

Análise espacial de área aplicada a produtividade de soja na região oeste do Paraná utilizando o software R

RESUMO

Rodolfo Seffrin

seffrinrodolfo@gmail.com

orcid.org/0000-0003-1133-2986

Universidade Tecnológica Federal do Paraná (UTFPR), Medianeira, Paraná, Brasil.

Everton Coimbra de Araújo

everton@utfpr.edu.br

orcid.org/0000-0002-2498-1224

Universidade Tecnológica Federal do Paraná (UTFPR), Medianeira, Paraná, Brasil.

Claudio Leones Bazzi

bazzi@utfpr.edu.br

orcid.org/0000-0002-7726-144X

Universidade Tecnológica Federal do Paraná (UTFPR), Medianeira, Paraná, Brasil.

Este trabalho apresenta um estudo referente à produtividade de soja na região oeste do Paraná, demonstrando como aplicar algumas técnicas de análise espacial de área utilizando o *software* R. O estudo foi realizado com dados de produtividade de soja dos anos 2007/2008 para 48 municípios. Para a verificação de correlação e autocorreção espacial foram utilizadas as técnicas de Índice de Moran Global, Índice de Moran Local e Índice de Geary, todos esses procedimentos foram realizados com as bibliotecas *spdep*, *sp*, *maptools* e *rcolorbrewer* do *software* R. Por meio deste estudo foi possível demonstrar a aplicabilidade do *software* R para análise espacial de área.

PALAVRAS-CHAVE: Autocorrelação Espacial. Matriz de Vizinhaça. Índice de Moran. Índice de Geary

INTRODUÇÃO

Compreender fenômenos resultantes da distribuição espacial de dados no espaço é um grande desafio para diversas áreas do conhecimento. Vários estudos estão sendo realizados para se conseguir adquirir algum padrão que esteja associado à distribuição espacial, seja na saúde, em ambiente, em geologia ou na agricultura, entre outras áreas. Portanto, tais estudos estão se tornando cada vez mais comuns em consequência da disponibilidade dos Sistemas de Informação Geográfica (SIG) que, em alguns casos, se encontram com baixo custo e possuem uma interface amigável. Assim, esses sistemas fazem com que seja possível a visualização espacial de variáveis, como: população de indivíduos, índices de qualidade de vida e vendas de empresas em uma determinada região, sendo ilustradas por mapas. Com isso, é necessário se ter um banco de dados com informações de uma base geográfica (como um mapa de municípios), e em torno disso o SIG possibilita visualizar algum padrão espacial do fenômeno por meio de mapas. (DRUCK et al., 2004)

Neste contexto, a estatística espacial situa-se em um cenário amplo, a análise espacial de dados. O objetivo é determinar a propriedade e relacionamentos, levando em consideração a localização espacial do fenômeno em estudo. Assim, conceitos como autocorrelação espacial e dependência espacial são de suma importância. A dependência espacial é o fato do qual a grande maioria das ocorrências naturais ou sociais apresentam dependência em relação à distância (DRUCK et al., 2004). Assim sendo, o objetivo é analisar como a dependência espacial varia perante os valores das amostras dos seus vizinhos (FOOK, 2005).

A Estatística espacial de Área (EEA) é um recurso estatístico que utiliza a referência geográfica no modelo, como as coordenadas espaciais no processo de coleta, descrição e análise dos dados. Portanto, o foco está concentrado nos processos que acontecem no espaço, e as técnicas estatísticas procuram compreender e analisar o comportamento desses processos (ARAÚJO et al., 2014).

A escolha referente a cultura da soja se deve ao fato da sua importância no cenário brasileiro. Desde 1976, ela vem se destacando no país, tornando o Brasil o segundo maior produtor mundial de soja, perdendo apenas para os Estados Unidos. Em 2010, o Brasil foi responsável por 26,3% da produção mundial de soja (CONAB, 2010), valor de 67,5 milhões toneladas, sendo cultivada em uma área referente a 24,3 milhões de hectares. Em relação à economia, a soja propiciou aproximadamente 9% das exportações brasileiras, girando um valor de R\$ 17,5 bilhões. Outra informação importante sobre o PIB, é que a *comodity* respondeu por 5,6% de um total de R\$ 821,8 Bilhões, correspondendo à porcentagem de 1,25% de todo PIB nacional. Em 2012/13 o Brasil se tornou o maior exportador de soja do mundo, exportando 42 milhões de toneladas de grãos, já para o ciclo 2013/14 a perspectiva é que se consolide o maior produtor de soja do planeta (SEAB, 2013).

Foi escolhido o *Software R* para realizar este estudo por se tratar de uma linguagem de programação voltada para um ambiente de computação estatística e gráfica, e por sua licença ser gratuita, além de ter uma grande abrangência, podendo ser utilizado por diversos sistemas operacionais. O R possui vários

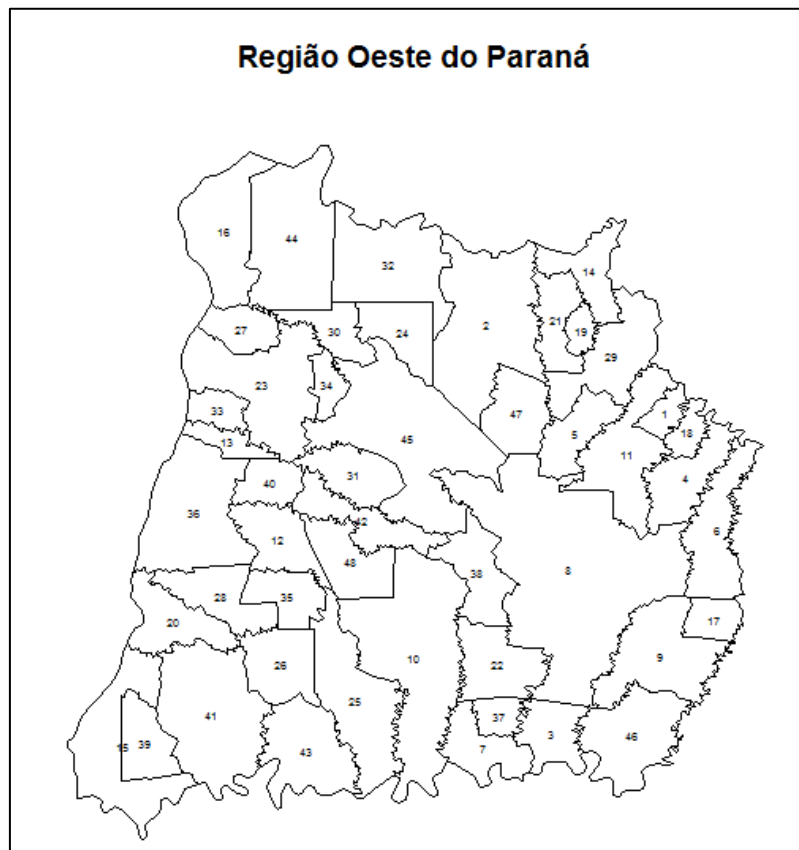
pacotes que podem ser estendidos, como por exemplo: bibliotecas para funções específicas ou áreas de estudos específicas (SILVA; DINIZ; BORTOLUZZI, 2009).

Portando, o trabalho tem como objetivo demonstrar as principais funcionalidades do *software* R aplicadas a estatística espacial de área, utilizando como variável de estudo a produtividade de soja da região oeste do Paraná. As técnicas aplicadas foram: Índice de Moran Global, Índice de Moran Local e Índice de Geary para verificar a existência de autocorrelação espacial na área estudada.

MATERIAIS E MÉTODOS

A área de estudo deste trabalho abrange 48 municípios da região oeste do Estado do Paraná, conforme Figura 1. Como variável de estudo foi utilizado a produtividade de soja dos anos safra 2007/2008.

Figura 1 – Municípios da Região Oeste do Paraná



(1) Anahy, (2) Assis Chateaubriand, (3) Boa Vista da Aparecida, (4) Braganey, (5) Cafelândia, (6) Campo Bonito, (7) Capitão Leônidas Marques, (8) Cascavel, (9) Catanduvas, (10) Céu Azul, (11) Corbélia, (12) Diamante D'Oeste, (13) Entre Rios do Oeste, (14) Formosa do Oeste, (15) Foz do Iguaçu, (16) Guaíra, (17) Ibema, (18) Iguatu, (19) Iracema do Oeste, (20) Itaipulândia, (21) Jesuítas, (22) Lindoeste, (23) Marechal Cândido Rondon, (24) Maripá, (25) Matelândia, (26) Medianeira, (27) Mercedes, (28) Missal, (29) Nova Aurora, (30) Nova Santa Rosa, (31) Ouro Verde do Oeste, (32) Palotina, (33) Pato Bragado, (34) Quatro Pontes, (35) Ramilândia, (36) Santa Helena, (37) Santa Lúcia, (38) Santa Tereza do Oeste, (39) Santa Terezinha de Itaipu, (40) São José das Palmeiras, (41) São Miguel do Iguaçu, (42) São Pedro do Iguaçu, (43) Serranópolis do Iguaçu, (44) Terra Roxa, (45) Toledo, (46) Três Barras do Paraná, (47) Tupãssi e (48) Vera Cruz do Oeste. Fonte: Autoria Própria (2017).

As informações referentes à produtividade em (t ha⁻¹) foram fornecidas pela SEAB (2010). Para desenvolver a análise espacial de área foi utilizado o *software* R na versão 3.1.1(R, 2013).

Para o cálculo de autocorrelação espacial foram utilizados os índices de Moran Global e Local e Índice de Geary, com a finalidade de demonstrar como aplicar esses recursos com o *software* R e concluir se existe autocorrelação espacial para a variável produtividade na região oeste do Paraná.

O índice I de Moran é o método mais usado para o cálculo de autocorrelação espacial. O indicador verifica se existe correlação espacial de uma variável de interesse. (ALBUQUERQUE, 2009)

A equação 1 é a fórmula que calcula a autocorrelação espacial global das variáveis, verificando se os dados estão autocorrelacionados no espaço.

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n z_i z_j w_{ij}}{S_0 \sum_{i=1}^n z_i^2} \quad (1)$$

onde n é o número de municípios (n polígonos); $z_i = (x_i - \bar{x})$ e $z_j = (x_j - \bar{x})$ para $i \neq j = 1 \dots n$, que se referem aos valores analisados das populações i e j centradas na média \bar{x} da variável de interesse; w_{ij} é o elemento da matriz de proximidade W , $n \times n$, pelo qual indica a relação espacial para as n populações e S_0 é definida pela Equação (2):

$$S_0 = \sum_{i=1}^n \sum_{j=1}^n w_{ij} \quad (2)$$

Segundo Druck et al. (2004) e Anselin et al. (2007), cada elemento w_{ij} da matriz de proximidade espacial W , representa uma medida de proximidade entre os municípios (polígonos) A_i e A_j , a qual pode ser calculada a partir de um dos seguintes critérios:

- Regra da Distância entre Centroides: $w_{ij} = 1$, quando o centroide de A_i está a uma determinada distância de A_j ; senão $w_{ij} = 0$; para $i \neq j = 1, \dots, n$;
- Regra de contiguidade (torre, rainha e bispo): $w_{ij} = 1$, quando A_i compartilha um lado comum com A_j , senão $w_{ij} = 0$; para $i \neq j = 1, \dots, n$;
- Regra de número de vizinhos mais próximos: $w_{ij} = A_{ij}/A_i$, onde A_{ij} é o comprimento da fronteira entre A_i e A_j e A_i é o perímetro de A_i ; para $i \neq j = 1, \dots, n$;

O índice I de Moran está entre -1 e +1, fornecendo uma medida geral de associação global, testando a hipótese nula de independência espacial ($I=0$) e a hipótese alternativa para a dependência espacial ($I \neq 0$). Em outras palavras, verifica se as áreas vizinhas mostram uma semelhança maior comparando a variável de interesse com um padrão aleatório. Portanto, o indicador é uma das

maneiras de analisar se as regiões possuem algum padrão espacial de agrupamento. (ALBUQUERQUE, 2009).

Valores negativos próximos de zero apontam a ausência de autocorrelação espacial significativa para os valores da variável de interesse na região analisada. Quando os valores são positivos e significativos apontam autocorrelação espacial positiva, ou seja, o valor da variável de interesse tende a ser semelhante nos vizinhos mais próximos na região estudada. Os valores que são negativos e significativos apontam para uma autocorrelação negativa, indicando que a variável de interesse tende a ter valores diferentes para os locais mais próximos na região analisada (KHAN,2010).

Os valores globais podem ocultar os padrões locais de autocorrelação espacial, sendo necessário analisar com maior detalhamento a dependência espacial existente. Por este motivo que existem os indicadores locais que permitem um maior detalhamento para cada região, tornando possível, a identificação de diferentes padrões espaciais, como agrupamentos com valores parecidos (*clusters*) ou divergentes (*outliers*). Ao estudar os padrões locais para autocorrelação, é possível ter-se informações sobre a presença de heterogeneidade espacial. Podendo então, descobrir associações espaciais do tipo alto-alto (AA) ou baixo-baixo (BB), e também do tipo alto-baixo (AB) ou baixo-alto(BA) (ALBUQUERQUE, 2009).

Para uma melhor contextualização, as associações do tipo AA, representam o agrupamento dos municípios com uma alta produtividade de soja, já as associações BB representam os municípios com uma baixa produtividade de soja, e os agrupamentos AB e BA são os municípios com correlação negativa (*outliers*), formando grupos locais com valores divergentes (ALBUQUERQUE, 2009).

Segundo Celebioglu e Dall’erba (2009), a estatística *LISA*, ou índice *I* de Moran local, pode ser explicado pela Equação (3):

$$I_i = \frac{x_i - \mu}{\sigma_0^2} \sum_{j=1}^n w_{ij}(x_j - \mu), \quad i = 1, \dots, n \quad (3)$$

em que, σ_0^2 a variância populacional da variável em estudo dos n municípios; x_i é a observação de uma variável de interesse no município i para $i = 1, \dots, n$, μ é a média dos n municípios (populações).

Outro teste que pode ser aplicado para verificar a existência de autocorrelação global é o índice de Geary, a diferença deste método com o Moran Global, é que ele considera a diferença entre os pares, já o Moran utiliza a diferença entre cada ponto e a média global (DRUCK et al. 2004). O índice de Geary pode ser explicado pela equação (4)

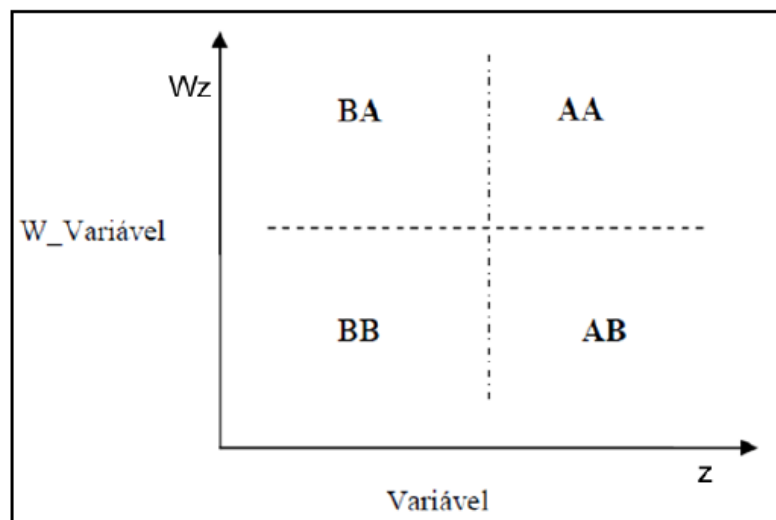
$$GR = \left[(N-1) \sum_{I=1}^N \sum_{J=1}^N C_{IJ} (X_I - X_J)^2 \right] / \left[2 \left(\sum_{I=1}^N \sum_{J=1}^N C_{IJ} \right) \sum_{I=1}^N (X_I - \bar{X})^2 \right] \quad (4)$$

Em que X_i e X_j são os valores da variável de interesse; \bar{X} é a média da variável de interesse; N é o número de cidades; e C_{ij} recebe valor um 1 caso X_i e X_j sejam vizinhos, senão recebe 0.

Quando aplicado o Índice de Moran global é importante realizar uma análise gráfica da autocorrelação espacial por meio do diagrama de dispersão de Moran (Scatterplot Moran). Segundo Almeida et al. (2005), o diagrama de dispersão de Moran é uma representação do coeficiente de regressão linear por Mínimos Quadrados Ordinários (MQO), mediante um gráfico de duas variáveis z e Wz , na qual o coeficiente da inclinação da curva de regressão é dado pela estatística I de Moran. A inclinação da curva é obtida pela regressão de Wz contra z , e essa inclinação fornece o grau de ajustamento.

Conforme pode ser visto na Figura 2, o diagrama de dispersão de Moran é uma maneira de visualizar o resultado calculado no indicador global da autocorrelação espacial. No eixo vertical é mostrada a defasagem espacial da variável de interesse, e já no eixo horizontal resulta no valor da variável de interesse. O diagrama está dividido em quatro partes: Alto-Alto(AA), Baixo-Baixo(BB), Alto-Baixo(AB) e Baixo-Alto (BA) (ARAÚJO et al., 2014).

Figura 2 - Estrutura do Diagrama de Dispersão onde W _Variável caracteriza a variável de interesse defasada espacialmente



Fonte: Araújo (2012).

Os quadrantes que apresentam os valores AA e BB significam que essas áreas possuem valores altos (acima da média) e/ou valores baixos da variável de interesse, rodeadas por áreas que apresentam valores também altos e/ou baixos. Os locais que resultam no quadrante BA representam áreas com valor baixo rodeada de outras áreas com valores altos, já o quadrante AB é ao contrário, são regiões com valores altos próximas de locais com valores baixos.

Outro procedimento importante para a análise de dados de área é a construção de uma matriz de vizinhança, que também pode ser chamada de matriz de distância, matriz de conectividade ou matriz de proximidade. Esta matriz de vizinhança aponta a relação espacial de cada área do estudo, podendo ser ligada pela lista de vizinhos de cada município, ou pelas distâncias dos

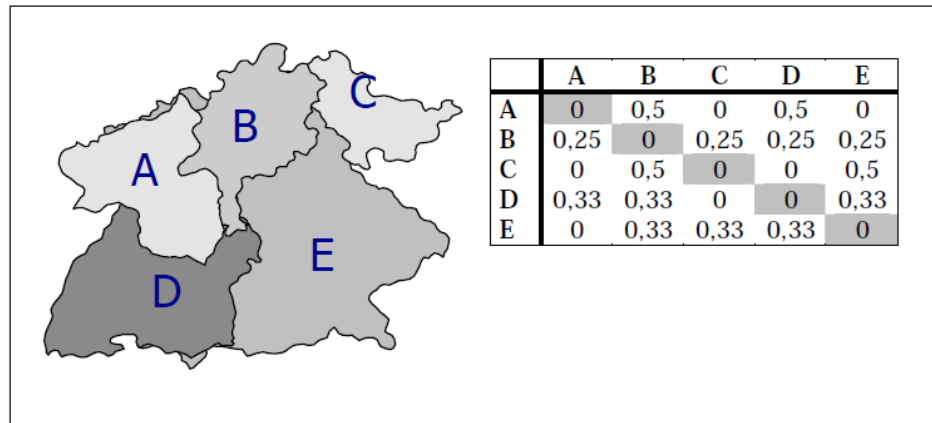
municípios ligados pelas estradas, e também pela conectividade ponderada pelo comprimento da fronteira comum. (ARAÚJO et al., 2014)

Por exemplo, para calcular um conjunto de n áreas $\{A_1, \dots, A_n\}$, pode-se construir a matriz $W^{(1)}(n \times n)$, em que cada um dos elementos W_{ij} é uma medida de proximidade entre A_i e A_j . Então, esta medida é calculada seguindo os critérios:

- $w_{ij}=1$, quando o centroide de A_i está a uma determinada distância de A_j ; senão $w_{ij}=0$;
- $w_{ij}=1$, quando A_i compartilha um lado comum com A_j , senão $w_{ij}=0$;
- $w_{ij}=l_{ij}/l_i$, onde l_{ij} é o comprimento da fronteira entre A_i e A_j e l_i é o perímetro de A_i ;

Como a matriz de vizinhança é usada para cálculos de indicadores quando está na fase de análise exploratória, é importante realizar a normalização das suas linhas, fazendo com que as somas dos pesos das linhas sejam iguais a 1. Isto simplifica os cálculos que são feitos pelos índices de autocorrelação espacial. (ARAÚJO et al., 2014). A Figura 3 mostra um exemplo de matriz de vizinhança onde os valores foram normalizados:

Figura 3 - Matriz de proximidade espacial de primeira ordem, normalizada pelas linhas.



Fonte: Carvalho et al. (2004).

É de suma importância convencionar as formas de vizinhança nas matrizes de vizinhança que consideram a contiguidade. Sendo assim, as regras seguem os movimentos de peças que existem no jogo de xadrez, tais como a rainha (*Queen*), a torre (*Rook*) e o bispo (*Bishop*) (ARAÚJO et al., 2014). A Figura 4 ilustra um exemplo dessas regras para identificar os vizinhos na área J.

Figura 4 - Representação dos tipos de contiguidade entre áreas. (a) Contiguidade *Queen* (rainha), (b) Contiguidade *Rook* (torre) e (c) Contiguidade *Bishop* (bispo)

A	B	C	D
E	F	G	H
I	J	K	L
M	N	O	P

(a)

A	B	C	D
E	F	G	H
I	J	K	L
M	N	O	P

(b)

A	B	C	D
E	F	G	H
I	J	K	L
M	N	O	P

(c)

Fonte: Araújo (2012).

A Figura 4a mostra a regra *Queen*, aonde todas as áreas que têm intersecção não-nula com a área J serão vizinhas de J. Já na Figura 4b detalha-se a regra *Rook*, que considera como vizinho apenas as áreas com lado em comum da área J. E por último, a Figura 4c que demonstra a regra *Bishop*, que usa como critério de vizinhança as áreas que se localizam nas diagonais da área J.

Para ilustrar o agrupamento das regiões em relação a variável produtividade ($t\ ha^{-1}$) de soja, utilizou-se o teste de *Tukey*, que calcula os resultados estatísticos: *minimum*, *lower-hinge*, *median*, *upper-hinge*, *maximum*. Este teste está disponível na função *fivenum()* do pacote *stats* no *software R*. (R, 2013).

RESULTADOS E DISCUSSÕES

Para realizar os procedimentos deste estudo foram utilizados os seguintes pacotes do *software R*:

- *Spdep*: Calcular os índices de Moran Global e Local, índice de Geary, índice de Moran Bayesiano Empírico;
- *Maptools*: Construir os mapas da região oeste do Paraná e carregar o arquivo *shape* referente a produtividade de soja;
- *Sp*: Foi utilizado a função *splot* para gerar os mapas referente ao cálculo do índice de Moran local;
- *RColorBrewer*: Criar as paletas de cores para os mapas temáticos;

É necessário realizar a instalação de cada um deles utilizando a função `install.packages("<nome do pacote>")`, informando o nome do pacote desejado. Por padrão serão instalados os pacotes dependentes. Concluída a instalação é possível realizar todos os procedimentos deste estudo. Portanto, a fim de demonstrar a aplicabilidade do *software R* na estatística espacial de área, o referente trabalho foi desenvolvido seguindo as seguintes etapas:

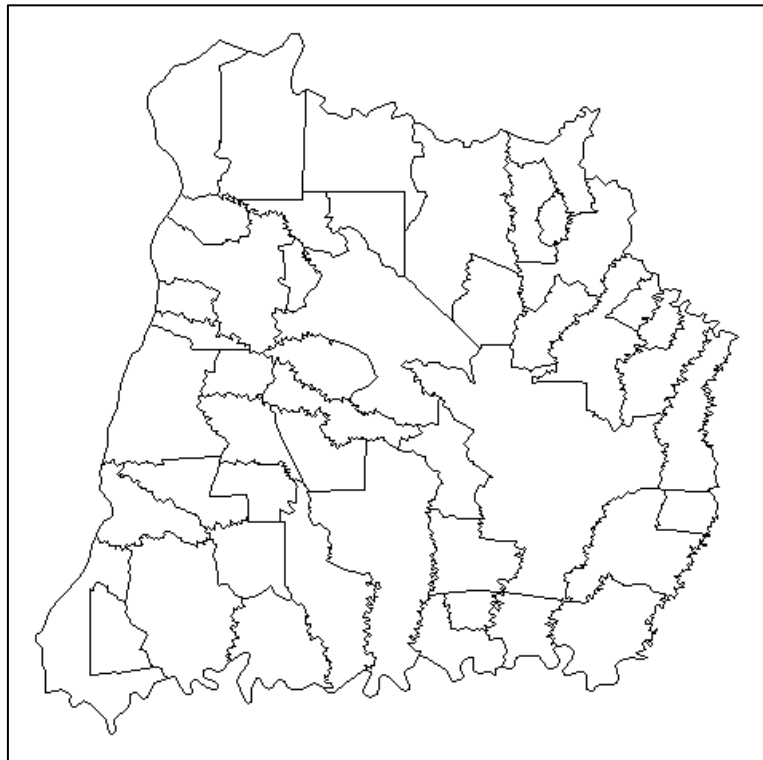
- Carregar o arquivo *shape* e demonstrar alguns procedimentos para gerar o mapa do *shape* carregado;
- Demonstrar a relação de vizinhança da região Oeste do Paraná por meio das técnicas: contiguidade *queen*, vizinhos mais próximos e distância entre centroides;
- Calcular o Índice de Moran Global e Índice de Geary;
- Gerar o *scatterplot* Moran Global e construir o mapa com as cidades que resultaram com um valor divergente das suas vizinhas (*outliers*);

- Calcular o Índice de Moran Local e construir os mapas com os padrões descobertos localmente;

Depois de instalados todos os pacotes descritos acima, é necessário carregá-los para utilizar suas funções, para isto, é preciso chamar a função *require* (pacote *spdep*) ou *library*(pacote *spdep*) passando como parâmetro o nome do pacote. Após isso, o primeiro passo é visualizar o mapa da região oeste do Paraná, portanto, é necessário carregar o arquivo *shape* que contém as cidades por meio da função *readShapePoly()* do pacote *mapproj*, e exibi-lo na tela (Figura 5) invocando a função *plot*, conforme script abaixo:

```
> prodMap<- readShapePoly("C:\\produtividade\\dados.shp"[1])
> plot(prodMap)
> plot(prodMap, col="red") # Construir o mapa com a cor vermelha
```

Figura 5 - Mapa dos municípios da região Oeste do Paraná



Fonte: Autoria própria (2017).

Para calcular a estatística descritiva da variável produtividade (Tabela 1), pode-se utilizar a função *summary()*, passando como parâmetro a variável que representa o arquivo *shape* carregado anteriormente, junto com o nome da coluna que identifica os dados referentes à produção de soja dos anos 2007/2008:

```
> summary(prodMap$PRODU07_08)
```

Tabela 1 - Estatística Descritiva para a produtividade de soja (t ha⁻¹)

Minimo	Primeiro Quartil	Mediana	Média	Terceiro Quartil	Máximo
2.500	3.188	3.286	3.273	3.471	3.699

Fonte: Autoria própria (2017).

Como pode ser visto na tabela 1, a média de produtividade de soja na região oeste do Paraná foi de 3.273 (t ha⁻¹). Araújo et. al. (2013) realizou um estudo nos anos de 2007/2008 para classificar áreas associadas a produtividade de soja na região oeste do Paraná utilizando a lógica Fuzzy, os valores da estatística descritiva gerados por Araújo et. al. (2013) foram semelhantes aos resultados da tabela 1.

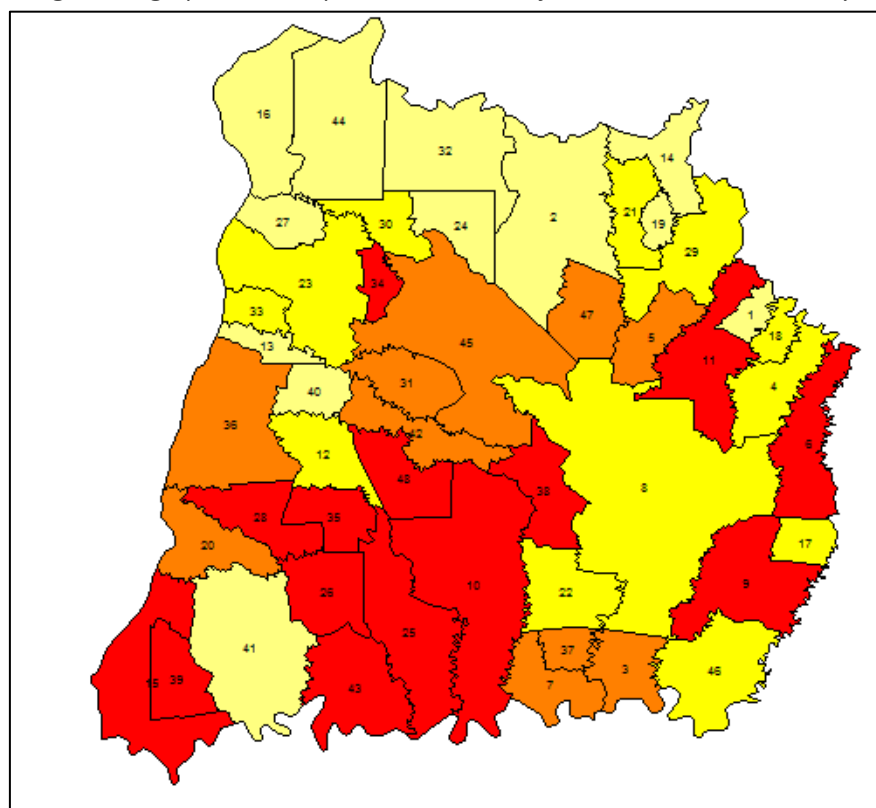
Foi utilizado o teste de *Tukey* para calcular os resultados estatísticos: *minimum*, *lower-hinge*, *median*, *upper-hinge*, *maximum*. Esses resultados foram usados para construir o mapa de agrupamento da área em estudo. Desta maneira, é possível visualizar se houve semelhança na produtividade de soja nas regiões mais próximas.

Conforme a figura 6, percebe-se que algumas áreas ocorreram agrupamento, ou seja, isto significa que segundo teste de *Tukey*, cidades próximas obtiveram produtividade semelhante.

O cálculo do teste de *Tukey* foi realizado em meio a função *fivenum()* do pacote *stats*. O script utilizado para gerar o mapa:

```
> res<- prodMap$PRODU07_08  
> brks<- round(fivenum(prodMap$PRODU07_08), digits=3)  
> cols<- rev(heat.colors(4))  
> plot(prodMap, col=cols[findInterval(res, brks, all.inside=TRUE)])  
> text(coordinates(prodMap), labels=as.character(prodMap$POLY_IDB), cex=0.4)
```

Figura 6 -Agrupamento da produtividade de soja utilizando o teste de *Tukey*



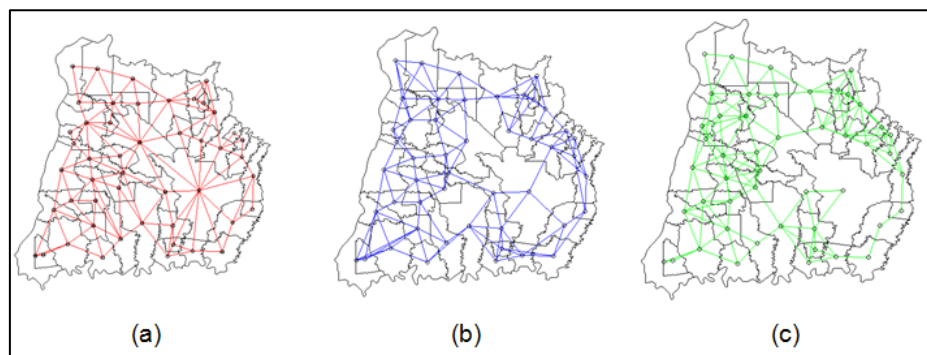
Fonte: Autoria própria (2017).

A função *findInterval()*, do pacote *base*, retorna em qual intervalo (entre os 5 valores do teste de *tukey* realizado pela função *fivenum*) o polígono se encontra.

Esses intervalos são entre os valores *minimum lower-hinge*, *lower-hinge* e *median*, *mediane upper-hinge*, *upper-hinge* e *maximum*. Para identificar cada polígono com seu respectivo id, foi utilizada a função *text()*, passando como parâmetro o centroide da cidade e o valor desejado como descrição.

Uma técnica importante para a análise exploratória dos dados e também para utilizar no cálculo do índice de autocorrelação espacial é a matriz de vizinhança. Para a área de estudo foi realizado três cálculos, utilizando contingência *queen*, distância entre centroides e vizinho mais próximo para a matriz de proximidade *W*. Araújo et. al. (2014) também utilizou em seu estudo essas matrizes de vizinhança para calcular os índices de Moran global e local. A Figura 7 ilustra a conectividade com os três métodos.

Figura 7- Matriz de conectividade: Contiguidade *queen* (a), Vizinho mais próximo (b), Distância entre centroide (c)



Fonte: Autoria própria (2017).

Pode-se perceber a diferença entre as conexões dos polígonos nos três mapas, o primeiro utilizando a regra *queens*, considerou como vizinho apenas as áreas divisas de cada cidade, desconsiderando qualquer critério de distância. Já a regra do vizinho mais próximo, foi considerada os quatro polígonos mais próximos de cada município. O último critério foi à distância entre os centroides, neste teste utilizou-se um peso as distâncias entre as duas áreas mais próximas, caso a distância dos polígonos sejam menores que o peso permitido, são considerados vizinhos, portanto essa regra considera a distância entre as áreas entre os dois vizinhos mais próximos.

Para a contiguidade *queen* foi executado o seguinte *scripts*:

```
> prod.nb<- poly2nb(prodMap,queen = TRUE)
> plot(prodMap)
> plot(prod.nb, coordinates(prodMap), add=TRUE, col="red")
> text(coordinates(prodMap), labels=as.character(prodMap$POLY_IDB), cex=0.4)
```

O cálculo do vizinho mais próximo foi desenvolvido conforme *script*:

```
> coordprodMap<- coordinates(prodMap)
> prodMapIDs<-row.names(as(prodMap, "data.frame"))
> prodMap_kn1<-knn2nb(knearneigh(coordprodMap,k=1), row.names= prodMapIDs)
> prodMap_kn2<-knn2nb(knearneigh(coordprodMap,k=2), row.names= prodMapIDs)
> prodMap_kn4<-knn2nb(knearneigh(coordprodMap,k=4),row.names= prodMapIDs)
> plot(prodMap)
> plot(prodMap_kn4, coordprodMap, add=T,col="blue")
```

Para calcular a distância baseada no centroide foi executado o seguinte *script*:

```
> coordprodMap<- coordinates(prodMap)
> prodMapdist<-unlist(nbdists(prodMap_kn2, coordprodMap))
> summary(prodMapdist)
> max_k1<-max(prodMapdist)
> prod_kd4<-dnearneigh(coordprodMap, d1=0, d2=max_k1, row.names=
prodMapIDs)
> plot(prodMap)
> plot(prod_kd4, coordprodMap, add=T,col="green")
```

Para verificar as informações geradas de cada cálculo é necessário executar a função *summary(prod.nb)*, passando como parâmetro o objeto da matriz de vizinhança.

Foram utilizados para a verificação de autocorrelação espacial global os índices de Moran e Geary, utilizando como variável de interesse os dados de produtividade de soja dos anos 2007 e 2008. Wrublack (2013) utilizou o índice de Moran global para verificar a dependência espacial da produtividade de Canola no estado do Paraná, o autor concluiu que a cultura também apresenta autocorrelação espacial significativa a 5%, Bohórquez (2011) usou o índice de Moran para identificar áreas com possíveis ocorrências de queimadas que ocasionaram desmatamento no Parque Nacional de Tuparro na Colombia. Zhang (2008) utilizou o índice de Moran para verificar a autocorrelação espacial da presença de carbono orgânico no solo no sudoeste da Irlanda, no ano de 2004 Zhang investigou os pontos críticos de contaminação do solo por presença de Chumbo (Pb) na cidade de Galway na Irlanda. Araújo (2014) realizou um estudo para estimar a produtividade de soja na região oeste do Paraná e utilizou o índice de Moran para verificar a autocorrelação espacial da produtividade de soja nos anos de 2007/2008. Portanto, o índice de Moran é utilizado em diversos estudos que envolvem a descoberta de algum padrão espacial que está relacionado com a localização geográfica dos eventos nas áreas de interesse.

Para calcular o índice de autocorrelação espacial global de Moran e Geary é preciso utilizar a função *moran.test()* e *geary.test()* do pacote *spdep*, passando como argumentos os dados da produtividade e matriz de conectividade desejada, conforme *script*:

```
> moran.test(prodMap$PRODU07_08, nb2listw(prod_kd4))
> geary.test(prodMap$PRODU07_08, nb2listw(prod_kd4))
```

Na Tabela 1, são apresentados os índices de autocorrelação espacial Global de Moran e Geary para a variável produtividade, e o nível de significância (entre parênteses), segundo os critérios de contingência *queen*, distância entre centroides e vizinho mais próximo para a matriz de proximidade *W*. Pode-se verificar que os níveis descritivos (*p*-valor) são menores que 0,05 (nível de significância), concluindo que todas as variáveis têm autocorrelação espacial significativa a 5% de probabilidade.

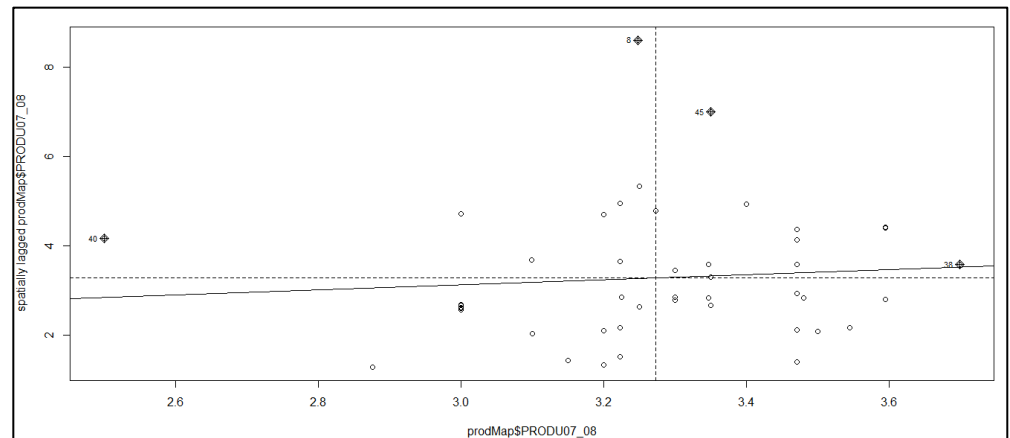
Tabela 2 - Teste de autocorrelação espacial global Moran e Geary

Índice	Contiguidade (<i>queen</i>)	Distância entre centróides	Vizinhos mais próximos
moran.test()	0.3112 (0.001)	0.2200 (0.001)	0.3051 (0.001)
geary.test()	0.6570 (0.001)	0.8096 (0.001)	0.6665 (0.001)

Fonte: Autoria própria (2017).

Um processo importante para se realizar é a geração do diagrama de dispersão de Moran, conhecido como *scatterplot* Moran Global, ele mostra a defasagem espacial da variável de interesse (ou seja, a média do atributo nos vizinhos) no eixo vertical e o valor da variável de interesse no eixo horizontal. O diagrama da Figura 8 foi construído baseado no cálculo da matriz de vizinhança contiguidade *queen*.

Figura 8 - *Scatterplot* Moran Global

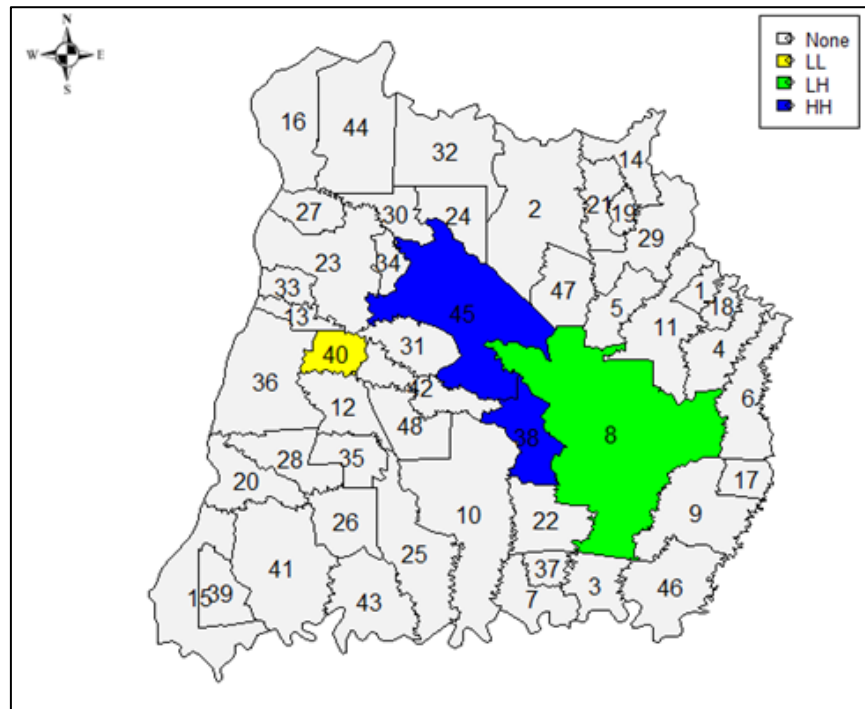


Fonte: Autoria própria (2017).

Pode-se verificar que os municípios de São José dos Palmares (40) e Cascavel (8), se distanciaram muito do restante, isso significa que a produtividade de soja dessas cidades é divergente para as áreas próximas a elas, tornando-se assim pontos discrepantes (outliers) da região em estudo. Os municípios de Toledo (45) e Santa Tereza do Oeste (38) apresentaram alta produtividade de soja, formando um agrupamento do tipo HH (ALTO – ALTO).

Para melhor destaque dos outliers e regiões de agrupamento do tipo HH (ALTO - ALTO), foi construído um mapa ilustrando esses municípios, conforme Figura 9.

Figura 9—Municípios que apresentaram características diferentes de seus vizinhos (outliers)



Fonte: Autoria própria (2017).

Para gerar o diagrama de dispersão de Moran foi utilizado a função *moran.plot()* do pacote *spdep*, passando como parâmetro os dados de produtividade e a matriz de vizinhança contiguidade *queen*, conforme script:

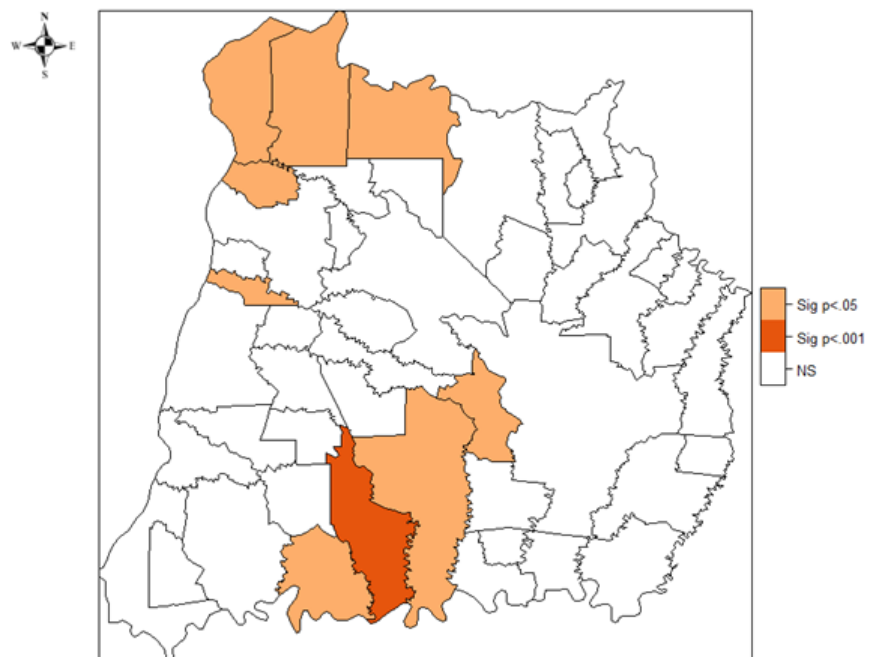
```
> moran.plot(prodMap$PRODU07_08, listw=nb2listw(prodMap.nb,style="C"))
```

O mapa destacando os pontos discrepantes da região em estudo foi construído utilizando o seguinte *script*:

```
> nci<- moran.plot(prodMap$PRODU07_08, labels=prodMap$POLY_IDB,
listw=nb2listw(prodMap.nb,style="C"))
> infl<- apply(nci$is.inf, 1, any)
> x <- prodMap$PRODU07_08
> lhx<- cut(x, breaks=c(min(x), mean(x), max(x)), labels=c("L", "H"),
include.lowest=TRUE)
> wx<- lag(nb2listw(prodMap.nb), prodMap$PRODU07_08)
> lhw<- cut(wx, breaks=c(min(wx), mean(wx), max(wx)), labels=c("L", "H"),
include.lowest=TRUE)
> lhl<- interaction(lhx, lhw, infl, drop=TRUE)
> cols<- rep(1, length(lhl))
> cols[lhl == "L.L.TRUE"] <- 2
> cols[lhl == "L.H.TRUE"] <- 3
> cols[lhl == "H.H.TRUE"] <- 4
> grey.colors(4, 0.95, 0.55, 2.2)
> plot(prodMap, col=grey.colors(4, 0.95, 0.55, 2.2)[cols])
> legend("topright", legend=c("None", "LL", "LH", "HH"), fill=grey.colors(4, 0.95, 0.55,
2.2), bty="n", cex=0.8, y.intersp=0.8)
> text(coordinates(prodMap), labels=as.character(prodMap$POLY_IDB), cex=1.0)
```

Conforme Druck et al. (2004), para um grande número de áreas, como neste caso com 48 municípios, é importante utilizar indicadores de associação espacial local que permitam identificar melhor os agrupamentos. Para o Índice de Moran Local (LISA) foram classificados os municípios em função dos seus níveis de significância. Na Figura 10, observa-se os municípios não significativos ao nível de 5% de probabilidade (sem cor). Já os municípios que apresentaram LISA significativo são exibidos com diferentes tons de marrom, tanto a 5% significância como a 1% de significância, e são os que apresentam uma característica própria, que merecem uma análise detalhada. Para calcular o Índice de Moran Local foi utilizado a matriz de vizinhança baseada no vizinho mais próximo.

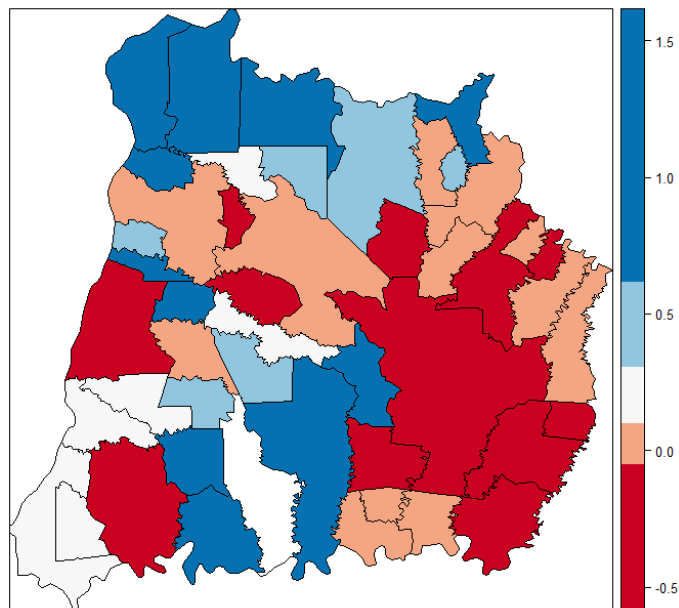
Figura 10 - Indicador local de autocorrelação espacial (LISA) para a variável Produtividade de Soja.



Fonte: Autoria própria (2017).

Outra amostragem importante é a ilustração, por meio de um mapa, como ficou calculado o Índice de Moran local para cada município da região. A Figura 11 detalha o cálculo para a área em estudo.

Figura 11 - Cálculo do índice de Moran local



Fonte: Autoria própria (2017).

O cálculo do índice de Moran local foi gerado por meio da função *localmoran()* do pacote *spdep*. Foi passado como argumento os dados da produtividade de soja e o cálculo da matriz de vizinhança baseado no vizinho mais próximo, conforme *script*:

```
> local.mi.prod<- localmoran(prodMap$PRODU07_08, nb2listw(prodMap_kn4))
```

Os mapas gerados para o índice de Moran local foram construídos por meio do pacote *sp* utilizando a função *spplot()*, e também foi carregado o pacote *RcolorBrewer* usando a função *brewer.pal()* para paleta de cores, conforme *script*:

```
> prodMap$lmi<-local.mi.prod[,1]
> prodMap$lmi.p<-local.mi.prod[,5]
> prodMap$lmi.p.sig<-as.factor(ifelse(local.mi.prod[,5]<.001,"Sig p<.001",
ifelse(local.mi.prod[,5]<.05,"Sig p<.05", "NS" )))
> require("RColorBrewer")
> require("sp")
> spplot(prodMap, "lmi", at=summary(prodMap$lmi), col.regions=brewer.pal(5,
"RdBu"), main="Local Moran's I")
> spplot(prodMap, "lmi.p.sig", col.regions=c("white", "#E6550D", "#FDAE6B"))
```

CONCLUSÃO

A partir deste estudo concluiu-se que houve autocorrelação espacial da produtividade da soja nas safras de 2007/2008 em meio da análise exploratória espacial por áreas, usando técnicas estatísticas como o índice I de Moran Univariado e I de Geary. Para o caso global dos Índices de Geary e Moran, ocorreu autocorrelação espacial entre os 48 municípios da região oeste do Paraná, constatando áreas com agrupamentos e também de transição. Na análise local, utilizando o índice LISA, áreas com características individuais correlacionadas também foram possíveis de identificação.

No diagrama de dispersão de Moran foi detectado que quatro municípios não apresentaram produtividade de soja semelhante a suas regiões próximas, sendo classificados como pontos discrepantes.

O *software* R demonstrou-se ser uma ferramenta muito útil para a análise espacial de áreas, pelo fato de ser gratuita e oferecer diversos pacotes para aplicação de técnicas de análise espacial, além de possuir funções completas para a geração de mapas temáticos.

Portanto, de forma geral, os métodos estatísticos espaciais que foram aplicados para este trabalho mostraram-se eficientes na identificação de padrões de área, e para a análise de autocorrelação espacial.

Spatial analysis of area applied to soybean productivity in the west region of Paraná using software R

ABSTRACT

This paper presents a study related to soybean productivity in the west region of Paraná, demonstrating how to apply some spatial analysis techniques of area using the software R. The study was conducted with data of soybeans in the period 2007/2008 to 48 municipalities. The Moran Global Index, Moran Local Index and Geary Index techniques were used to verify correlation and spatial autocorrelation, these procedures were performed with the libraries `spdep`, `sp`, `maptools` and `rcolorbrewer` of the software R. Through this study it was possible to demonstrate the applicability of the software R for spatial analysis of the area, and the steps of how to generate maps for better understanding of the information studied.

KEYWORDS: Spatial Autocorrelation. neighborhood Matrix. Moran Index. Geary Index.

REFERÊNCIAS

ALBUQUERQUE, P. H. M. **Curso de Econometria Espacial Aplicada**. Fortaleza:IPECE, 2009.

ALMEIDA, E. S.; HADDAD, E. A.; HEWINGS, G. J. D. **The spatial pattern of crime in Minas Gerais: an exploratory analysis**. *Economia Aplicada*, v. 9, n. 1, p.1-27, 2005.

ARAÚJO, Everton Coimbra. **Análise da produtividade da soja associada a fatores agrometeorológicos, por meio de estatística espacial de área, na região oeste do estado do paraná**. 2012. 113f. Dissertação de Mestrado - Universidade Estadual do Oeste do Paraná. Cascavel/PR. 2012.

ARAÚJO, E. C.; URIBE-OPAZO, M. A.; JOHANN, J. A. **Modelo de regressão espacial para estimativa da produtividade da soja associada a variáveis agrometeorológicas na região oeste do estado do Paraná**. *Eng. Agríc., Jaboticabal*, v. 34, n. 2, p. 286-299, abr. 2014. <http://dx.doi.org/10.1590/S0100-69162014000200010>.

ARAUJO, E. C. et al . **Classification of areas associated with soybean yield and agrometeorological variables through fuzzy clustering**. *Cienc. Inv. Agr., Santiago*, v. 40, n. 3, p. 617-627, dez. 2013. <http://dx.doi.org/10.4067/S0718-16202013000300014>.

BIVAND, R et al. **spdep: Spatial dependence: weighting schemes statistics and models**. 2014. Disponível em: <<https://CRAN.R-project.org/package=spdep>>. Acesso em: 10 nov. 2014.

BIVAND, R. et al. **maptools: Tools for reading and handling spatial objects**. 2014. Disponível em: <<https://CRAN.R-project.org/package=maptools>>. Acesso em: 07 nov. 2014.

BOHÓRQUEZ, L.; GÓMEZ, I.; SANTA, F. **Methodology for the discrimination of areas affected by forest fires using satellite images and spatial statistics**. *Procedia Environmental Sciences*, v. 7, p. 389–394, 2011.

CARVALHO, M. S. et al. **ANÁLISE ESPACIAL DE ÁREAS**. Brasília, DF. EMBRAPA. 2004.

CELEBIOGLU, F.; DALL'ERBA, S. **Spatial Disparities across the regions of Turkey: an exploratory spatial data analysis**, *The Annals of Regional Sciences*. v.45, n.2, p.379-400, 2009.

DRUCK, S. et al. **Análise espacial de dados geográficos**. Brasília, DF. EMBRAPA. 2004.

ESRI. **ArcGIS Spatial Analyst**. 2011. Disponível em: <<http://www.esri.com/software/arcgis/extensions/spatialanalyst/surface.html>>. Acesso em: 9 nov. 2014.

FOOK, K. D. **INTEGRAÇÃO DA ESTATÍSTICA ESPACIAL EM AMBIENTES GIS**. Pós-Graduação em Computação Aplicada. INPE. São José do Campos, SP. 2005.

KHAN, A. S. **DEPENDÊNCIA ESPACIAL E EMPREGO FORMAL: O que é possível afirmar para indústria cearense?**. In. Anais do VI Encontro de Economia do Ceará em Debate. Fortaleza - CE, novembro de 2010.

SEAB. **Soja – Análise da Conjuntura Agropecuária**. 2013. Disponível em: <<http://www.agricultura.pr.gov.br/modules/conteudo/conteudo.php?conteudo=32>>. Acesso em: 07 set. 2014.

SEAB. **Secretaria de Estado da Agricultura e do Abastecimento**. 2010.

SILVA, B. F.; DINIZ, J.; BORTOLUZZI, M. A. **Introdução ao software R**. 2009. Disponível em: <http://www.uft.edu.br/engambiental/prof/catalunha/arquivos/r/r_bruno.pdf>. Acesso em: 01 jul. 2014.

WRUBLACK, S. C. et al. **Spatial distribution of Canola culture in the State of Paraná (Brazil) between the agricultural years of 2005 and 2009**. Ciencia e investigación agraria, v. 40, n. 3, p. 523–535., 2013.

ZHANG, C.; MCGRATH, D. **Geostatistical and GIS analyses on soil organic carbon concentrations in grassland of southeastern Ireland from two different periods**. Geoderma, v. 119, n. 3–4, p. 261–275, 2004.

ZHANG, C. et al. **Use of local Moran's I and GIS to identify pollution hotspots of Pb in urban soils of Galway, Ireland**. Science of the Total Environment, v. 398, n. 1–3, p. 212–221, 2008.

TEAM, R. C. R Core Team. 2014. Disponível em R Core Team: <<http://www.R-project.org/>>. Acesso em 25 de out. 2016.

Recebido: 18 mai. 2017

Aprovado: 22 mar. 2018

DOI: 10.3895/rbgeo.v6n1.5912

Como citar: SEFFRIN, R.; ARAUJO, E. C.; BAZZI, C. L.. Análise espacial de área aplicada a produtividade de soja na região oeste do Paraná utilizando o software R. **R. bras. Geom.**, Curitiba, v. 6, n. 1, p. 23-43, jan/mar. 2018. Disponível em: <<https://periodicos.utfpr.edu.br/rbgeo>>. Acesso em: XXX.

Correspondência:

Rodolfo Seffrin

Rua Bahia,1323, CEP 85884-000, Medianeira, Paraná, Brasil.

Direito autoral: Este artigo está licenciado sob os termos da Licença Creative Commons-Atribuição 4.0 Internacional.

