

Pesquisa em Desenvolvimento Rural

**Técnicas, Bases de Dados
e Estatística Aplicadas aos
Estudos Rurais**

VOLUME 2

**Guilherme Francisco Waterloo Radomsky
Marcelo Antonio Conterato
Sergio Schneider**

ORGANIZADORES

Pesquisa em Desenvolvimento Rural



UNIVERSIDADE
FEDERAL DO RIO
GRANDE DO SUL

Reitor

Carlos Alexandre Netto

Vice-Reitor e Pró-Reitor
de Coordenação Acadêmica

Rui Vicente Oppermann

EDITORA DA UFRGS

Diretor

Alex Niche Teixeira

Conselho Editorial

Carlos Pérez Bergmann

Claudia Lima Marques

Jane Fraga Tutikian

José Vicente Tavares dos Santos

Marcelo Antonio Conterato

Maria Helena Weber

Maria Stephanou

Regina Zilberman

Temístocles Cezar

Valquiria Linck Bassani

Alex Niche Teixeira, presidente

Pesquisa em Desenvolvimento Rural

Técnicas, Bases de Dados
e Estatística Aplicadas aos
Estudos Rurais

VOLUME 2

Guilherme Francisco Waterloo Radomsky
Marcelo Antonio Conterato
Sergio Schneider

ORGANIZADORES

© dos autores
1ª edição: 2015

Direitos reservados desta edição:
Universidade Federal do Rio Grande do Sul

Capa: Carla M. Luzzatto
Revisão: Carlos Batanoli Hallberg
Editoração eletrônica: Fernando Piccinini Schmitt

P438 Pesquisa em desenvolvimento rural: técnicas, base de dados e estatística aplicadas aos estudos rurais – volume 2 / Organizadores Guilherme F. W. Rodomsky, Marcelo Antonio Conterato [e] Sergio Schneider. Porto Alegre: Editora da UFRGS, 2015.
344 p. : il. ; 16x23cm

(Série Estudos Rurais)

Inclui figuras, quadros e tabelas.

Inclui referências.

1. Agricultura. 2. Desenvolvimento Rural – Pesquisa - Técnicas de investigação. 3. Pesquisas socioeconômicas – Amostragem. 4. Censo Agropecuário – 2006. 5. Avaliação – Políticas Públicas – Gestão do Conhecimento – Programa de Desenvolvimento Sustentável de Territórios Rurais (PRONAT) 6. Estatística – Investigação rural. 7. Técnicas de investigação – Análise de Resultados. I. Rodomsky, Guilherme Francisco Waterloo. II. Conterato, Marcelo Antonio. III. Schneider, Sergio. IV. Série.

CDU 631.1:316.324.5:001.891

CIP-Brasil. Dados Internacionais de Catalogação na Publicação.
(Jaqueline Trombin – Bibliotecária responsável CRB10/979)

ISBN 978-85-386-0266-8

Técnicas de análise de dependência espacial de dados socioeconômicos¹

Iván G. Peyré Tartaruga

Introdução

Nos últimos anos verifica-se um crescente interesse no uso dos Sistemas de Informações Geográficas (SIG) e, conseqüentemente, das técnicas de análise de dados espaciais (georreferenciados) dos mais diversos aspectos da realidade (sociais, demográficos, econômicos e ambientais).

Este interesse se explica principalmente, por um lado, pela notável disponibilidade de dados territoriais e de bases cartográficas digitais. Apenas no Brasil podem-se obter, em muitos casos gratuitamente, estas informações no Instituto Brasileiro de Geografia e Estatística (IBGE), no Instituto Nacional de Pesquisas Espaciais (INPE), dentre outras instituições. E, por outro lado, pela considerável quantidade de programas computacionais para processamento e exame de dados espaciais à venda, como ArcGIS, IDRISI, MapInfo, dentre outros, ou gratuitos, como TerraView, GeoDa, gvSIG e outros.

¹ O autor agradece a leitura atenta da economista Fernanda Queiroz Sperotto da versão preliminar deste texto, sendo de inteira responsabilidade do autor a versão publicada.

Entretanto, esta facilidade de acesso a dados e ferramentas pode levar a um contexto de riqueza de informações concomitante a uma pobreza teórico-metodológica no estabelecimento de padrões, relacionamentos e hipóteses, como alerta Anselin (1993). Em razão disto, uma das metas do presente texto é revelar as dificuldades e, portanto, sinalizar as limitações do uso deste tipo de informações.

Tendo em mente tais possibilidades e limitações dos dados territoriais foram desenvolvidas diversas técnicas estatísticas específicas para seu tratamento. Assim, paralelamente à estatística convencional surgiu a estatística espacial em torno do início da década de 1950, com os trabalhos de Moran (1948) e Geary (1954), conforme Sabater, Tur e Azorín (2011). Anos antes, Sviatlovsky e Eells (1937) já sustentavam que para a resolução dos problemas concernentes à geografia, especificamente, e à análise regional, em termos gerais, somente o uso do mapa não era suficiente, nem tão-somente dos dados estatísticos. Para eles era evidente que os dois instrumentos deveriam ser trabalhados juntos.

O objetivo fundamental deste artigo é introduzir alguns métodos estatísticos que têm em conta a dimensão espacial dos fenômenos. Deste modo, o trabalho divide-se em quatro seções principais. A primeira que termina neste parágrafo e serve de preparação aos temas do texto. A segunda que mostra algumas características específicas, possíveis usos e, também, restrições referentes aos dados georreferenciados. A terceira que aprofunda o entendimento da dependência espacial, e apresenta e aplica suas técnicas estatísticas de análise. E a quarta seção que conclui o artigo.

○ uso dos dados espaciais: possibilidades e limitações

As informações estatísticas espaciais possuem características próprias que impõem a necessidade de alguns cuidados na sua utilização e até mesmo algumas restrições (Rogerson, 2012; Sabater, Tur e Azorín, 2011; Anselin, 1988 e 1992). Dentre os aspectos mais pertinentes para a análise de dados espaciais estão (a) o problema da unidade de área modificável, (b) os problemas de fronteira e os dois efeitos espaciais – (c) a heterogeneidade espacial e (d) a dependência espacial.

O *problema da unidade de área modificável* (em inglês *the modifiable areal unit problem – MAUP*) diz respeito à sensibilidade das análises estatísticas ao tipo de regionalização utilizado para agregar os dados no espaço. Openshaw e Taylor (1979) foram os primeiros a identificar este problema ao analisar a correlação entre a votação recebida por candidatos do Partido Republicano e a

proporção de idosos, num conjunto de condados dos EUA. Eles demonstraram categoricamente que a escolha entre diferentes sistemas de agregação espacial dos dados levavam a diferentes interpretações do mesmo fenômeno. Rogerson (2012) apresenta um interessante esquema que exemplifica este problema por meio de dois zoneamentos para examinar fluxos migratórios (Figura 1), onde os deslocamentos dos migrantes são representados por setas com origem e destino. Enquanto, no esquema (a) não ocorre migração entre as diferentes zonas, no (b) se constata um forte movimento geral para o sul. Portanto, fica evidente que para duas regionalizações, de um mesmo território, as respectivas interpretações do mesmo fenômeno são totalmente diferentes.

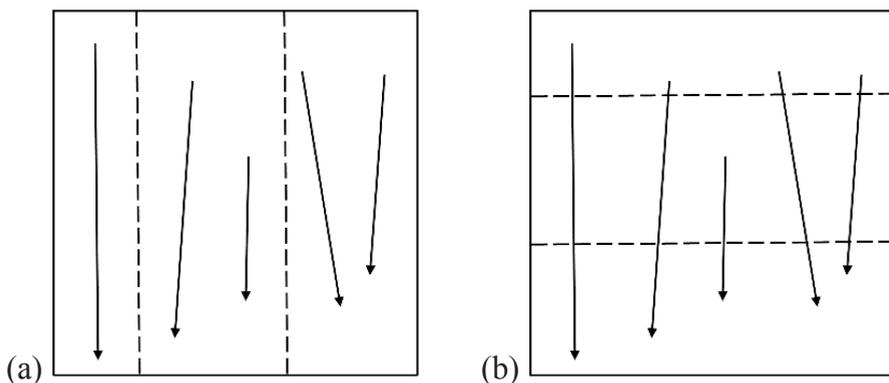


Figura 1 – Esquemas de dois zoneamentos diferentes para fluxos de migração demonstrando o problema da unidade de área modificável (as setas identificam a origem e o destino dos migrantes)
Fonte: Adaptado de Rogerson (2012, p. 16).

Como alerta Rogerson (2012), deve-se ter ciência de que ao estabelecer os limites e a escala geográfica de um conjunto de zonas define-se uma estrutura espacial que corresponde a uma interpretação específica, e talvez única, de um dado fenômeno no território. O pesquisador, ao analisar um fato ambiental ou social, deve buscar a alternativa de zoneamento mais adequada para representá-lo. Assim, o caso da Figura 1 mostra claramente a questão da escolha de diferentes limites. No âmbito escalar, por exemplo, para examinar movimentos pendulares diários é preferível um quadro de zonas em uma escala de pequena dimensão, pois zonas de menor área acabam captando melhor estes deslocamentos. Ao passo que para deslocamentos aéreos a escala mais adequada é a de maior dimensão.

Com as geotecnologias atuais fica facilitado o uso de diversas regionalizações para a análise de dados, como municípios, micro ou macrorregiões do

IBGE, possibilitando testar a sensibilidade dos resultados para cada caso. Vale destacar que a abordagem conceitual do estudo é crucial na eleição dos limites e da escala do zoneamento, portanto, somente com o suficiente conhecimento das características próprias do fenômeno em tela é que será possível fazer tal escolha.

Os *problemas de fronteira*, o segundo aspecto, se referem às influências sofridas por uma área de estudo em razão de agentes ou estruturas exteriores (Rogerson, 2012). Ao analisar, como exemplo, o desempenho econômico de um município específico deve-se verificar se os consumidores e os setores (primário, secundário ou terciário) dos congêneres vizinhos atuam de modo importante naquele município. Neste ponto, é importante observar a forma e o tamanho dos limites, e as relações e as vias de comunicação entre os espaços próximos. Porém, deve-se olhar, também, mais além dos entornos adjacentes aos territórios em análise. Atualmente, com a contínua intensificação dos processos econômicos da globalização, cada vez mais agentes (empresas, governos, instituições) distantes agem fortemente sobre o local.

Ao focalizar aspectos específicos relativos aos dados e aos modelos utilizados na análise espacial, Anselin (1988 e 1992) denomina estes de efeitos espaciais sobre os dados e os divide em dois tipos gerais: a heterogeneidade espacial e a dependência espacial.

O efeito da *heterogeneidade espacial* trata da diferenciação regional, ou da unicidade de cada lugar, na qual cada localização é possuidora de características únicas. A heterogeneidade está associada a uma ausência de estabilidade do comportamento das variáveis sobre o espaço em questão e, assim, os parâmetros que caracterizam as unidades variam, na direção e/ou intensidade, com a localização.

A *dependência espacial*, também conhecida como associação espacial ou autocorrelação espacial, tem como fundamento a primeira lei da geografia de Tobler, qual seja, “todas as coisas estão relacionadas com todas as outras, porém coisas próximas estão mais relacionadas do que as distantes” (Tobler, 1970, p. 236). Nesse sentido, pode-se considerar que, em um conjunto de entes geográficos (pontos, linhas ou polígonos), valores similares de uma variável tendem a estar próximos espacialmente, caracterizando um aglomerado (*cluster*). A título de exemplo, veja-se o caso de um bairro metropolitano com altas taxas de criminalidade. Assumindo a lei de Tobler, provavelmente os bairros vizinhos também possuem um grande número de crimes. Portanto, a análise de dependência espacial busca verificar a associação entre diferentes unidades territoriais que depende do conceito de distância, em que a noção de espaço (relativo) pode ir além da ideia de distância (euclidiana), ou seja, pode-se con-

siderar a distância associada ao espaço social, político, econômico (custo de deslocamento), filosófico, etc., mostrando um grande leque de possibilidades de estudos para as ciências sociais em geral.

Análise da dependência espacial

Alguns aspectos da análise de dependência espacial

O conceito de dependência espacial tem por base a ideia de que no espaço geográfico todos os elementos estão relacionados entre si, porém, os elementos mais próximos territorialmente estão mais relacionados do que os mais distantes, como estabelece a primeira lei da geografia. Esse princípio mais claramente se manifesta com variáveis naturais, contudo com variáveis humanas (sociais, demográficas, econômicas, dentre outras) deve-se verificar e medir esse aspecto para cada caso particular, pois os comportamentos socioespaciais podem afastar-se dessa lei (Buzai e Baxendale, 2006).

A análise de autocorrelação espacial pode ser realizada nas formas univariada ou bivariada. Na univariada se tenta medir a relação da variável de um determinado espaço com a mesma variável dos espaços contíguos, ou seja, o comportamento da mesma variável em diferentes unidades territoriais, por isso o uso do prefixo “auto”. Já no modo bivariado o procedimento efetua-se com duas variáveis diferentes, semelhante à correlação da estatística clássica, no qual, se busca descobrir o relacionamento de uma variável em uma unidade espacial com os valores de outra variável nas unidades vizinhas.

Além do mais, há dois usos do exame de autocorrelação: um não trivial e outro trivial. No âmbito do exercício não trivial, a ocorrência de autocorrelação espacial entre diferentes unidades territoriais, muito frequente em situações reais, colocam fortes implicações sobre as interpretações da estatística padrão. Efetivamente, a estatística inferencial, usada para interpretar uma população (universo) a partir de uma amostra, possui como pressuposto básico que as observações devem ser independentes entre si. Entretanto, a dependência espacial das variáveis afeta a análise de regressão levando a erros de interpretação.

O uso trivial do procedimento de autocorrelação está dentro do rol de técnicas de Análise Exploratória de Dados Espaciais (AEDE) (em inglês *Exploratory Spatial Data Analysis – ESDA*). Estas técnicas visam à utilização da estatística na análise espacial, de modo a compreender as relações entre variáveis distribuídas em um conjunto de unidades espaciais quando não há hipóteses ou expectativas claras sobre o fenômeno estudado (Sabater, Tur e Azorín, 2011). Por conseguinte, procura-se descobrir algum sentido nas informações como

arranjos, aglomerações (*clusters*) ou dispersão, em outras palavras, verificar se a localização influencia o comportamento da distribuição da variável em estudo. Desta forma, ao examinar o relacionamento entre os valores de uma unidade espacial e de suas unidades vizinhas podem surgir as seguintes situações:

- *Autocorrelação espacial positiva*: quando há semelhança entre os valores das unidades espaciais vicinais, situação que pode indicar a tendência de aglomeração;

- *Autocorrelação espacial negativa*: quando se apresentam valores não semelhantes, nos quais os valores das unidades vizinhas são distintos da unidade principal;

- *Inexistência de autocorrelação espacial*: quando não ocorre nenhuma das situações anteriores. Os valores das unidades contíguas se produzem de forma aleatória, não há relação espacial.

Para efetuar o exame da autocorrelação de dados distribuídos no território deve-se estabelecer a *matriz de proximidade* (ou de pesos) da distribuição. Esta é obtida, primeiramente, escolhendo o critério de proximidade que pode ser pela distância, quando se elege uma distância (euclidiana) determinada, na qual a partir de uma unidade espacial todas as unidades espaciais dentro do raio desta distância são consideradas vizinhas; ou pela distância por rodovias, por ferrovias, via fluvial ou marítima, dentre outras. Ou pela contiguidade (para o caso de polígonos), ou seja, os vizinhos são aqueles polígonos em contato direto (contíguos), critério muito utilizado em estudos sociais e econômicos.

O critério de contiguidade pode ser de dois tipos: o tipo *rook* (a “torre” do jogo de xadrez) ou *queen* (“rainha”) (ver Figura 2). O primeiro significa que somente serão considerados contíguos os polígonos (municípios, regiões, etc.) que possuírem um limite maior do que um ponto em comum, enquanto o segundo considera contíguos tanto os polígonos anteriores quanto os de contato pontual (único ponto).

Junto ao critério de contiguidade também deve ser definida a ordem de contiguidade (1, 2, 3,...), no qual, a ordem “um” considera vizinhos somente os polígonos em contato direto com a unidade espacial principal, a ordem “dois”, aqueles da ordem “um” e os em contato destes (vizinhos dos vizinhos), e assim por diante.

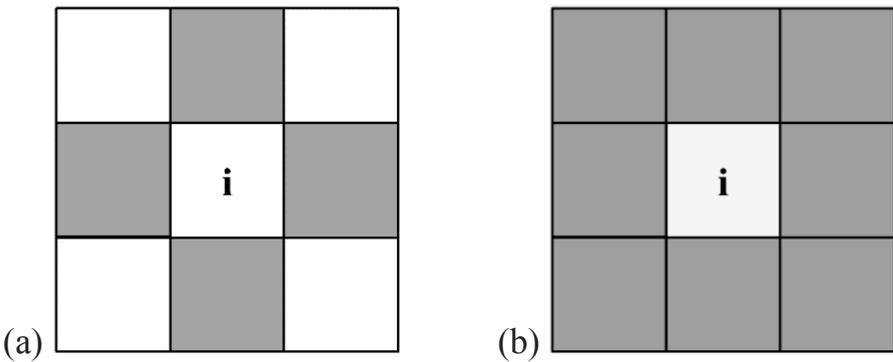


Figura 2 – Representações esquemáticas dos critérios de contiguidade (a) *rook* (torre) e (b) *queen* (rainha) para as relações de vizinhança com a unidade espacial “i”
 Fonte: Adaptado de Ribeiro (2011, p. 302).

Com o critério de contiguidade definido, *rook* ou *queen*, usa-se a seguinte notação:

- Quando duas unidades são vizinhas, $w_{ij} = 1$, onde i e j indicam as unidades espaciais e w , a existência de vizinhança ou “peso” entre as unidades (o “w” vem da palavra *weight*, peso em inglês);
- Quando duas unidades não são vizinhas, $w_{ij} = 0$.

Com estas informações pode-se, finalmente, obter a matriz de proximidade ou matriz de pesos de um conjunto de unidades territoriais que se quer estudar, em que nesta matriz quadrada estão indicados quais são vizinhos (ver exemplo na Figura 3), cuja formulação é a seguinte:

$$W = \begin{pmatrix} w_{11} & w_{12} & \cdots & w_{1j} \\ w_{21} & w_{22} & \cdots & w_{2j} \\ \vdots & \vdots & \cdots & \vdots \\ w_{i1} & w_{i2} & \cdots & w_j \end{pmatrix}, i = 1, 2, \dots, n \text{ e } j = 1, 2, \dots, n; \quad (1)$$

onde W é a matriz de proximidade $n \times n$, n é o número de unidades espaciais e w_{ij} é a medida de proximidade (ou peso) entre as unidades i e j .

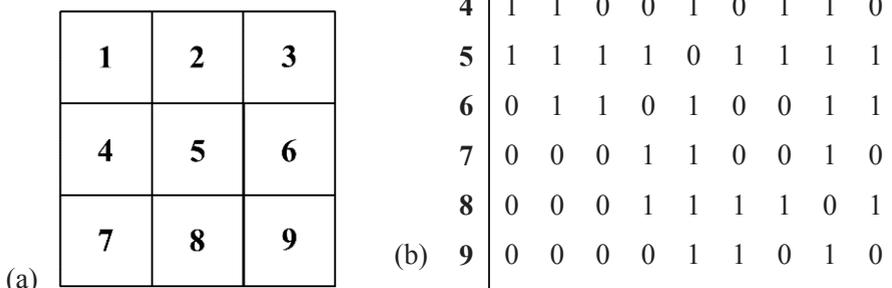


Figura 3 – Representação esquemática de um território com nove unidades espaciais (a) e a respectiva matriz de proximidade tipo *queen* e ordem de contiguidade um (b)
 Fonte: Elaborado pelo autor.

A partir da matriz de proximidade, uma das formas de determinar a autocorrelação espacial é através da estatística I de Moran ou índice de Moran, que pode ser global ou local.

Autocorrelação espacial global univariada

A estatística I de Moran visa medir o comportamento geral (global) de um dado espaço subdividido em um conjunto de unidades espaciais, este índice varia de -1 a $+1$, em que valores próximos a $+1$ indicam autocorrelação espacial positiva (denotando um padrão espacial), próximos a -1 , autocorrelação negativa, e próximos a zero significam ausência de autocorrelação ou de padrão espacial (Anselin, 1995; Rogerson, 2012; Sabater, Tur e Azorín, 2011).

A formulação matemática do índice de Moran (I) é a seguinte:

$$I = \frac{n \sum_i \sum_j w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\left(\sum_i \sum_j w_{ij}\right) \sum_i (y_i - \bar{y})^2}, \quad i = 1, 2, \dots, n \text{ e } j = 1, 2, \dots, n; \quad (2)$$

em que y_i e y_j são os respectivos valores da variável em estudo das regiões i e j , e \bar{y} é a média aritmética da variável.

Uma forma de simplificação da Equação 2 é através da transformação da variável em questão em um escore z (processo também conhecido como normalização).²

Desta forma, o índice I de Moran é obtido por meio da expressão:

$$I = \frac{n \sum_i^n \sum_j^n w_{ij} z_i z_j}{(n-1) \sum_i^n \sum_j^n w_j}, \quad i = 1, 2, \dots, n \text{ e } j = 1, 2, \dots, n; \quad (3)$$

onde z_i e z_j são os escores z dos respectivos dados das regiões i e j . A interpretação da Equação 3 fica mais visível na parte do seu numerador, em que está a soma dos produtos dos escores z dos valores das regiões contíguas. Portanto, os pares de regiões em que os escores estiverem ao mesmo tempo acima da média, ou abaixo da média, vão contribuir positivamente no numerador da Equação 3, resultando em uma autocorrelação espacial positiva; já os pares onde uma região possui escore acima da média (positivo) e a outra, escore abaixo da média (negativo) para o numerador, originará uma autocorrelação espacial negativa.

O índice adquire sua significância estatística por meio da realização de um teste de hipóteses, no âmbito dos métodos da estatística inferencial, que tem por objetivo verificar se a distribuição espacial de uma variável se produz aleatoriamente. Neste teste a hipótese nula (H_0) é definida para o caso *da configuração espacial se originar de maneira aleatória*, ou de aleatoriedade espacial, e a hipótese alternativa (H_1), para o caso *da configuração espacial não se originar de forma aleatória*.

Com as hipóteses estabelecidas deve-se determinar o *nível de significância* (denotado por α ou erro Tipo I), que indica a probabilidade de rejeição da hipótese nula, sendo ela verdadeira. O nível de significância, ou seja, o grau de risco de cometer um erro, é escolhido em função da importância do problema em questão. De modo geral, em análises espaciais utilizam-se níveis de 5% (0,05) e 1% (0,01) (Buzai e Baxendale, 2006), ou, em outros termos, com uma *confiança* de 95% e 99%, respectivamente. O resultado para o teste de hipóteses é designado como *p-valor*.³

² Os escores z , ou valores normalizados, são obtidos a partir do seguinte fórmula: $z = \frac{x - \bar{x}}{s}$, onde z é o escore z do valor, x é o valor original, \bar{x} é a média aritmética do conjunto dos valores x_i , e s é o desvio padrão dos valores x_i , i variando de um a n , e n o número de unidades espaciais. Os valores dos escores são interpretados como unidades de desvio padrão acima (quando positivos) ou abaixo (negativos) da média.

³ Ao se estabelecer, por exemplo, um nível de significância de 5% ($\alpha = 0,05$) para um caso e o *p-valor* encontrado é de 4%, pode-se rejeitar a hipótese nula (e aceitar a hipótese alternativa) com uma probabilidade de erro em 4% das ocorrências. Caso contrário, ao obter-se um *p-valor* de 6%,

No âmbito do I de Moran, a hipótese nula indica a ausência de autocorrelação espacial, pois a distribuição dos dados no espaço está sujeita ao acaso, sem nenhum padrão espacial, ao passo que a ocorrência da hipótese alternativa aponta para uma autocorrelação positiva ou negativa.

Autocorrelação espacial local univariada

Enquanto a autocorrelação espacial global é uma medida do comportamento geral da distribuição da variável no território, a autocorrelação local visa determinar as áreas onde há padrões espaciais. Para identificar estes padrões é possível utilizar o método LISA (*Local Indicators of Spatial Association*) (Anselin, 1995), no qual seu cálculo sintetiza, a partir do comportamento de cada unidade espacial, o índice I de Moran global. Desta maneira, quando o exame global, de um lado, apontar a aleatoriedade territorial dos dados, os testes locais podem ser importantes para descobrir locais isolados de maior concentração. De outro, quando o teste global indicar autocorrelação espacial, a análise local pode ser útil para verificar se o espaço analisado é homogêneo, isto é, se há um padrão espacial em toda a área, ou se existem locais extremos (*hot* ou *cold spots*)⁴ que exercem forte influência na estatística global. Portanto, como lembra Rogerson (2012), esta estatística local pode ser utilizada tanto para confirmar hipóteses de estudo quanto de forma exploratória, dentro das técnicas de AEDE.

A estatística I de Moran local (I_j) é definida por

$$I_j = n(y_i - \bar{y}) \sum_j^n w_j (y_j - \bar{y}), j = 1, 2, \dots, n \quad (4)$$

e a soma dos índices Moran locais de um conjunto de unidades espaciais é igual ao Moran global, ou seja, $\sum I_j = I$.⁵

A maneira de verificar e visualizar a dependência espacial (global e local) é através do *diagrama de espalhamento de Moran*. Esse diagrama é elaborado a partir dos valores normalizados (escores z) da variável a ser analisada (Z). Com posse dos valores normalizados é construído um gráfico bidimensional

portanto, superior ao nível de significância, deve-se rejeitar a hipótese alternativa e, ao mesmo tempo, aceitar a hipótese nula de aleatoriedade da configuração territorial.

⁴ *Hot spots* (zonas quentes) são áreas onde há uma concentração da variável em análise em alta intensidade, enquanto *cold spots* (zonas frias) são localizações onde a variável encontra-se com valores muito baixos.

⁵ Na verdade, este somatório em alguns casos resulta em um valor proporcional ao índice global.

de Z (dados normalizados) – eixo horizontal (abscissas) –, por WZ (média dos dados normalizados dos vizinhos) – eixo vertical (ordenadas) –, com base na matriz de proximidade correspondente.⁶ No gráfico a seguir (Figura 4) a linha de tendência relacionada aos pontos locados corresponde ao indicador I de Moran (global), que é equivalente ao coeficiente de regressão linear (inclinação da reta de regressão) de WZ em Z , que varia de -1 a $+1$. E os quatro quadrantes do gráfico (Q_1 , Q_2 , Q_3 e Q_4) podem ser interpretados das seguintes formas, correspondendo ao índice de Moran local (LISA):

- Q_1 (valores e médias positivos) e Q_2 (valores e médias negativos): os pontos presentes nestes quadrantes indicam autocorrelação espacial positiva destas unidades espaciais (os vizinhos possuem valores semelhantes);
- Q_3 (valores positivos e médias negativas) e Q_4 (valores negativos e médias positivas): nestes os pontos indicam autocorrelação negativa (vizinhos desta unidade possuem valores diferentes).

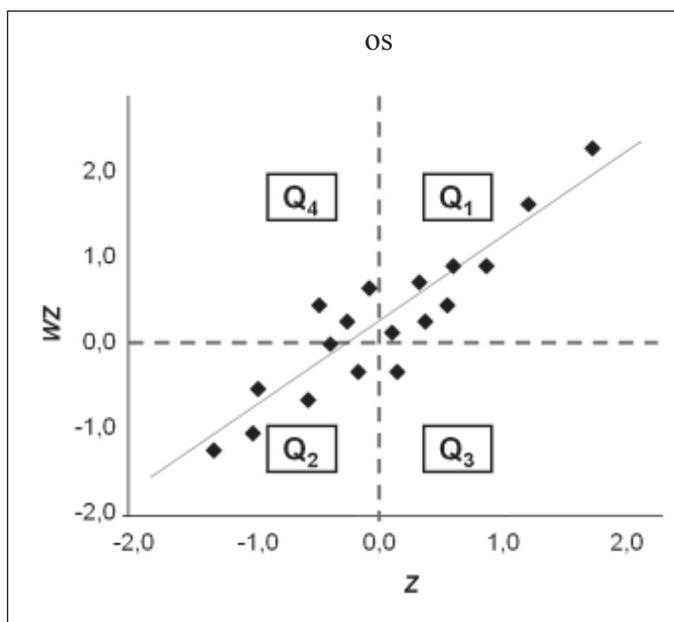


Figura 4 – Esquema de um diagrama de espalhamento de Moran para dados hipotéticos
 Fonte: Elaborado pelo autor.

⁶ O conjunto WZ é geralmente nomeado de atraso espacial (*spatial lag*), em referência às unidades espaciais vizinhas à unidade principal em análise.

As unidades espaciais, na análise local, podem ser apresentadas num mapa em quatro classes possíveis denominadas de:

- “Alto-Alto” (*High-High*): Z e WZ com valores positivos ou ambos acima da média (quadrante Q1)
- “Baixo-Baixo” (*Low-Low*): ambos, Z e WZ , com valores negativos ou abaixo da média (quadrante Q2)
- “Alto-Baixo” (*High-Low*): Z com valores positivos (acima da média) e WZ negativos (abaixo da média) (quadrante Q3)
- “Baixo-Alto” (*Low-High*): Z negativos (abaixo da média) e WZ positivos (acima da média) (quadrante Q4).

Para o exame do nível de significância calcula-se o índice de Moran para cada unidade espacial e após permutam-se aleatoriamente os valores das demais unidades diversas vezes para obter-se, desse modo, uma pseudodistribuição dos índices calculados e, então, verifica-se a significância do índice real. Com essas informações (níveis de significância) é possível determinar as unidades espaciais que apresentam uma autocorrelação significativamente diferente ao restante das informações. Estas regiões podem ser aglomerações de não estacionaridade, portanto, regiões com dinâmica espacial particular merecedoras de uma análise mais detalhada.

Autocorrelação bivariada

A análise bivariada, tanto global como local, é a verificação da correlação espacial entre duas variáveis diferentes, em que o diagrama de dispersão de Moran é construído de forma que o eixo Z (valores normalizados) corresponda a primeira variável (independente) e WZ (média dos vizinhos), a segunda variável (dependente). Portanto, as formulações matemáticas e os diagramas de dispersão são semelhantes às correspondentes para uma única variável (univariada), com a diferença de que agora se busca o grau de associação entre uma variável e os valores das unidades vizinhas de outra variável.

Aplicação

A título de exemplo, veja-se o exame da autocorrelação espacial para o percentual de estabelecimentos agropecuários com baixo nível de instrução da pessoa que dirige o estabelecimento, por municípios do estado do RS no ano de 2006.⁷ A análise de dependência espacial deste tipo de informação pode

⁷ Dados obtidos no censo agropecuário de 2006 do IBGE. Neste trabalho, o baixo grau de instrução é definido para aqueles dirigentes que não sabem ler e escrever, ou que possuem ensino fundamental

auxiliar no planejamento de uma política pública de educação para o meio rural-agrário gaúcho.

Fazendo uso do *software* GeoDa⁸ chegou-se aos resultados a seguir para a análise de Moran, com a matriz de proximidade construída a partir do critério de contiguidade do tipo *queen* e de ordem “um” de contiguidade, e um nível de significância de 5% (0,05), logo uma probabilidade de acerto de no mínimo 95%.

O índice *I* de Moran global foi de aproximadamente 0,4818 (Figura 5), o que corresponde ao ângulo da reta de tendência, revelando a existência de autocorrelação espacial positiva. Em termos práticos, esta reta mostra que, de maneira geral, os municípios possuem proporções de estabelecimentos com baixo nível de instrução do dirigente (representados no eixo *Z*) semelhantes às médias das proporções dos respectivos vizinhos (representados no eixo *WZ*). Em termos interpretativos, a tendência geral da distribuição da variável no território foi de que a localização importa, ou seja, os valores (proporções) dos municípios provavelmente têm influência (relação) sobre seus respectivos vizinhos; por exemplo, se uma municipalidade apresentar um percentual alto (ou baixo) da variável (percentual de estabelecimentos) há grande probabilidade de que os municípios limítrofes apresentem igualmente um valor alto (ou baixo).

O teste de significância concluiu que a estrutura de autocorrelação espacial ($I = 0,4818$) é significativa de cerca de 99,9%, pois o *p-valor* foi de 0,001 (0,1%), por conseguinte, inferior ao nível de significância arbitrado que era de 5%. Em outras palavras, há pouca probabilidade (apenas 0,1%) de que a estrutura espacial de relacionamentos se produza ao acaso (aleatoriamente). Na Figura 6 pode-se observar o histograma com os vários *I* de Moran gerados nas 999 permutações (quando os valores de instrução são intercambiados nas unidades espaciais de forma aleatória) e o índice da distribuição dos dados verdadeira; somente 0,1% das simulações foram semelhantes ao *I* real.

incompleto ou alfabetização de adultos, ou que sabem ler e escrever, porém não tiveram nenhuma instrução formal.

⁸ Foi utilizado o OpenGeoDa (versão 1.2.0), *software* desenvolvido pela equipe de um dos maiores estudiosos, atualmente, da análise de dados geográficos, Luc Anselin (disponível gratuitamente em <<http://geodacenter.asu.edu/>>).

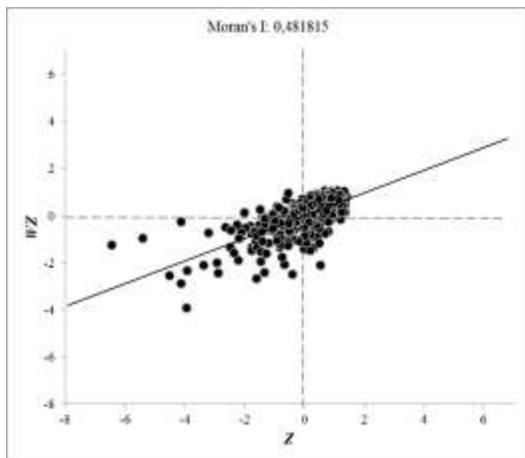


Figura 5 – Diagrama de espalhamento de Moran da proporção (%) de estabelecimentos agropecuários com baixo nível de instrução da pessoa que dirige o estabelecimento (elaborado a partir do GeoDa)

Fonte: Elaborado pelo autor a partir do GeoDa.

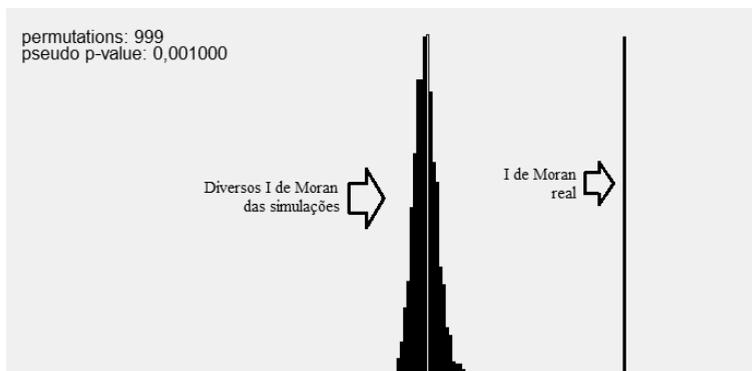


Figura 6 – Significância estatística da autocorrelação espacial (elaborado a partir do GeoDa)

Fonte: Elaborado pelo autor a partir do GeoDa.

Tendo garantida a significância estatística, a análise LISA (local) dos dados determinou os municípios que formam aglomerações significativas (Figura 7). Para os objetivos da aplicação há interesse nos 68 municípios do tipo “Alto-Alto” que configuram os agrupamentos com os níveis mais baixos de instrução dos dirigentes de estabelecimentos agropecuários. Estas municipalidades conformam basicamente três aglomerações principais: ao norte, no centro e outra mais ao sul do Estado. Assim, estas áreas deveriam ser analisadas como potenciais focos de uma política educacional específica,

de forma a otimizar os esforços, principalmente, em uma situação de recursos escassos para tal ação.

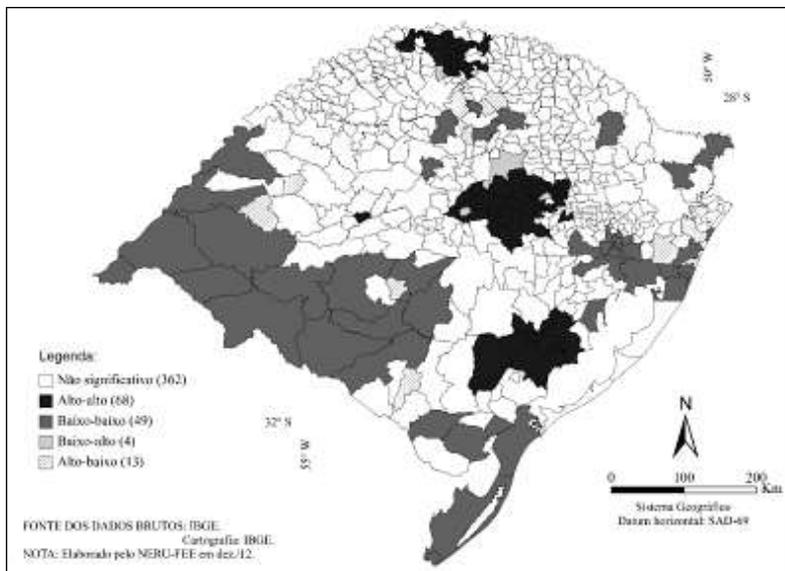


Figura 7 – Mapa LISA de aglomerações
Fonte: NERU-FEE, a partir de dados de cartografia IBGE

Como comentado anteriormente, os métodos do I de Moran global e local (LISA) podem ser utilizados para fins exploratórios, como no exemplo em tela, em que o objetivo é mais de sugerir soluções para o problema. Nesse sentido, o teste global indicou um grau significativo de aglomerações de municípios com respeito à variável analisada no território gaúcho (Figura 5), porém sem apontar as possíveis localizações exatas de tais aglomerados, esta tarefa ficando a cargo do exame localista.

O teste local (Figura 7) mostrou que a área de estudo não é homogênea, isto é, as estatísticas locais não se apresentam parecidas em toda a área, e, portanto, há lugares específicos onde estão as aglomerações significativas. Entretanto, é importante ressaltar, o método estatístico em si próprio não explica o porquê de tal configuração espacial, ele tão-somente está indicando estatisticamente que, nessas localizações, cada município é influenciado ou tem um relacionamento forte com sua vizinhança – o espaço importa. O método encerra um caráter relacional e não explicativo ou causal. Por conseguinte, a partir destes procedimentos devem seguir outras análises, com base em co-

nhecimento teórico-empírico específico do fenômeno em estudo, objetivando agora sim a explicação dos fatos. No âmbito explicativo, esta metodologia pode, no máximo, servir para a verificação de hipóteses.

Considerações Finais

Apesar de todos os obstáculos para o uso dos métodos espaciais de análise de dados (*MAUP*, problemas de fronteira,...), são promissoras as possibilidades de aplicações importantes e úteis tanto no âmbito das pesquisas socioeconômicas e ambientais, quanto nos campos da gestão e do planejamento públicos e privados. Além das conhecidas utilizações das técnicas territoriais acerca da criminologia, das questões urbanas, da saúde coletiva, dentre outras temáticas, vislumbram-se possíveis estudos na área de geografia econômica sobre as interações socioespaciais no desenvolvimento territorial de cidades, regiões e países.

Especificamente aos espaços rurais, a utilização das técnicas de análise exploratória de dados espaciais também é promissora. Com efeito, são inúmeras as aplicações destas análises no sentido de verificar se uma distribuição espacial de uma característica social ou econômica (nível de instrução no meio rural, produtividade, estrutura fundiária, etc.) ocorre de forma aleatória ou se há autocorrelação na configuração espacial. Ou se existe associação espacial na distribuição de uma variável em relação à outra (forma bivariada), como, por exemplo, ao investigar se a produtividade de uma área tem influência na renda rural de seu entorno.

A partir deste trabalho espera-se que mais gestores e pesquisadores, das mais diferentes formações, levem em conta a dimensão espacial dos respectivos fenômenos de interesse e, acima de tudo, tenham consciência de que os atributos espaciais absolutos (localização) e relativos (distância, arranjo espacial) dos objetos ou eventos estudados podem ser elementos importantes para sua compreensão e para conjecturas sobre a sua evolução.

Referências

ANSELIN, L. *Spatial econometrics: methods and models*. Dordrecht: Kluwer Academic Publishers, 1988.

_____. *Spatial data analysis with GIS: an introduction to application in the social sciences*. Technical Report 92-10, 1992.

_____. *The Moran Scatterplot as an ESDA Tool to Assess Local Instability in Spatial Association*. Paper presented at the GISDATA Specialist Meeting on GIS and Spatial Analysis, Amsterdam, Dec. 1-5, 1993.

_____. Local Indicators of Spatial Association – LISA. *Geographical Analysis*, v. 27, n. 2, p. 93-115, apr. 1995.

BUZAI, G. D.; BAXENDALE, C. A. *Análisis socioespacial con Sistemas de Información Geográfica*. Buenos Aires: Lugar Editorial/GEPAMA, 2006.

GEARY, R. C. The Contiguity Ratio and Statistical Mapping. *The Incorporated Statistician*, London, v. 5, n. 3, p. 115-145, 1954.

MORAN, P. A. P. The Interpretation of Statistical Maps. *Journal of the Royal Statistical Society – Series B (Methodological)*, v. 10, n. 2, p. 243-251, 1948.

OPENSHAW, S.; TAYLOR, P. J. A million or so correlation coefficients: three experiments on the modifiable areal unit problem. In: WRIGLEY, N. (Org.). *Statistical Applications in the Spatial Sciences*. London: Pion Limited, 1979.

RIBEIRO, A. Modelos de regressão territorial. In: COSTA, J. S.; DENTINHO, T. P.; NIJKAMP, P. (Coord.). *Compêndio de Economia Regional*. Vol. II: métodos e técnicas de análise regional. Parede: Princípia, 2011.

ROGERSON, P. A. *Métodos estatísticos para geografia: Um guia para o estudante*. 3.ed. Porto Alegre: Bookman, 2012.

SABATER, L. A.; TUR, A. A.; AZORÍN, J. M. N. Análise Exploratória de Dados Espaciais (AEDE). In: COSTA, J. S.; DENTINHO, T. P.; NIJKAMP, P. (Coord.). *Compêndio de Economia Regional*. Vol. II: métodos e técnicas de análise regional. Parede: Princípia, 2011.

SVIATLOVSKY, E. E.; EELLS, W. C. The Centrophraphical Method and Regional Analysis. *Geographical Review*, v. 27, n. 2, p. 240-254, 1937.

TOBLER, W. R. A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, v. 46, p. 234-240, 1970.