Medidas de similaridade e pesos espaciais

Patrícia de Siqueira Ramos

UNIFAL-MG, *campus* Varginha

23 de Outubro de 2017

Medidas de similaridade

Estatísticas para medir a autocorrelação:

- Devem capturar a similaridade de valor e de localização
- Similaridade de valor: y em i (sua localização) e em j (seus vizinhos)
- Esses índices tentam resumir o quanto observações similares tendem a estar próximas

Medidas de similaridade

O que esses índices globais de autocorrelação tentam fazer é resumir a similaridade de valores em locais i e j e ponderam a similaridade de acordo com a proximidade:

- alta similaridade + alto peso = valores similares próximos
- baixa similaridade + alto peso = valores dissimilares próximos

Medidas de similaridade

As medidas de similaridade podem envolver três tipos de cálculo:

- produto cruzado: $y_i \cdot y_j$ (I de Moran é desse tipo)
- diferenças ao quadrado: $(y_i - y_j)^2$
- módulo das diferenças: $|y_i - y_j|$

Todas essas medidas globais sugerem se há agrupamentos, mas não identificam áreas específicas onde eles se encontram (para isso existem medidas locais - LISA)

Medidas de similaridade

Produto cruzado:

$$y_i \cdot y_j$$

- sob H_0 (aleatoriedade espacial), esse produto não deve ser muito grande ou muito pequeno
- quando valores altos estão sempre juntos, o produto é alto e vice-versa

Medidas de dissimilaridade

Quadrado das diferenças e módulo das diferenças:

$$(y_i - y_j)^2, \qquad |y_i - y_j|$$

- sob H_0 (aleatoriedade espacial), o resultado não deve dar muito alto ou baixo
- quando valores altos ou baixos estão sistematicamente juntos, o resultado é pequeno
- por isso, é uma medida de dissimilaridade (quanto menor o resultado, maior dissimilaridade)

I de Moran

Estatística mais utilizada para medir a autocorrelação espacial:

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

em que $z_i = y_i - \bar{y}$ representa o desvio em relação à média da variável.

I de Moran

Estatística mais utilizada para medir a autocorrelação espacial:

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

em que $z_i = y_i - \bar{y}$ representa o desvio em relação à média da variável. Sua fórmula se parece com o coeficiente de correlação:

$$r = \frac{\sum_i (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_i (X_i - \bar{X})^2} \sqrt{\sum_i (Y_i - \bar{Y})^2}},$$

mas o valor de I depende dos pesos w_{ij}.

I de Moran

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

I de Moran

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

em que $z_i = y_i - \bar{y}$ representa o desvio em relação à média da variável

- y_i: valor da variável em um determinado local i
- n: número de observações
- $S_0 = \sum \sum w_{ij}$
- $E[I] = -1/(n-1)$

I de Moran

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

em que $z_i = y_i - \bar{y}$ representa o desvio em relação à média da variável

- y_i: valor da variável em um determinado local i
- n: número de observações
- $S_0 = \sum \sum w_{ij}$
- $E[I] = -1/(n-1)$

Obs.: se usarmos a matriz de pesos ($\boldsymbol{W}$) normalizada:

- $S_0 = \sum \sum w_{ij} = n$
- $\frac{n}{S_0} = 1$

I de Moran

Interpretação:

- I próximo de $+1$: autocorrelação positiva
- I próximo de 0: ausência de autocorrelação espacial (aleatoriedade)
- I próximo de -1: autocorrelação negativa

Obs.: $E[I] = -1/(n-1) \approx 0$

c de Geary

Outra estatística que pode ser utilizada para medir a autocorrelação espacial:

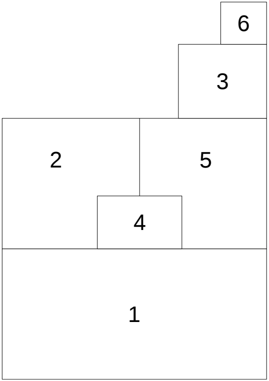
$$c = \frac{n-1}{2\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij}(y_i - y_j)^2}{\sum_i (y_i - \bar{y})^2}.$$

- O valor do c situa-se entre 0 e 2
- Valor esperado $= 1$
- $c < 1$: autocorrelação positiva (concentração)
- $c > 1$: autocorrelação negativa (dispersão)

Pesos espaciais

- expressão formal de similaridade local
- força da interação: efeito combinado do coeficiente de autocorrelação + pesos

Ex.: Representação de vizinhos



Seis polígonos - vizinhos compartilham fronteira

Matriz de pesos W

- Matriz positiva $N \times N$ com elementos w_{ij}
- Matriz binária:
  1 se tem fronteira entre os pares
  0 se não tem fronteira
- Por convenção, $w_{ii} = 0$ (uma região não é vizinha de si mesma)

Matriz de pesos *W*

Do exemplo:

Matriz de pesos $\boldsymbol{W}$

Do exemplo:

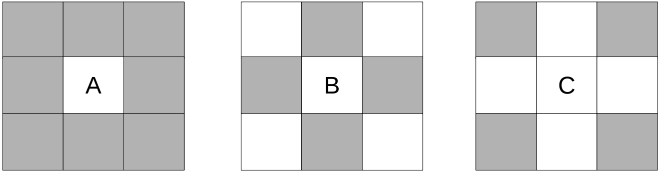
$$\boldsymbol{W} = \begin{bmatrix} 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

(simétrica)

Matriz de pesos W

- Há diferentes convenções de contiguidade (definição de vizinhos):
  - rainha (*queen*): além das fronteiras, os vértices também são considerados
  - torre (*rook*): apenas fronteiras físicas com extensão diferente de zero
  - bispo (*bishop*): apenas os vértices são levados em conta

  Obs.: rainha e torre são as mais usadas



Rainha Torre Bispo

Matriz de pesos W^*

- Usar w_{ij} como 0's e 1's pode não ser uma boa ideia (pode trazer problemas ao calcular a média dos vizinhos)
- Padronizar de forma que $\sum_j w_{ij} = 1$
- Do exemplo:

Matriz de pesos W^*

- Usar w_{ij} como 0's e 1's pode não ser uma boa ideia (pode trazer problemas ao calcular a média dos vizinhos)
- Padronizar de forma que $\sum_j w_{ij} = 1$
- Do exemplo:

$$\mathbf{W}^* = \begin{bmatrix} 0 & 1/3 & 0 & 1/3 & 1/3 & 0 \\ 1/3 & 0 & 0 & 1/3 & 1/3 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 \\ 1/3 & 1/3 & 0 & 0 & 1/3 & 0 \\ 1/4 & 1/4 & 1/4 & 1/4 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

(não simétrica)

Observações sobre a matriz de pesos W

- Há também pesos baseados em distância geográfica (quanto mais próximas duas regiões, maior o peso)
- Um problema que pode ocorrer em situações práticas: ilhas (regiões sem vizinhos)
  - ilhas devem ser removidas da análise
- Número grande de vizinhos também pode ser ruim: perde-se variabilidade