

# Uma introdução à econometria espacial

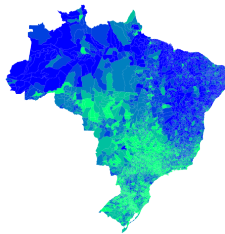
Walef Machado de Mendonça

PPGEAB

20 de Abril de 2021

# Introdução

- A econometria espacial tem como objetivo **especificar**, **estimar**, **testar** e **prever** modelos teóricos influenciados pelos efeitos espaciais (ALMEIDA, 2012).
- Na econometria espacial as observações representam regiões. Ex.: bairros, distritos, setores censitários, regiões urbanas etc



Geralmente a estrutura espacial contribui para a modelagem de duas formas:

- 1 O processo gerador dos dados é explicitamente espacial. (ex.: preços de casas) → a geografia é uma variável e contribui **diretamente** para o modelo.
- 2 Uma alternativa é utilizar a localização espacial como uma variável **instrumental**. (ex.: o erro é maior em algumas áreas do que em outras)

# As hipóteses de Gauss-Markov

$$y = X\beta + \varepsilon \quad (1)$$

- 1 Linearidade dos parâmetros
- 2 Não colinearidade perfeita
- 3 Média condicional zero

$$\mathbb{E}(\varepsilon_i|X) = 0$$

- 4 Homocedasticidade

$$\text{Var}(\varepsilon|X) = \sigma^2$$

- 5 Independência dos Erros

$$\mathbb{E}(\varepsilon_i\varepsilon_j|X) = 0 \text{ para } i \neq j$$

## ■ Dependência espacial

- As unidades de corte transversal – domicílios, municípios, regiões – não são independentes entre si.
- O valor de uma variável numa região  $i$ ,  $y_i$ , depende do valor dessa mesma variável nas regiões vizinhas  $j$ ,  $y_j$ .

$$y_i = f(y_j, X), \quad i, j = 1, \dots, n \quad \text{e} \quad i \neq j$$

$$y_1 = \rho_2 y_2 + \beta X_1 + \varepsilon_1$$

## ■ Heterogeneidade espacial

- Ocorre quando há instabilidade estrutural através das regiões → diferentes respostas dependendo da localidade.

$$y_i = f_i(X_i, \beta_i, \xi_i) \quad \xi_i \sim (0, \Omega)$$

## 1 Lineariedade dos parâmetros

- Na presença de heterogeneidade espacial deve-se alterar a especificação do modelo para permitir que os coeficientes sejam variáveis de acordo com a localização do fenômeno.

## 2 Não colinearidade perfeita

- Essa hipótese continua válida. Porém, ao se utilizar a defasagens espaciais a colnearidade pode aumentar.

## 3 Homocedasticidade

- É comum que a variância não seja constante devido aos efeitos espaciais.

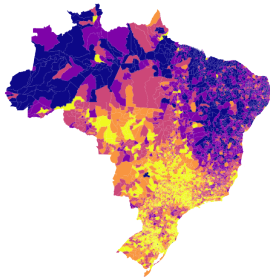
## 1 Média condicional zero

- Pressupõe-se que o erro da região  $i$  não esteja correlacionado com a variável explicativa na região  $i$ .
- Omissão de defasagens espaciais relevantes
- Multidirecionalidade e caráter endógeno da defasagem espacial  $\rightarrow$  estimador  $MQO$  é viesado e inconsistente.

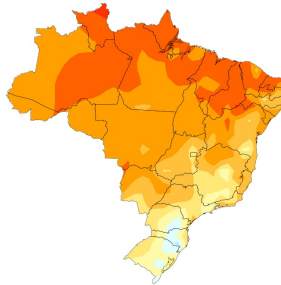
## 2 Independência dos Erros

- Quando as observações são unidades geográficas é provável que essa hipótese seja violada. (Amostra aleatória de regiões?)
- O erro de uma região  $i$  pode estar correlacionado linearmente com o erro de outra região  $j$ .  
 $\rightarrow$  Ineficiência, viés e inconsistência do estimador  $MQO$

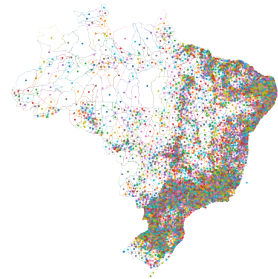




**(a)** Áreas com contagem e taxas agregadas



**(b)** Superfícies contínuas



**(c)** Eventos ou padrões pontuais

**Figura:** Exemplos de dados espaciais

# Matriz de ponderação espacial

- É uma medida de proximidade entre as regiões.
- Para um conjunto de  $n$  áreas é definida como:

$$W = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1n} \\ w_{21} & w_{22} & \dots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \dots & w_{nn} \end{bmatrix},$$

Cada elemento  $w_{ij}$  representa uma medida de proximidade entre as áreas  $i$  e  $j$ .

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

Por convenção,  $w_{ij} = 0, \quad \forall i = j$

## ■ Defasagem

$$y_{ij} \rightarrow y_{i-1,j}, y_{i,j-1}, y_{i+1,j}, \dots$$

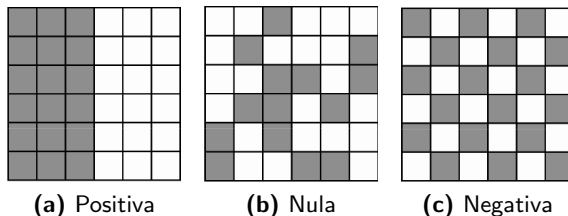
$$\mathbf{W} \times y = \mathbf{W}y$$

em que  $\mathbf{W}$  é a matriz de pesos espaciais,  $y$  é o vetor da variável e  $\mathbf{W}y$  é a defasagem espacial da variável

- $\mathbf{W}y$  é a média do valor da variável nas regiões vizinhas se a matriz de pesos for normalizada nas linhas.

# 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

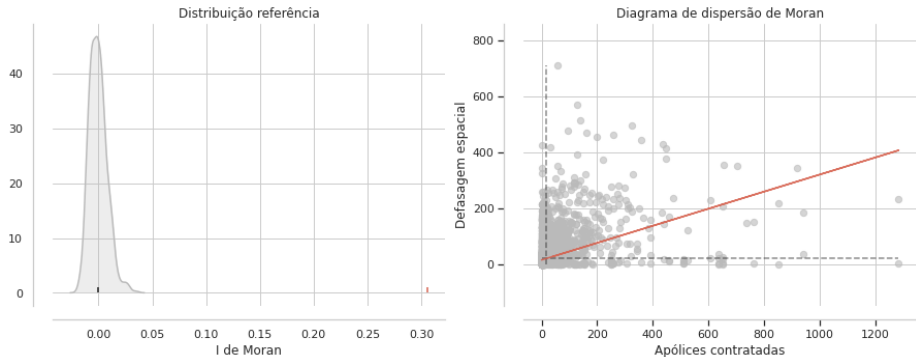
Fonte: Radil (2011)

- 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).

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

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: I de Moran para Apólices contratadas**

- 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.

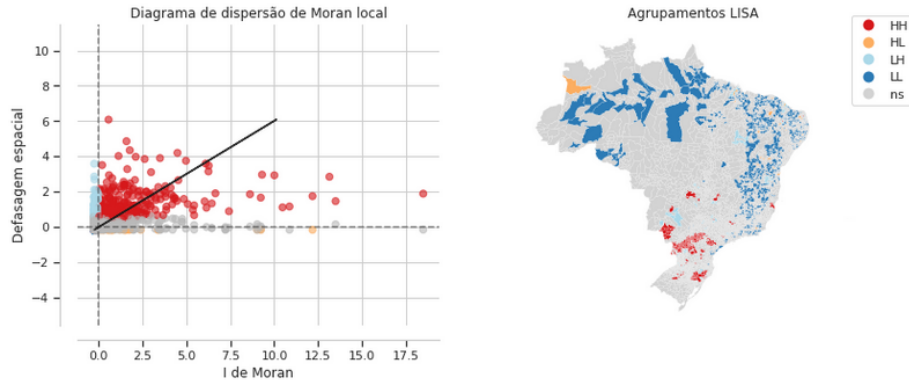
Um indicador local de associação espacial do tipo *LISA* é o  $I$  de Moran local que é expresso por:

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

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

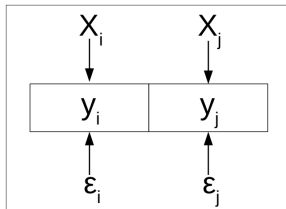


**Figura:** / de Moran local para Apólices contratadas

## ■ Modelo Clássico de Regressão Linear

$$y = X\beta + \varepsilon \quad \varepsilon \sim N(0, \sigma^2 \mathbf{I})$$

- **Processo a-espacial:**

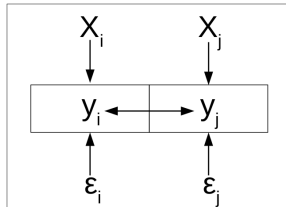


## ■ Modelo de defasagem espacial (SAR – Spatial Auto Regressive)

$$y = \rho Wy + \varepsilon \quad \text{ou} \quad y_i = \rho \sum_{j=1}^n w_{ij} y_j + \varepsilon_i$$

em que  $Wy$  é um vetor de defasagens espaciais (variável endógena) e  $\rho$  é o coeficiente auto-regressivo espacial.

- **Processo espacial:**



## ■ Modelo de defasagem espacial (SAR – Spatial Auto Regressive)

Incluindo variáveis exógenas (SAR misto)

$$y = \rho Wy + X\beta + \varepsilon \quad \text{ou} \quad y_i = \rho \sum_{j=1}^n w_{ij} y_j + \sum_{j=1}^p x_{ij} \beta_j + \varepsilon_i$$

$X$  é a matriz de variáveis explicativas exógenas

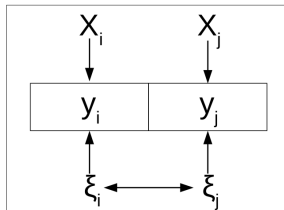
## ■ Modelo de erro auto-regressivo espacial (SEM – Spatial Error Model)

$$y = X\beta + \xi \quad \text{ou} \quad y_i = \sum_{j=1}^p x_{ij}\beta_j + \xi_i$$

$$\xi = \lambda W\xi + \varepsilon \quad \text{ou} \quad \xi_i = \lambda \sum_{j=1}^n w_{ij}u_j + \varepsilon$$

## ■ Modelo de erro auto-regressivo espacial (SEM – Spatial Error Model)

- Processo espacial:



## ■ Modelo de defasagem espacial com erro autorregressivo espacial (SAC – Spatial autoregressive combined)

Há interação na variável de interesse  $y$  e no erro  $\xi$ .

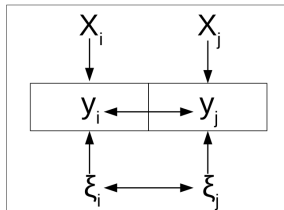
$$y = \rho W_1 y + X\beta + \xi \quad \text{ou} \quad y_i = \rho \sum_{j=1}^n w_{1ij} y_j + \sum_{j=1}^p x_{ij} \beta_j + \xi_i$$

$$\xi = \lambda W_2 \xi + \varepsilon \quad \text{ou} \quad \xi_i = \lambda \sum_{j=1}^n w_{2ij} \xi_j + \varepsilon_i$$

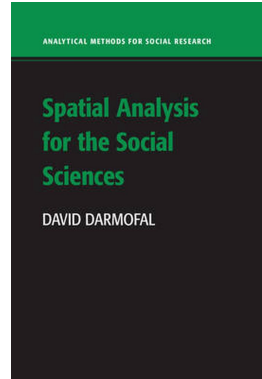
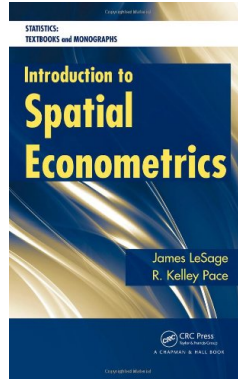
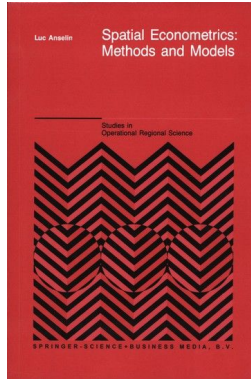
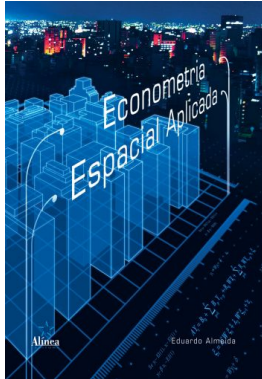
$$|\rho| < 1 \text{ e } |\lambda| < 1.$$

## ■ Modelo de defasagem espacial com erro autorregressivo espacial (SAC – Spatial autoregressive combined)

- Processo espacial:







- [https://darribas.org/gds\\_scipy16/](https://darribas.org/gds_scipy16/)
- <https://spatial.uchicago.edu/>
- <https://www.youtube.com/user/GeoDaCenter>
- <https://geopandas.org/index.html>
- <https://geodacenter.github.io/index.html>
- <https://pysal.org/packages/>
- <https://www.liverpool.ac.uk/geographic-data-science/>
- <https://geographicdata.science/book/intro.html>
- **Repositório da apresentação no GitHub:**  
[https://github.com/walefmachado/spreg\\_rural\\_insurance](https://github.com/walefmachado/spreg_rural_insurance)

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.

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

PYTHON. **The Python programming language**. Disponível em: [github.com/python/cpython](https://github.com/python/cpython) Acesso em: 18 jul. 2017.

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.