

Agrupamento de dados multivariados de seguro rural no Brasil com base em medidas locais de autocorrelação espacial.

Walef Machado de Mendonça

PPGEAB

6 de Dezembro de 2020

Introdução – Seguro rural

- Ambiente de elevado risco e grande incerteza.
- Oscilações na renda do setor.
- Gerenciamento de risco – contratação de seguro rural
- Programa de Subvenção ao Prêmio do Seguro Rural (PSR), instituído pela Lei 10.823/2003 e decreto no 5.121/2004 (BRASIL, 2018)

■ Gerais:

- 1 Agrupar espacialmente os municípios do Brasil segundo variáveis relativas ao seguro rural.

■ Específicos:

- 1 Apresentar um procedimento para o agrupamento espacial com base na distribuição espacial de dados multivariados.
- 2 Aplicar o procedimento de agrupamento utilizando as medidas de autocorrelação local I de Moran local e G_i local.
- 3 Subsidiar a tomada de decisões sobre políticas públicas de estímulo à demanda por produtos de seguro específicas para cada grupo de municípios.

Dados referentes a apólices de seguro rural dos municípios brasileiros no ano de 2018.

Sigla	Variável
apolices_contratadas	Total de apólices contratadas
total_segurado_mil	Soma da importância segurada (R\$ milhão)
soma_premio_total_mil	Soma dos prêmios (R\$ milhão)
total_subvencao_mil	Total de subvenção (R\$ milhão)
valor_indenizacoes_pagas_mil	Soma das indenizações pagas (R\$ milhão)
sinistralidade_media	Total de sinistros pagos / total do prêmio arrecadado
taxa_media	Taxa média aplicada às apólices
apolices_indenizadas	Número de apólices indenizadas

Os dados utilizados estão disponíveis no site do Ministério da Agricultura, Pecuária e Abastecimento (MAPA). Também foram utilizados dados geográficos, como a posição e o formato, do território brasileiro disponíveis no site do Instituto Brasileiro de Geografia e Estatística (IBGE)

	apolices_contratadas	total_segurado_mil	soma_premio_total_mil	total_subvencao_mil
Média	31.24	6,182,709.17	425,507.89	182,321.48
Desvio padrão	68.31	12,827,391.88	1,065,823.27	456,709.66
Min	1.00	7,812.00	335.83	151.12
25%	2.00	385,000.00	17,141.87	7,519.30
50%	7.00	1,572,481.89	85,647.30	37,273.52
75%	29.00	6,444,608.00	403,375.43	174,454.50
Max	918.00	180,919,573.67	23,268,328.48	10,432,116.36

	valor_indenizacoes_pagas_mil	sinistralidade_media	taxa_media	apolices_indenizadas
Média	453,285.73	7,486.30	455.36	8.68
Desvio Padrão	1,496,503.65	13,706.05	673.90	29.90
Min	0.00	0.00	0.01	0.00
25%	0.00	0.00	61.00	0.00
50%	1,850.94	2.28	171.00	1.00
75%	253,649.12	8,974.00	520.09	4.00
Max	37,003,064.14	128,373.00	4,456.53	771.00

Tabela: Resumo estatístico das variáveis

Autocorrelação espacial global

- Um índice de autocorrelação espacial mede a associação espacial nos dados considerando o **local** e o **valor** da variável em estudo.
- Hipótese: Aleatoriedade na distribuição espacial da variável.

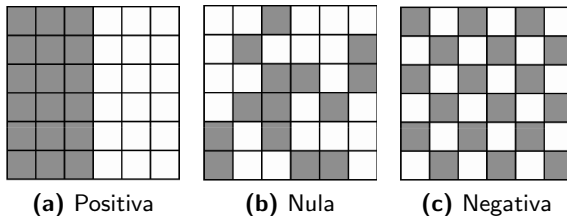


Figura: Padrões de autocorrelação espacial

Fonte: Radil (2011)

- É necessário usar uma estatística de teste para a aleatoriedade da distribuição espacial global da variável.

O I de Moran mede a relação do desvio padronizado de uma variável z numa área i com o desvio padronizado das áreas vizinhas j para a mesma variável z (ALMEIDA, 2012).

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

em que $z_i = (x_i - \bar{x})$, w_{ij} é uma medida de contiguidade entre i e j , n é o número de regiões e S_0 é a soma dos pesos espaciais (w_{ij})

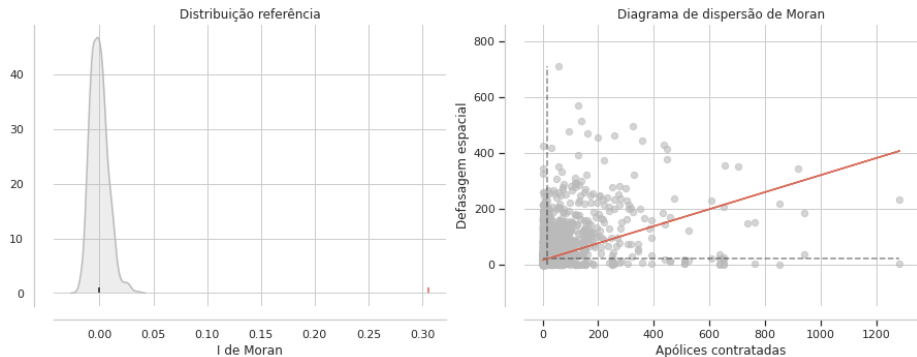


Figura: / de Moran para Apólices contratadas

Outra medida de autocorrelação espacial global é o G de Getis e Ord (ALMEIDA, 2012).

$$G(d) = \frac{\sum_i \sum_j w_{ij}(d) y_i y_j}{\sum_i \sum_j y_i y_j}, \quad \text{para } j \neq i \quad (2)$$

em que $w_{ij}(d)$ é uma medida de contiguidade entre i e j dentro de uma distância d e y_i é o valor observado para a variável.

A estatística (2) assume valores em $[0, 1]$, em que valores próximos de 1 indicam concentração de valores altos e valores próximos de 0 indicam concentração de valores baixos.

- Estatísticas globais fornecem padrões de associação espacial em todo o conjunto de dados
- Medidas de autocorrelação local buscam identificar padrões no interior de uma região de estudo
- Podem informar a existência de um *cluster* de valores autocorrelacionados em nível local
- Podem informar sobre existência de *outliers* locais
- Para Anselin (1995), um indicador local de associação espacial (*LISA*) deve satisfazer a dois critérios:
 - 1 Deve indicar clusters espaciais estatisticamente significativos;
 - 2 A soma dos indicadores locais deve ser levar ao indicador global.

Um indicador local de associação espacial do tipo *LISA* é o I de Moran local que é expresso por:

$$I_i = z_i \sum_j w_{ij} z_j,$$

em que z_i e z_j são os valores da variável padronizada nas regiões i e j , w_{ij} é uma matriz de pesos espaciais.

/ de Moran local

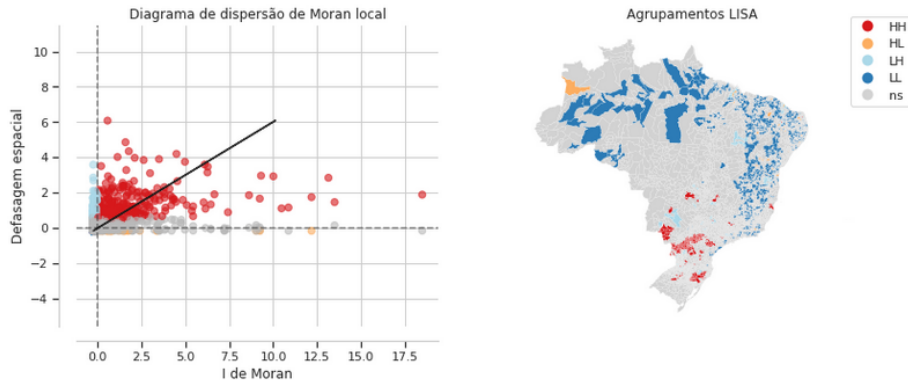


Figura: / de Moran local para Apólices contratadas

Outra forma de se analisar de forma local a associação espacial é o G_i local proposto por Getis e Ord (ALMEIDA, 2012). Essa estatística permite encontrar padrões do tipo *hot spot* ou *cool spot*.

$$G(d) = \frac{\sum_j w_{ij}(d)y_j}{\sum_j y_j}, \quad \text{para } j \neq i$$

em que j são os vizinhos próximos de i .

Agrupamentos com base medidas locais de autocorrelação espacial

- 1 Dada uma matriz de pesos espaciais W , calcule para cada variável as estatísticas I de Moran e Getis-Ord locais. Reuna esses valores em uma matriz Z de dimensão $(n \times p)$. Cada coluna de Z expressa o padrão de autocorrelação local para uma variável, enquanto cada linha de Z fornece o perfil de agrupamento em torno de cada unidade local.
- 2 Aplique o algoritmo k-médias neste conjunto de novas variáveis. Esta etapa permite agrupar observações com base em seus perfis espaciais multivariados que contêm informações de localização e das variáveis.

Resultados – I de Moran local

Grupo	apolices_contratadas	total_segurado_mil	soma_premio_total_mil	total_subvencao_mil
0	0.12	0.18	0.11	0.11
1	94.46	20.55	17.49	19.40
2	79.81	18.01	13.74	15.25
3	11.51	5.60	4.86	5.11
4	15.08	36.11	48.06	47.82

Grupo	valor_indenizacoes_pagas_mil	sinistralidade_media	taxa_media	apolices_indenizadas
0	0.07	0.21	0.29	0.05
1	333.82	51.43	12.34	294.29
2	123.04	12.37	7.80	147.08
3	5.55	14.82	16.58	8.92
4	11.13	6.01	8.94	4.63

Tabela: Média dos valores de I de Moran local das variáveis nos grupos

Resultados – I de Moran local

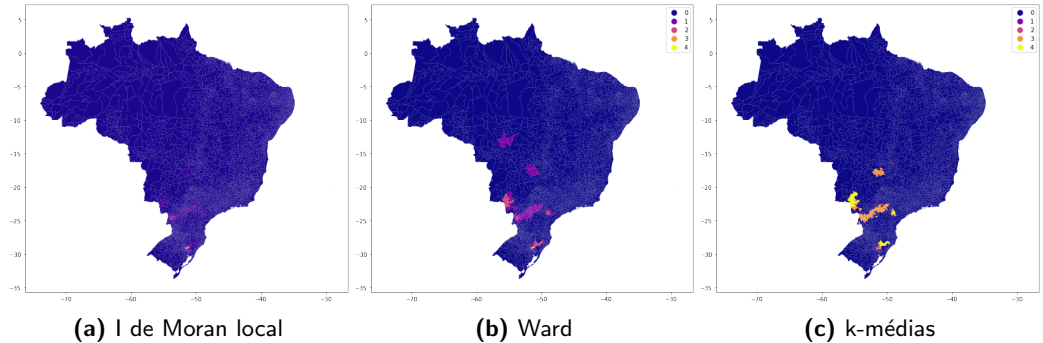


Figura: Resultados com I de Moran local

Resultados – G_i local

Grupos	apolices_contratadas	total_segurado_mil	soma_premio_total_mil	total_subvencao_mil
0	-0.19	-0.17	-0.17	-0.17
1	3.53	2.85	2.85	2.90
2	0.90	1.10	0.94	0.92

Grupos	valor_indenizacoes_pagas_mil	sinistralidade_media	taxa_media	apolices_indenizadas
0	-0.15	-0.21	-0.23	-0.15
1	2.75	3.26	3.47	3.08
2	0.66	1.14	1.28	0.52

Tabela: Média dos valores de G_i local das variáveis nos grupos

Resultados – G_i local

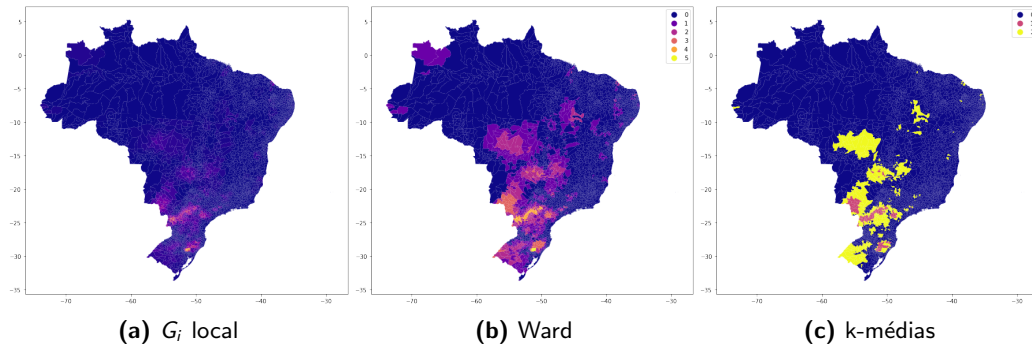


Figura: Resultados com G_i local

ALMEIDA, E. **Econometria Espacial Aplicada**. Campinas-SP: Alínea, 2012.

ANSELIN, L. Local indicators of spatial association - LISA. **Geographical analysis**, v. 27, p. 93-115, 1995.

BRASIL. **Programa de Subvenção ao Prêmio do Seguro Rural: Relatório Geral 2019**. Brasília, DF: Brasília: Ministério da Agricultura, Pecuária e Abastecimento (MAPA), 2019

JORDAHL, K. **GeoPandas**: Python tools for geographic data. 2014. Disponível em: github.com/geopandas/geopandas, Acesso em: 28 jul. 2020.

PYTHON. **The Python programming language**. Disponível em: github.com/python/cpython Acesso em: 18 jul. 2017.

RADIL, S. M. Spatializing social networks: making space for theory in spatial analysis. 2011. Tese de Doutorado. University of Illinois at Urbana-Champaign.

REY, S. J.; ANSELIN, L. PySAL: A Python library of spatial analytical methods. **Review of Regional Studies**, v. 37, n. 1, p. 5–27, 2007.

SCRUCCA, L. Clustering multivariate spatial data based on local measures of spatial autocorrelation: an application to the labour market of Umbria. **Quaderni del Dipartimento di Economia, Finanza e Statistica**, p. 1–25, 2005. Disponível em: <http://www.ec.unipg.it/DEFS/uploads/spatcluster.pdf>.