.

Gráfico 7 – Área de café colhida de 2002 a 2017 por região geográfica imediata de MG



Fonte: Elaboração própria

Gráfico 8 – Produtividade de café de 2002 a 2017 por região geográfica imediata de MG



Fonte: Elaboração própria

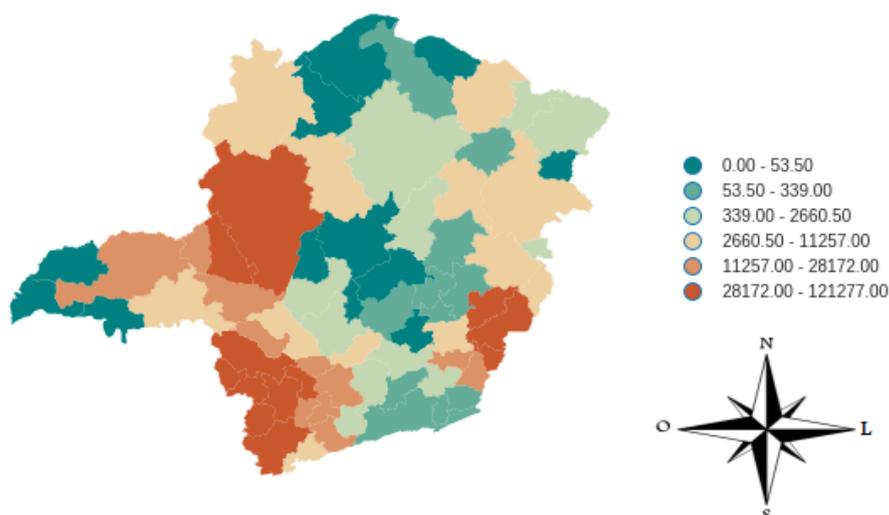
Pela análise dos Gráficos 7 e 8 percebe-se que as três regiões com maior área colhida ao longo do tempo não apresentaram os maiores índices de produtividade. Ao mesmo tempo, as três regiões que estiveram em boa parte do tempo com os maiores índices de produtividade apresentaram algumas das menores áreas colhidas ao longo da série. Cabe ressaltar também que

as regiões de Varginha, Três Pontas-Boa Esperança, Alfenas e Manhuaçu, mesmo estando em patamares distantes de área colhida, apresentaram desempenho de produtividade muito semelhante ao longo da série. Intuitivamente poderia se supor que as regiões com as maiores áreas de café teriam os maiores índices de produtividade, por se tratarem de regiões “especialistas” que, teoricamente, poderiam contar com técnicas mais avançadas devido ao foco na produção em larga escala. Contudo, não é o que foi identificado, ao menos nas regiões de maior produção e maior produtividade, sugerindo que o tamanho da área colhida não seja um fator determinante para o nível de produtividade.

4.2 MAPAS TEMÁTICOS

A segunda parte da análise trata da visualização dos mapas temáticos, de forma a ilustrar e buscar identificar visualmente o padrão espacial da produtividade. O primeiro mapa, apresentado na Figura 18, apresenta a distribuição da plantação de café em MG no ano de 2017, medida pela área total colhida em cada região geográfica em hectares. Para a construção do mapa os valores foram divididos em 6 intervalos, valor que será adotado posteriormente.

Figura 18 – Área total de café colhida, por região, em 2017



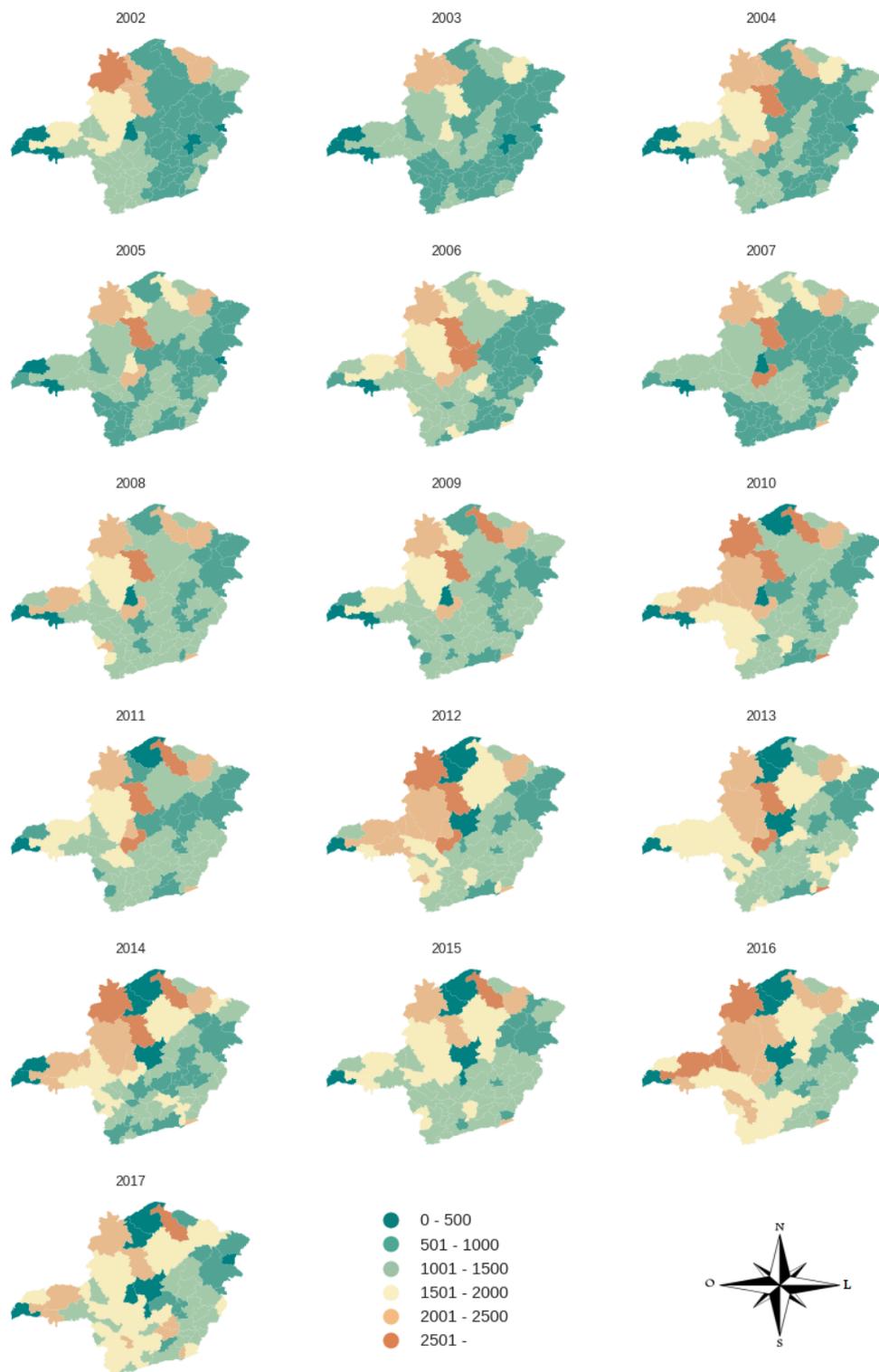
Fonte: Elaboração própria

Pela análise do mapa da Figura 18 percebe-se que, apesar de algumas concentrações com maiores plantações e áreas sem produção, existia em 2017 produção relevante de café

por praticamente todo o estado, sendo que das 70 regiões geográficas imediatas, apenas 7 não colheram café.

A Figura 19 mostra os mapas temáticos para a produtividade (kg/hectare) de café por região geográfica, de 2002 a 2017. Para a construção dos mapas foram definidos valores fixos para todos os anos, de modo que os mapas de diferentes anos pudessem ser comparáveis. Os valores foram definidos de modo que pudessem ser criados 6 intervalos e os intervalos ficassem do mesmo tamanho, com faixas de 500 kg/hectare, à exceção da faixa mais elevada, que contempla todos os valores acima de 2500 kg/hectare.

Figura 19 – Produtividade de café por regiões geográficas imediatas de MG, de 2002 a 2017



Fonte: Elaboração própria

Pela análise da Figura 19 é possível visualizar que o padrão espacial de produtividade por região modificou-se ao longo do tempo. Nos primeiros anos, poucas regiões tinham produtividade média acima de 1.500 kg/hectare e estavam localizadas, predominantemente, no Norte e Noroeste de MG, mais especificamente nas regiões intermediárias de Montes Claros e Patos de Minas. Com o passar dos anos, outras regiões do estado passaram a apresentar índices mais elevados de produtividade e a distribuição das regiões de alta produtividade ficou mais espalhada pelo estado. É possível que o processo tenha como um dos fatores a influência da vizinhança das regiões que já possuíam níveis mais elevados, já que as primeiras regiões a apresentarem produtividade mais elevada ao longo dos anos foram as vizinhas das regiões que já estavam nessa condição e assim sucessivamente, gerando um efeito de espalhamento da “alta produtividade” ao longo do estado. Analisando-se gradativamente a evolução, com uma visão aprofundada das regiões envolvidas em alguns anos de destaque, é possível identificar esse movimento, conforme descrito a seguir. Para fins de avaliação será definida (apenas para comparação entre as regiões aqui citadas) como alta produtividade as regiões presentes nos três intervalos superiores, ou seja, com produtividade média acima de 1.500 kg/hectare.

Em 2002, das 70 regiões geográficas imediatas, apenas oito apresentaram produtividade média acima de 1.500 kg/hectare. Essas regiões estavam distribuídas em quatro regiões geográficas intermediárias, sendo que quatro regiões imediatas estavam na região intermediária de Montes Claros (Salinas, Pirapora, Espinosa e São Francisco), duas na região intermediária de Patos de Minas (Patos de Minas e Unai), uma na região intermediária de Uberlândia (Uberlândia) e uma na região intermediária de Uberaba (Araxá). Já em 2006, o número de regiões imediatas com produtividade acima de 1.500 kg/hectare subiu de oito para quinze. Das oito regiões que estavam nesse patamar em 2002, seis se mantiveram, e duas, Espinosa e Araxá (em 2006 a produtividade média foi de 1.492 kg/hectare) saíram. Além destas seis, nove outras regiões entraram no grupo. Das nove, cinco eram vizinhas de regiões que já possuíam produtividade acima de 1.500 kg/hectare, são elas: Janaúba (vizinha de Salinas e Espinosa), Curvelo (vizinha de Pirapora), Dolores do Indaiá (vizinha de Patos de Minas e de Araxá), Abaeté (vizinha de Patos de Minas) e Monte Carmelo (vizinha de Uberlândia e de Araxá). Analisando pelo recorte das regiões intermediárias, em 2006 as regiões geográficas de produtividade elevada estavam distribuídas em oito regiões intermediárias. A região de Montes Claros possuía quatro regiões imediatas com produtividade acima de 1.500 kg/hectare, as regiões de Belo Horizonte, Patos de Minas, Uberlândia, e Divinópolis possuíam duas cada e as regiões de Varginha, Pouso

Alegre e Juiz de Fora, uma cada. Destaca-se que, apesar do aumento expressivo (oito para quinze) de regiões imediatas com alta produtividade, o número de regiões no quinto e sexto intervalos, ou seja, regiões imediatas com produtividade média acima de 2.000 kg/ hectares, não foi alterado, eram cinco em 2002 e cinco em 2006.

Em 2010, mais uma vez houve aumento no número de regiões geográficas imediatas com produtividade acima de 1.500 kg/hectare (quarto, quinto e sexto intervalos dos mapas), passando para 21. Das quinze regiões imediatas de 2006, dez se mantiveram. Das onze “novas” (Araxá, que já estava no grupo em 2002, voltou), oito eram vizinhas de regiões que já apresentavam produtividade elevada em 2006 e/ou 2002. As exceções são as regiões imediatas de Poços de Caldas, Alfenas e São João del Rei. Analisando as regiões geográficas intermediárias, as 21 regiões imediatas se distribuem da seguinte forma: cinco na região intermediária de Varginha, três nas regiões de Uberlândia, Patos de Minas e Montes Claros, duas nas regiões de Uberaba e Divinópolis e uma nas regiões de Pouso Alegre, Juiz de Fora e Barbacena.

No ano de 2013, o número de regiões geográficas imediatas que apresentaram alta produtividade média foi de 26. Destas, quinze já constavam no grupo em 2010 e uma, a região imediata de Abaeté, já havia aparecido em 2006. Das outras dez, seis eram vizinhas de regiões imediatas que estavam no grupo de alta produtividade nos anos anteriores. Em relação às quatro restantes, Viçosa, Ponte Nova, Manhuaçu e Muriaé, não são vizinhas de outras regiões que já apresentaram produtividade alta, mas são vizinhas entre si, todas pertencentes à região intermediária de Juiz de Fora que até então só apresentava a região de Além Paraíba no grupo em questão.

Por fim, no último ano avaliado, 2017, o número de regiões do grupo de alta produtividade foi de 34, um aumento bem expressivo em relação ao primeiro ano do estudo. Das 34, 17 estavam no grupo em 2013, seis não estavam em 2013, mas já estiveram em um dos outros três anos destacados. Das onze restantes, sete eram vizinhas de regiões que pertenceram ao grupo de alta produtividade em 2013. Diante do exposto, percebe-se que a característica de alta produtividade foi se espalhando pelo estado, em parte, através das regiões vizinhas das que já possuíam alta produtividade. A Tabela 2 mostra a evolução do número de regiões imediatas no grupo de alta produtividade em cada região geográfica intermediária nos anos destacados e a Tabela 3 mostra a evolução do número de regiões geográficas imediatas em cada intervalo nos anos selecionados.

Tabela 2 - Número de regiões geográficas imediatas de alta produtividade (acima de 1.500 kg/hectare) por região intermediária em cada ano destacado

Região Intermediária	2002	2006	2010	2013	2017
Barbacena	0	0	1	0	2
Belo Horizonte	0	2	0	1	2
Divinópolis	0	2	2	3	3
Governador Valadares	0	0	0	0	1
Ipatinga	0	0	0	0	0
Juiz de Fora	0	1	1	6	3
Montes Claros	4	3	3	3	4
Patos de Minas	2	3	3	3	2
Pouso Alegre	0	1	1	1	3
Teófilo Otoni	0	0	0	1	1
Uberaba	1	0	2	3	2
Uberlândia	1	2	3	3	3
Varginha	0	1	5	2	8
TOTAL	8	15	21	26	34

Fonte: Elaboração própria

A Tabela 2 mostra o efeito do espalhamento das regiões imediatas de alta produtividade. Em 2002, das treze regiões geográficas intermediárias, apenas quatro continham alguma região geográfica imediata de alta produtividade. Em 2006 o número aumentou, oito regiões intermediárias continham alguma região imediata de alta produtividade. Em 2010 o valor aumentou novamente, indicando o espalhamento das regiões de alta produtividade. Nesse ano, nove regiões intermediárias continham pelo menos uma região imediata de alta produtividade. Em 2013, com novo aumento, o número passou para dez. Por fim, em 2017, das treze regiões geográficas intermediárias, apenas uma, a região de Ipatinga, não continha uma região imediata com produtividade média acima de 1.500 kg/hectare. A Tabela 3 mostra em maiores detalhes a evolução do número de regiões imediatas.

Tabela 3 - Número de regiões geográficas imediatas em cada intervalo de produtividade por ano em Minas Gerais

Ano	4º intervalo 1501-2000 kg/hectare	5º intervalo 2001-2500 kg/hectare	6º intervalo Mais de 2500 kg/hectare	TOTAL
2002	3	4	1	8
2006	10	3	2	15
2010	11	6	4	21
2013	19	4	3	26
2017	25	8	1	34

Fonte: Elaboração própria

Pela Tabela 3, as variações entre o primeiro e último anos estudados, o número de regiões no quarto quantil (produtividade entre 1.501 e 2.000 kg/hectare) era de três em 2002 e passou para 25 em 2017, já o número de regiões no quinto quantil (produtividade entre 2.001 e 2.500 kg/hectare) era de quatro em 2002 e passou para oito em 2017. Por fim, o número de regiões no sexto quantil (produtividade acima de 2.500 kg/hectare) se manteve, uma em 2002 e uma em 2017. Assim, percebe-se que o aumento de produtividade do estado como um todo, já citado anteriormente, e o espalhamento das regiões de alta produtividade deram-se através do aumento do nível de produtividade de regiões que tinham níveis mais baixos e manutenção dos níveis de algumas regiões que já estavam em nível mais elevado, tornando a produtividade espacialmente mais homogênea no estado de Minas Gerais.

4.3 I DE MORAN

Após a análise visual do padrão espacial de distribuição dos dados, faz-se necessária a verificação por meio de uma medida numérica. Neste trabalho foi adotado o *I* de Moran, considerando a matriz de vizinhança do tipo rainha de primeira ordem. Para a análise foi calculado o *I* de Moran para cada ano. A Tabela 4 mostra os valores do *I* de Moran e o pseudo valor-*p* calculado a partir de 999 permutações aleatórias.

Tabela 4 - *I* de Moran para a produtividade do café em MG por regiões geográficas imediatas entre 2002 e 2017

ano	<i>I</i> de Moran	valor- <i>p</i>
2002	0.347	0.001
2003	0.370	0.001
2004	0.304	0.001
2005	0.308	0.001
2006	0.380	0.001
2007	0.202	0.010
2008	0.168	0.007
2009	0.182	0.015
2010	0.097	0.086
2011	0.132	0.031
2012	0.106	0.061
2013	0.009	0.377
2014	0.056	0.158
2015	-0.043	0.370
2016	0.099	0.065
2017	0.068	0.153

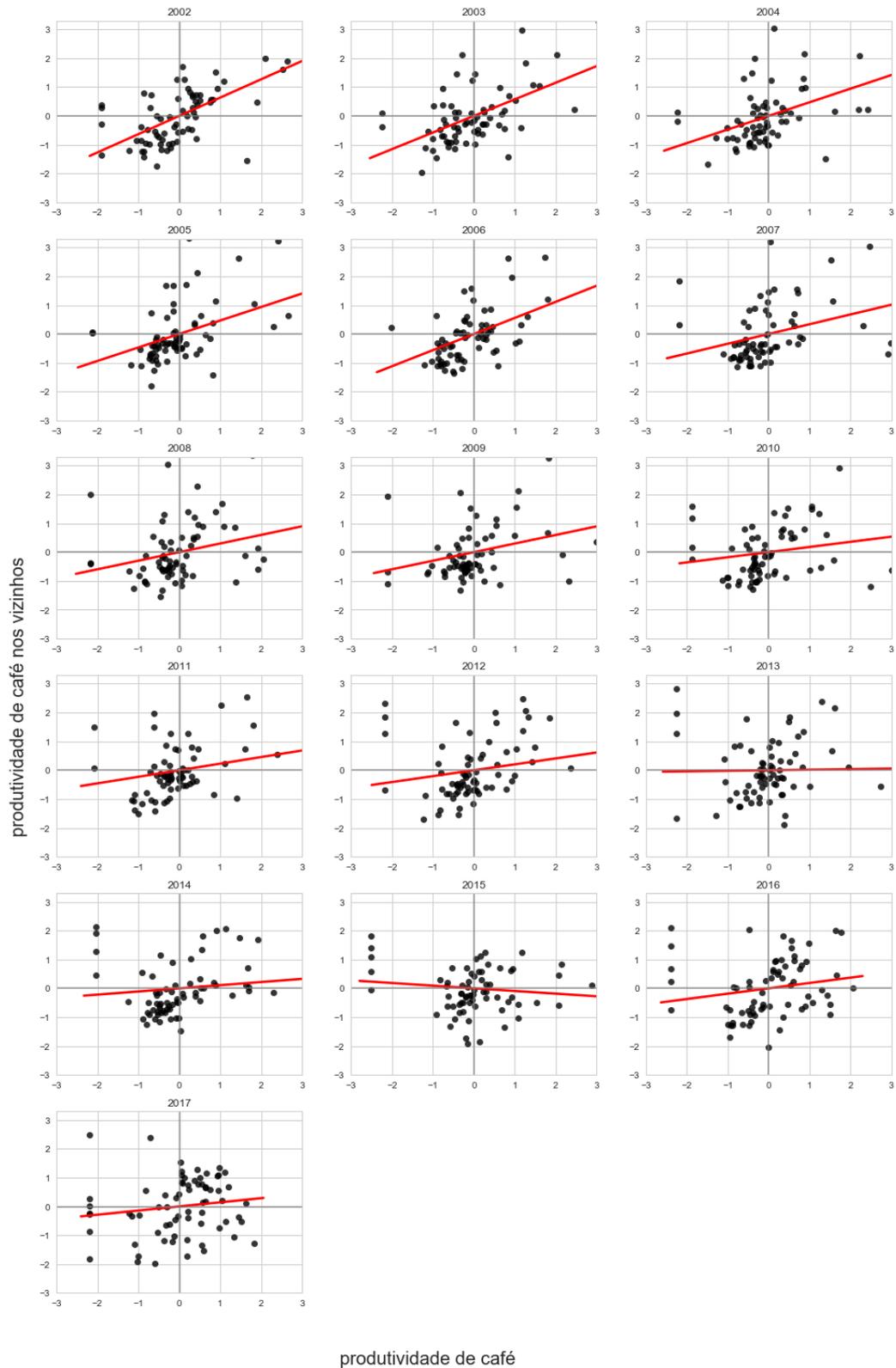
Fonte: Elaboração própria

Pelos resultados exibidos na Tabela 4, é possível dividir a análise do *I* de Moran em dois períodos: os 8 primeiros anos, de 2002 a 2009 e os oito últimos anos, de 2010 a 2017. No primeiro período, o *I* de Moran foi estatisticamente significativo ao nível de significância de 5% em todos anos, com sinal positivo, indicando a existência de autocorrelação positiva entre as regiões geográficas imediatas para a variável produtividade. Ou seja, o fato de uma região possuir produtividade alta era influenciado, dentre outros fatores, pela produtividade alta de uma região vizinha, o mesmo se aplicando à produtividade baixa. Já no segundo período, dos oito anos analisado, o índice foi significativo apenas em 2011 e em valor inferior ao dos primeiros anos. A análise indica que a dependência espacial foi sendo reduzida ao longo dos anos, o que pode ser explicado pela maior homogeneidade das regiões em relação à produtividade, conforme visto anteriormente.

4.3.1 Diagramas de dispersão de Moran

Em continuidade à análise, foram gerados os diagramas de dispersão de Moran para cada ano, exibidos no Gráfico 9.

Gráfico 9 – Diagramas de dispersão de Moran para os anos de 2002 a 2017 para a variável produtividade



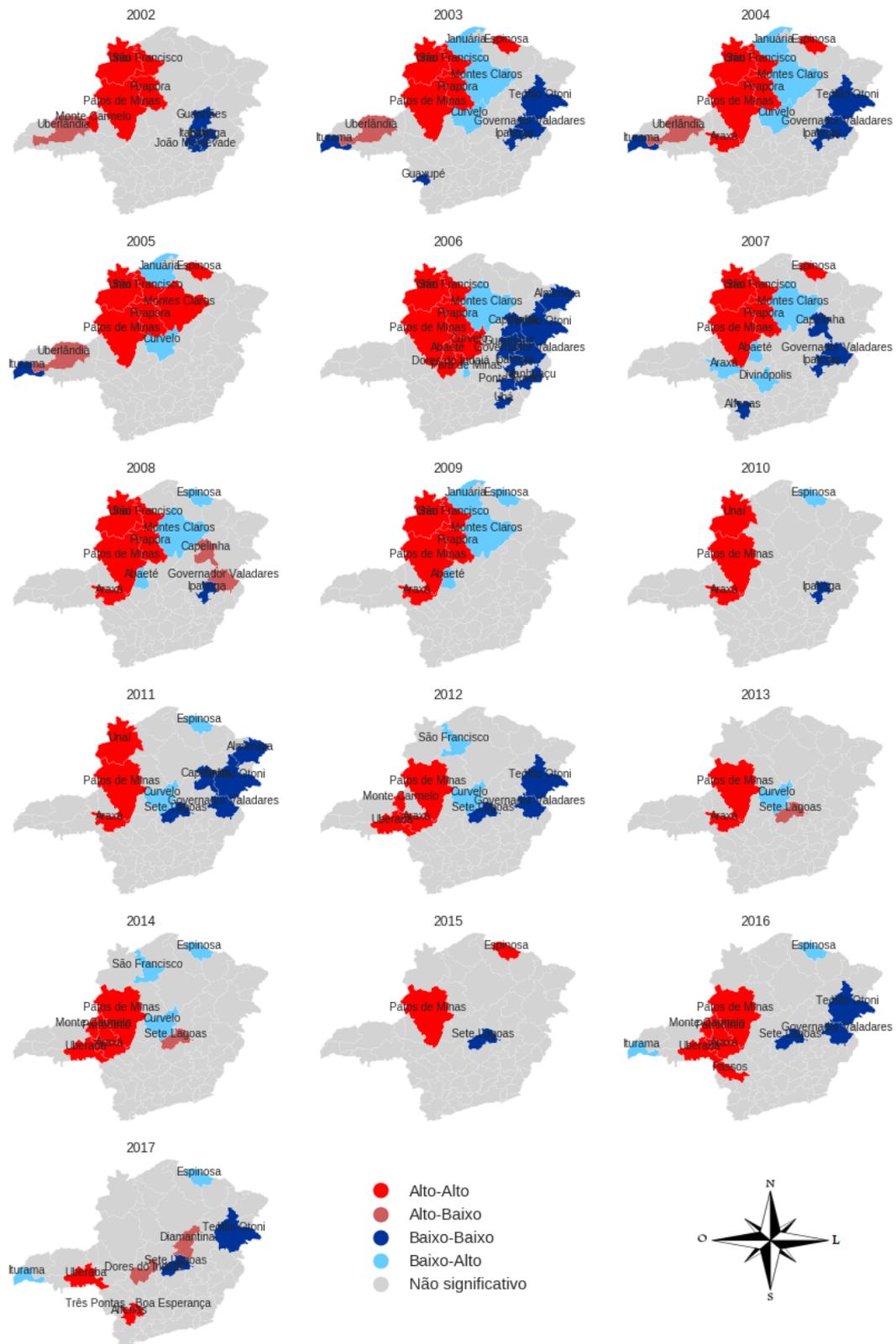
Fonte: Elaboração própria

Conforme Teixeira e Bertella (2015), os diagramas de dispersão permitem visualizar padrões locais de associação e não apenas a tendência geral dos dados. Os diagramas mostram que nos primeiros anos a maior parte das regiões estava concentrada nos quadrantes Alto-Alto e Baixo-Baixo. Em 2002, 57 das 70 regiões estavam em um desses dois quadrantes. Já nos últimos anos a distribuição está mais homogênea entre os quadrantes e a reta de ajuste não está tão inclinada, já que o coeficiente de inclinação é o valor do I de Moran. Mesmo com a visualização dos quadrantes a que pertencem cada uma das regiões, não é possível determinar, a partir dos diagramas, a existência de *clusters* ou *outliers*. Para isso devem ser visualizados os mapas LISA, que serão apresentados adiante.

4.4 MAPAS LISA (I DE MORAN LOCAL)

Foram gerados os mapas LISA para todos anos, exibidos na Figura 20. Nos mapas são mostrados os nomes das regiões que apresentaram valores LISA significativos ao nível de 5%.

Figura 20 – Mapas LISA de 2002 a 2017 para as regiões geográficas imediatas



Fonte: Elaboração própria

Os mapas da Figura 20 mostram que a quantidade de *clusters* e *outliers* significativos também apresentou tendência de redução ao longo do tempo analisado. Nos primeiros oito anos são visíveis grandes *clusters*, principalmente o *cluster* de alta produtividade na região Noroeste de MG. Com o passar dos anos, esse *cluster* foi se reduzindo.

Algumas regiões cuja produção era concentrada em poucos municípios ajudam a entender a diferença entre os cenários antes e depois de 2010. A região imediata de São Francisco caracterizava-se como uma região de alta produtividade de 2002 a 2009, sendo vizinha das regiões de Pirapora e Unaí, que se enquadraram no grupo de alta produtividade em todos os anos. A partir de 2010, a região de São Francisco apresentou baixa produtividade e, a partir de 2012, parou de produzir café, já que apenas um de seus seis municípios produzia café. Então, a partir de 2012 o padrão Alto-Alto passa a ser Alto-Baixo.

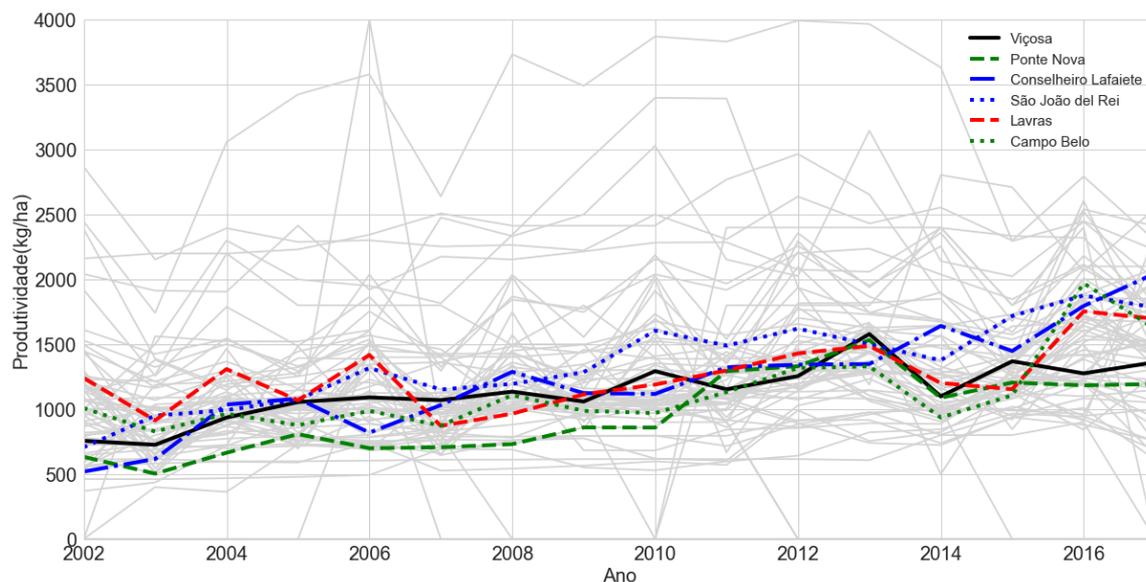
A região imediata de Pirapora ficou no grupo de alta produtividade em 12 dos 16 anos. Uma de suas regiões vizinhas, Januária, deixou de produzir café a partir de 2010. O padrão da vizinhança que foi classificado como Alto-Alto em alguns anos passou a ser Alto-Baixo a partir de então. Em Januária, em 2002, três municípios produziam café, de 2003 a 2005, dois municípios e de 2005 a 2009 apenas um município.

Na região imediata de Curvelo, em 2002, três municípios produziam café, a partir de 2003, apenas um, que deixou de produzir de 2012 em diante. A região é vizinha de duas regiões que apresentaram produtividade alta em quase todos os anos: Divinópolis e Pirapora, e teve produtividade acima da média nos anos de 2006, 2008 e 2009. Assim, as regiões Alto-Alto passaram a ser Alto-Baixo.

Os municípios observados de todas as regiões citadas que deixaram de produzir café tinham como característica em comum a pequena área colhida em todos os anos, em geral menos de 100 hectares. Ao que tudo indica, a tendência de redução da área colhida no geral, vista anteriormente, deu-se a partir da extinção de municípios cuja produção total era baixa e consequente concentração nos municípios/regiões de maior área plantada/colhida.

Por outro lado, algumas regiões que apresentavam baixa produtividade inicialmente, e chegaram a compor um grande *cluster* Baixo-Baixo em 2006, passaram a aumentar os seus níveis de produtividade. O Gráfico 10 mostra alguns desses exemplos.

Gráfico 10 – Produtividade de café de 2002 a 2017 por região geográfica imediata de MG



Fonte: Elaboração própria

Em geral, as tendências observadas a partir de todas as análises feitas foram: aumento da produtividade das regiões que possuem grande área plantada mas produtividade baixa em relação à média do estado, manutenção da produtividade das regiões com grande quantidade de café e que já possuíam níveis elevados de produtividade (em geral, próximo a 2.000 kg/hectare) e abandono da produção por regiões com pouca área plantada, mesmo que possuíssem produtividade elevada.

5 CONSIDERAÇÕES FINAIS

O objetivo do presente trabalho foi analisar a distribuição espacial da produtividade de café em Minas Gerais entre 2002 e 2017 e verificar a existência ou não de padrões de distribuição espacial, por meio da autocorrelação espacial, bem como a existência de agrupamentos locais de distribuição espacial.

Dentro do proposto, foi verificada a existência de autocorrelação espacial, a partir do *I* de Moran, em 9 anos dos 16 analisados. Contudo, a predominância foi na primeira metade do período de análise, quando, em todos os anos o índice se mostrou estatisticamente significativo. Os mapas do *I* de Moran Local, apresentados para a divisão em regiões geográficas imediatas, mostraram que existiam agrupamentos locais de alta produtividade estatisticamente significativos, mas que estes diminuíram ao longo do tempo.

Em geral, identificou-se uma tendência de saída do mercado de regiões com pequena área plantada de café, mesmo possuindo alta produtividade. Além disso, foi identificada a tendência de aumento de produtividade de regiões que possuíam grandes áreas de café plantada/colhida, mas baixa produtividade. Aparentemente, a tendência é de homogeneização da produtividade de café entre as regiões, com valores próximos a 2.000 kg/ha.

Este trabalho pode ser visto como um passo inicial, a partir do qual podem ser incorporadas novas técnicas e variáveis a fim de aprofundar os resultados. Dentre essas, destaca-se a investigação de outras variáveis que possam explicar a existência dos *clusters* encontrados e o uso da análise espacial bivariada. Em complemento à busca de variáveis explicativas podem ser ajustados modelos de regressão espacial, como o Modelo Espacial Autorregressivo Misto (SAR) e o Modelo do Erro Espacial (CAR) e/ou um modelo de convergência de produtividade do café.

REFERÊNCIAS

- ACOMPANHAMENTO DA SAFRA BRASILEIRA: café. Brasília, DF: Companhia Nacional de Abastecimento, v. 5, n. 4, p. 1-84, dez. 2018.
- ALMEIDA, E. **Econometria espacial aplicada**. Campinas: Alínea, 2012.
- ALMEIDA, E. S.; PACHECO, G. O.; PATROCÍNIO, A. P. B.; DIAS, S. M. Produtividade do café em Minas Gerais: uma análise espacial. *In*: CONGRESSO BRASILEIRO DE ECONOMIA E SOCIOLOGIA RURAL, 44., 2006, Fortaleza. **Anais [...]**. Fortaleza: Sociedade Brasileira de Economia e Sociologia Rural, 2006
- ANSELIN, L. Local Indicators of Spatial Autocorrelation - LISA. **Geographical Analysis**, v. 27, n. 2, Apr. 1995.
- ANSELIN, L. **Spatial econometrics: methods and models**. Boston: Kluwer Academic, 1988.
- ANSELIN, L. Interactive techniques and exploratory spatial data analysis. *In*: LONGLEY, P.; GOODCHILD, M.; MAGUIRE, D.; RHIND, D. (ed.). **Geographical information systems: principles, techniques, management and applications**. Nova Iorque: Wiley, 1999. p. 251-264.
- ÁVILA, R. P.; MONASTERIO, L. M. O MAUP e a análise espacial: um estudo de caso para o Rio Grande do Sul (1991-2000). **Revista Análise Econômica**, Porto Alegre, ano 26, n. 49, p. 233-259, mar. 2008.
- ASSUNÇÃO, R. M. **Estatística Espacial com Aplicações em Epidemiologia, Economia e Sociologia**. São Carlos: Associação Brasileira de Estatística, 2001.
- BAILEY, T. C.; GATRELL, A. C. **Interactive spatial data analysis**. Essex: Longman Scientific, 1995.
- CÂMARA, G. *et al.* Análise espacial de áreas. *In*: DRUCK, S.; CARVALHO, M. S.; CÂMARA, G.; MONTEIRO, A. M. V. **Análise espacial de dados geográficos**. Brasília, DF: Empresa Brasileira de Pesquisa Agropecuária Brasília, 2004. v. 2, cap. 5.
- CENTRO DE ESTUDOS AVANÇADOS EM ECONOMIA APLICADA (CEPEA) **PIB do Agronegócio Brasileiro**. Disponível em: <https://www.cepea.esalq.usp.br/br/pib-do-agronegocio-brasileiro.aspx> . Acesso em: 04 nov. 2018.
- CLIFF, A. D.; ORD, J. K.; **Spatial Process: model & applications** London:Taylor & Francis, 1981.
- CRESSIE, N. A. **Statistics for spatial data: Wiley series in probability and mathematical statistics**. New York: John Wiley & Sons, 1993.

GARCIA FILHO, J. **Análise de dados de área aplicada a indicadores de seguridade rural para a cultura de café**. 2018. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras, 2018.

FISCHER, M. M.; GETTIS, A. **Handbook of applied spatial analysis: software tools, methods and applications**. Berlin: Springer, 2009.

FOTHERINGHAM, A. S.; BRUNSDON, C.; CHARLTON, M. **Quantitative geography: perspectives on spatial data analysis**. London: Sage Publications, 2000.

HAINING, R. **Spatial data analysis: theory and practice**. Cambridge, UK: Cambridge University Press, 2003.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **Divisão Regional do Brasil em Regiões Geográficas Intermediárias e Regiões Geográficas Imediatas**: Rio de Janeiro: IBGE, 2017.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). Sistema IBGE de Recuperação Automática (SIDRA). **Tabela 1613**: área destinada à colheita, área colhida, quantidade produzida, rendimento médio e valor da produção das lavouras permanentes. Disponível em: <https://sidra.ibge.gov.br/tabela/1613>. Acesso em: 13 set. 2018.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). Sistema IBGE de Recuperação Automática (SIDRA). **Tabela 5457**: área plantada ou destinada à colheita, área colhida, quantidade produzida, rendimento médio e valor da produção das lavouras temporárias e permanentes. Disponível em: <https://sidra.ibge.gov.br/tabela/5457>. Acesso em: 13 jul. 2019.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **Cidades**. Disponível em: <https://cidades.ibge.gov.br/brasil/mg/panorama>. Acesso em: 14 jul. 2019.

INTERNATIONAL COFFEE ORGANIZATION (ICO). **Total production by all exporting countries**. Disponível em: <http://www.ico.org/historical/1990>

MINAS GERAIS. Secretaria de Estado de Agricultura, Pecuária e Abastecimento de Minas Gerais. **Projeções do agronegócio: Minas Gerais 2017 a 2027: projeções de longo prazo**. 3. ed. [Belo Horizonte: SEAPA, 2017]. Disponível em: http://www.agricultura.mg.gov.br/images/Arq_Relatorios/Publicacoes/projecoes_2017_a_2027.pdf. Acesso em: 9 maio 2019.

NEVES, M. C.; LUIZ, A. J. B. **Distribuição espacial da cultura do café no estado de São Paulo**. Jaguariúna: Embrapa Meio Ambiente, 2006.

PYTHON SOFTWARE FOUNDATION. **The Python Language reference: version 3.7**. c2019 Disponível em: docs.python.org Acesso em: 01 maio 2019.

ROBINSON, W. S. Ecological correlations and the behavior of individuals. **American Sociological Review**, v. 15, n. 3, p. 351-357, 1950.

SANTOS, F. L.; NANTES, J. F. D. Análise da distribuição espacial da cafeicultura no estado do Espírito Santo. *In: ENCONTRO NACIONAL DE ENGENHARIA DE PRODUÇÃO*, 29., 2009, Salvador. **Anais [...]**. Salvador: ENEGEP, 2009.

SCOLFORO, J. R. S.; CARVALHO, L. M. T.; OLIVEIRA, A. D. **Zoneamento ecológico-econômico do estado de Minas Gerais**: componentes geofísico e biológico. Lavras: Editora UFLA, 2008

SILVA, C. N. **Análise de dados de área aplicada a dois indicadores econômicos de mesorregiões do estado de Minas Gerais**. 2010. Dissertação (Mestrado em Estatística e Experimentação Agrícola) - Universidade Federal de Lavras, Lavras, 2010.

SOUZA, R. M.; PEROBELLI, F. S. Diagnóstico espacial da concentração produtiva do café no Brasil no período de 1991 a 2003. **Revista de Economia e Agronegócio**, v. 5, n. 3, p. 353-378, 2007.

TEIXEIRA, R. F. A. P.; BERTELLA, M. A. Distribuição espaço-temporal da produtividade média do café em Minas Gerais: 1997-2006. **Análise Econômica**, Porto Alegre, ano 33, n. 63, p. 275-299, mar. 2015.

TYSZLER, M. **Econometria espacial**: discutindo medidas para a matriz de ponderação espacial. 2006. Dissertação (Mestrado em Administração Pública e Governo) - Fundação Getúlio Vargas, São Paulo, 2006.

APÊNDICE A - Valores do I de Moran com matrizes de distância geográfica

Tabela 1 - I de Moran para a produtividade do café em MG por regiões geográficas imediatas entre 2002 e 2017 a partir de matriz de distância geográfica com os 4 vizinhos mais próximos

ano	I de Moran	valor- p
2002	0.280	0.002
2003	0.331	0.001
2004	0.281	0.001
2005	0.281	0.003
2006	0.275	0.003
2007	0.170	0.012
2008	0.114	0.051
2009	0.144	0.026
2010	0.081	0.099
2011	0.109	0.069
2012	0.135	0.023
2013	0.035	0.276
2014	0.069	0.145
2015	-0.024	0.478
2016	0.138	0.035
2017	0.159	0.022

Fonte: Elaboração própria

Tabela 2 - *I* de Moran para a produtividade do café em MG por regiões geográficas imediatas entre 2002 e 2017 a partir de matriz de distância geográfica com os 5 vizinhos mais próximos

ano	<i>I</i> de Moran	valor- <i>p</i>
2002	0.193	0.007
2003	0.288	0.001
2004	0.251	0.002
2005	0.245	0.001
2006	0.318	0.001
2007	0.143	0.022
2008	0.072	0.099
2009	0.118	0.034
2010	0.027	0.276
2011	0.082	0.079
2012	0.077	0.106
2013	0.012	0.317
2014	0.009	0.372
2015	-0.089	0.156
2016	0.092	0.081
2017	0.079	0.095

Fonte: Elaboração própria

Tabela 3 - *I* de Moran para a produtividade do café em MG por regiões geográficas imediatas entre 2002 e 2017 a partir de matriz de distância geográfica com os 6 vizinhos mais próximos

ano	<i>I</i> de Moran	valor- <i>p</i>
2002	0.166	0.006
2003	0.233	0.001
2004	0.207	0.002
2005	0.201	0.001
2006	0.291	0.001
2007	0.118	0.021
2008	0.060	0.101
2009	0.088	0.058
2010	0.003	0.362
2011	0.047	0.125
2012	0.043	0.159
2013	-0.024	0.435
2014	-0.002	0.399
2015	-0.096	0.081
2016	0.089	0.072
2017	0.069	0.100

Fonte: Elaboração própria

Tabela 4 - *I* de Moran para a produtividade do café em MG por regiões geográficas imediatas entre 2002 e 2017 a partir de matriz de distância geográfica com os 7 vizinhos mais próximos

ano	<i>I</i> de Moran	valor- <i>p</i>
2002	0.194	0.002
2003	0.229	0.002
2004	0.212	0.001
2005	0.210	0.002
2006	0.294	0.001
2007	0.131	0.018
2008	0.081	0.042
2009	0.098	0.029
2010	0.024	0.216
2011	0.059	0.098
2012	0.062	0.087
2013	-0.008	0.440
2014	0.021	0.263
2015	-0.080	0.117
2016	0.114	0.028
2017	0.085	0.042

Fonte: Elaboração própria