

Introdução à Estatística Espacial

CURSO CIEG - Semana II

Dados de área

Paulo Justiniano Ribeiro Jr

LEG - Laboratório de Estatística e Geoinformação, UFPR

paulojus@leg.ufpr.br

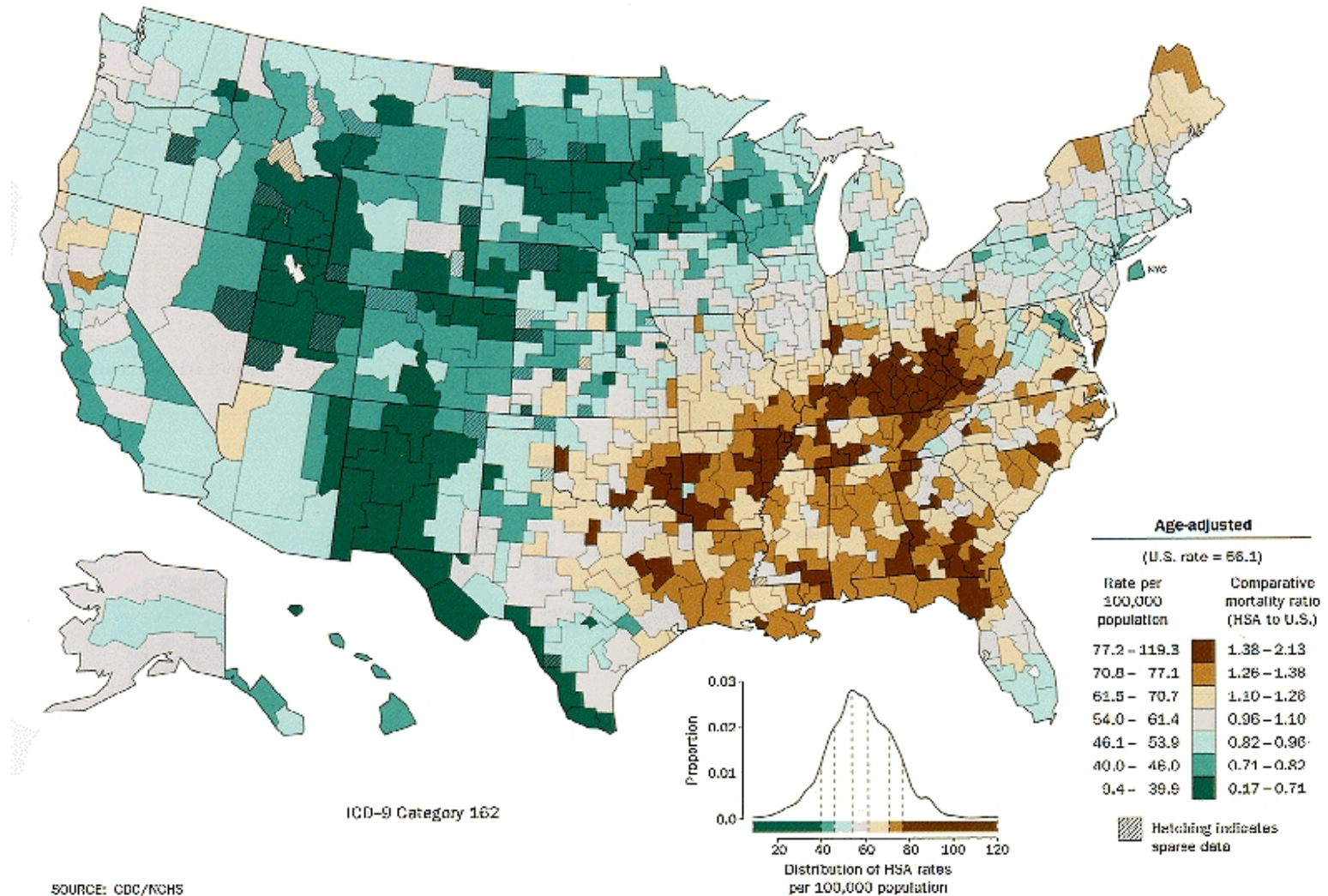
<http://www.leg.ufpr.br/cieg2010>

Baseado e materiais de:

Renato M. Assunção (LESTE - Laboratório de Estatística Espacial, UFMG)

DPI/INPE

Dados de Área - Cancer EUA



Dados de Área

Mapa de uma região \mathcal{R} particionado em n áreas.

Terminologia: Região e sub-região serão aglomerados de áreas.

Região $\mathcal{R} = \bigcup_{i=1}^n \mathcal{A}_i$ com $\mathcal{A}_i \cap \mathcal{A}_j = \emptyset$ se $i \neq j$

Na área i é feita uma observação aleatória Y_i .

Y_i envolve uma agregação ou uma soma sobre a área \mathcal{A}_i : número de homicídios na área i ; proporção de idosos na área i ; arrecadação de ICMS na área i .

Medições referem-se a toda a área \mathcal{A}_i , não a um ponto particular dentro dela

Vamos distinguir a variável aleatória de seu valor observado usando maiúsculas (Y) para a variável e minúsculas (y) para seu valor observado.

Padrão espacial é comum

Todo mapa apresenta sub-regiões com alguns valores relativamente mais altos aglomerados em algumas partes, enquanto que valores relativamente mais baixos ficam em outras partes do mapa.

Encontramos também muito "ruído": sub-regiões com valores altos e baixos misturados aleatoriamente sem nenhum arranjo espacial mais organizado.

Isto ocorre mesmo quando os valores observados no mapa são "jogados" de forma completamente casual nas áreas.

Padrões espaciais espúrios

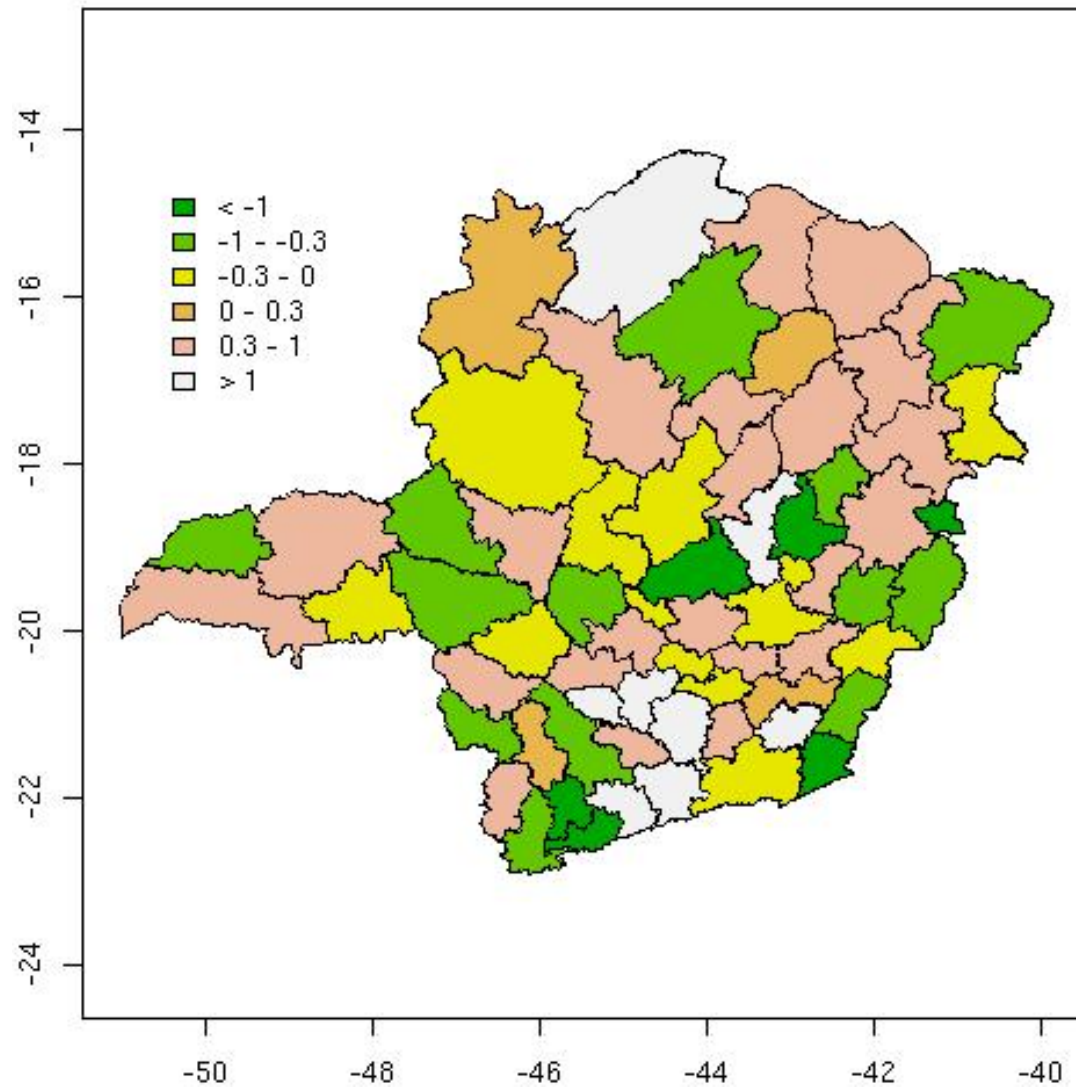
Por exemplo, role um dado bem balanceado para cada área e coloque uma cor associada com a face que aparecer.

Qualquer aglomeração espacial de valores altos ou baixos é completamente casual.

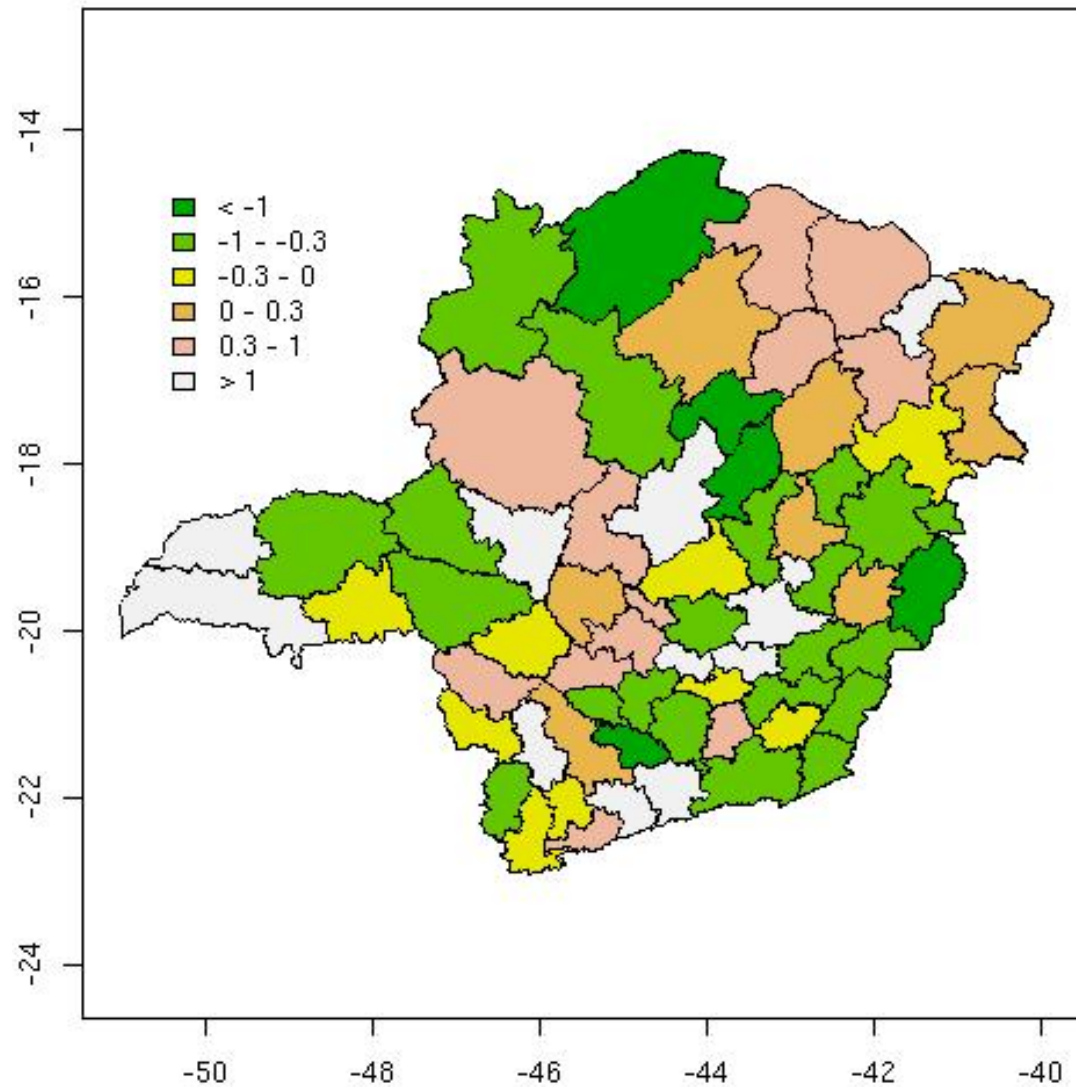
Não existe um mecanismo que, ao gerar os dados, estimule de alguma forma a organização espacial das cores.

Outro exemplo: SEM DIZER NADA SOBRE O MAPA, gere no computador números aleatórios independentes e com normal $N(0, 1)$. Aloque os números ao mapa de forma casual.

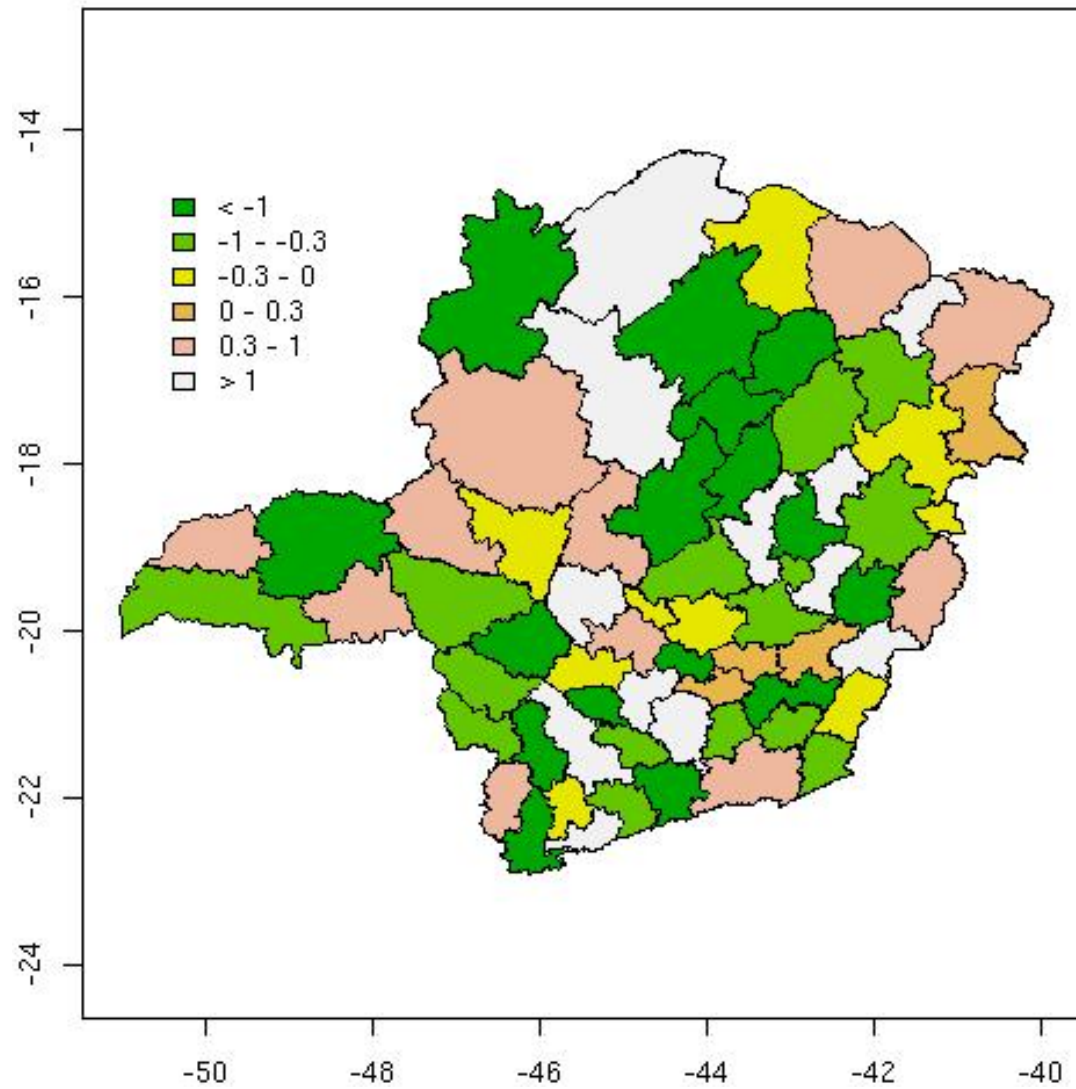
Exemplos: casual



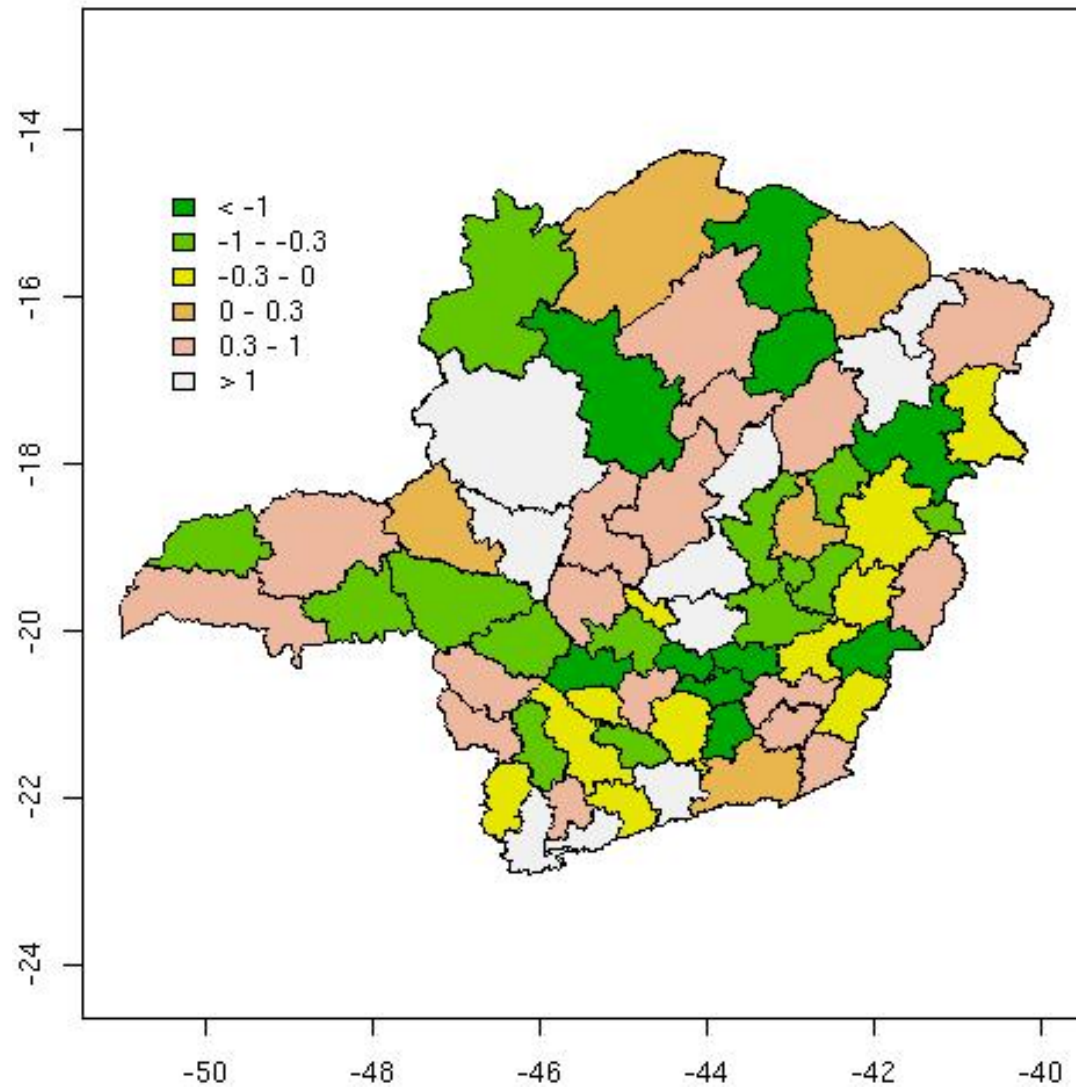
Exemplos: casual



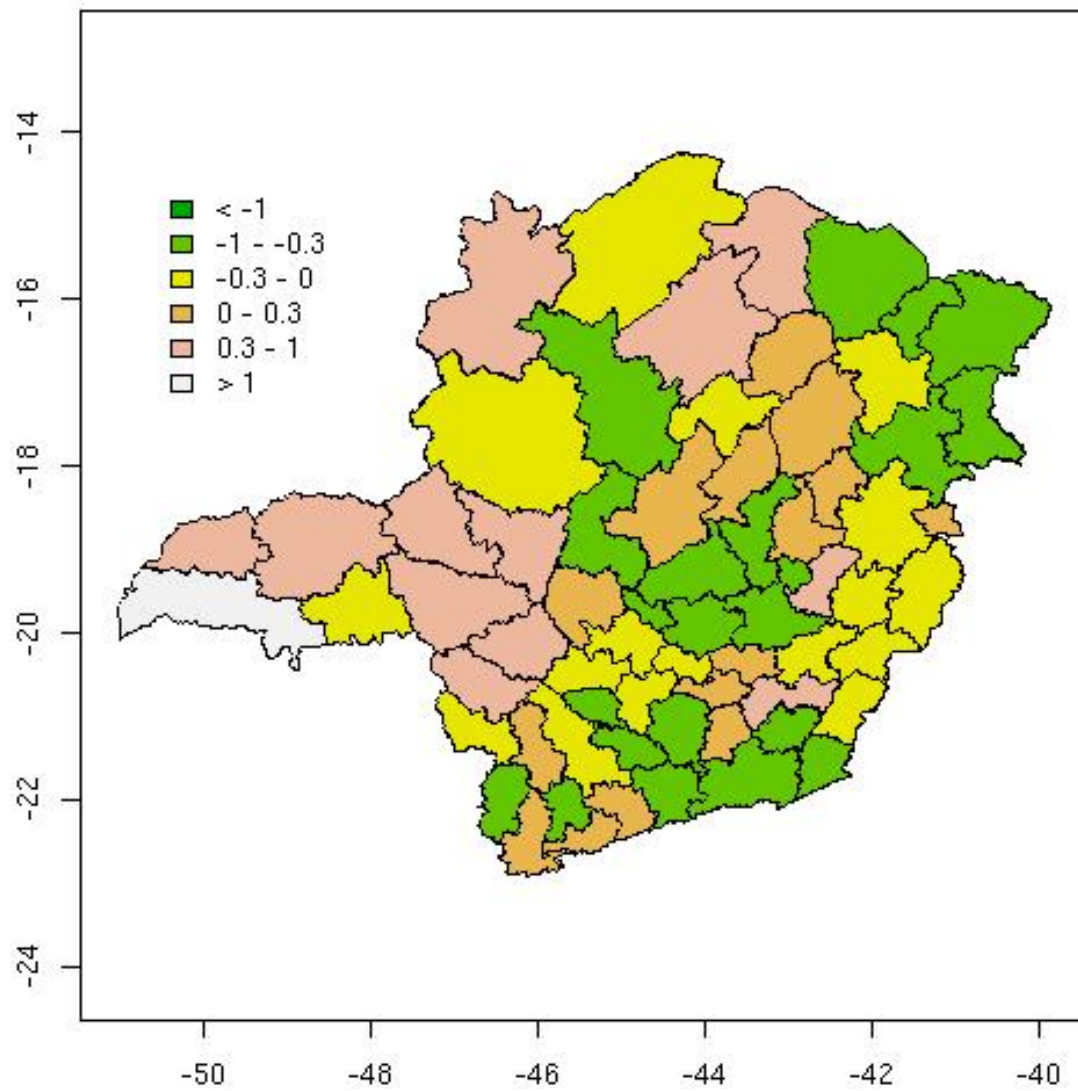
Exemplos: casual



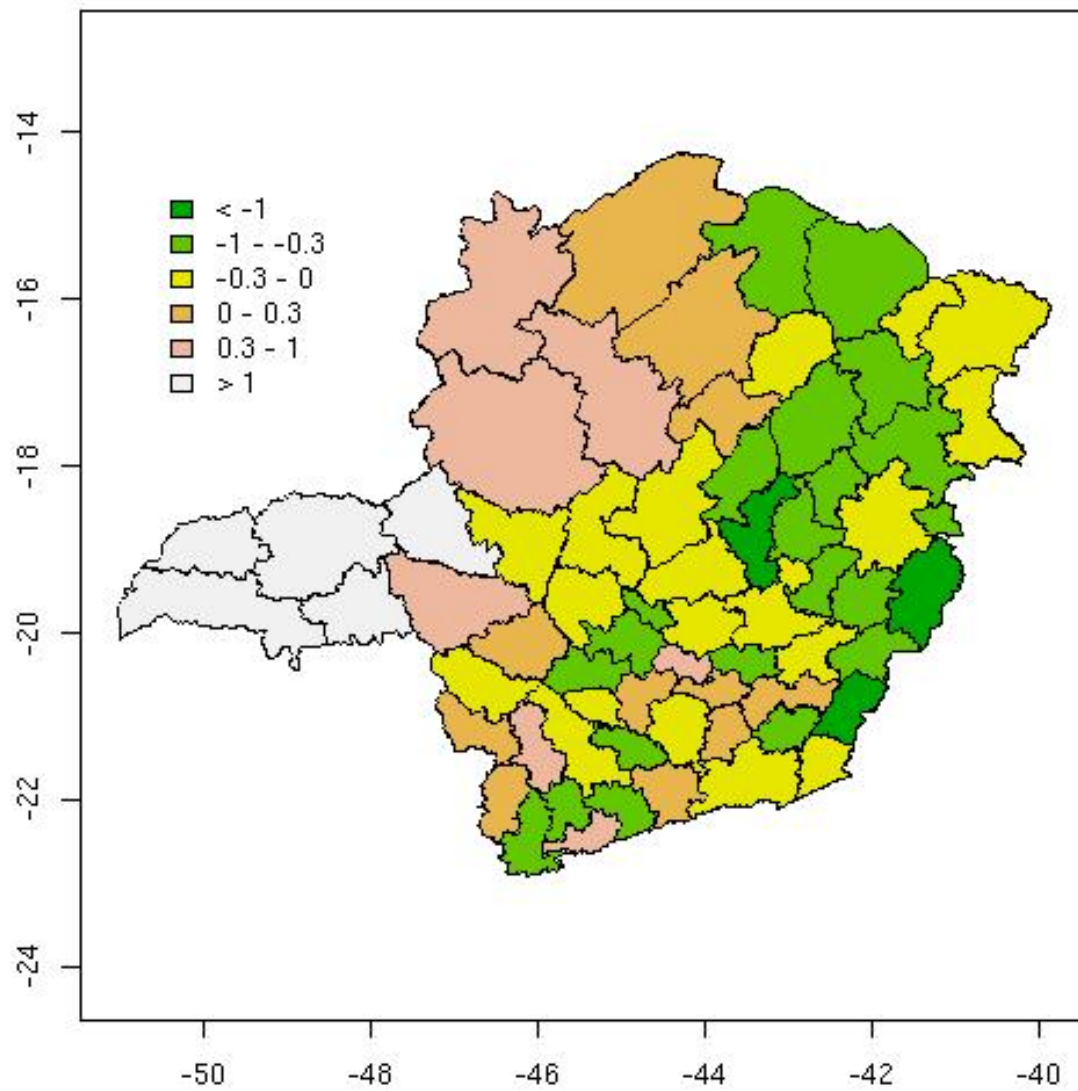
Exemplos: casual



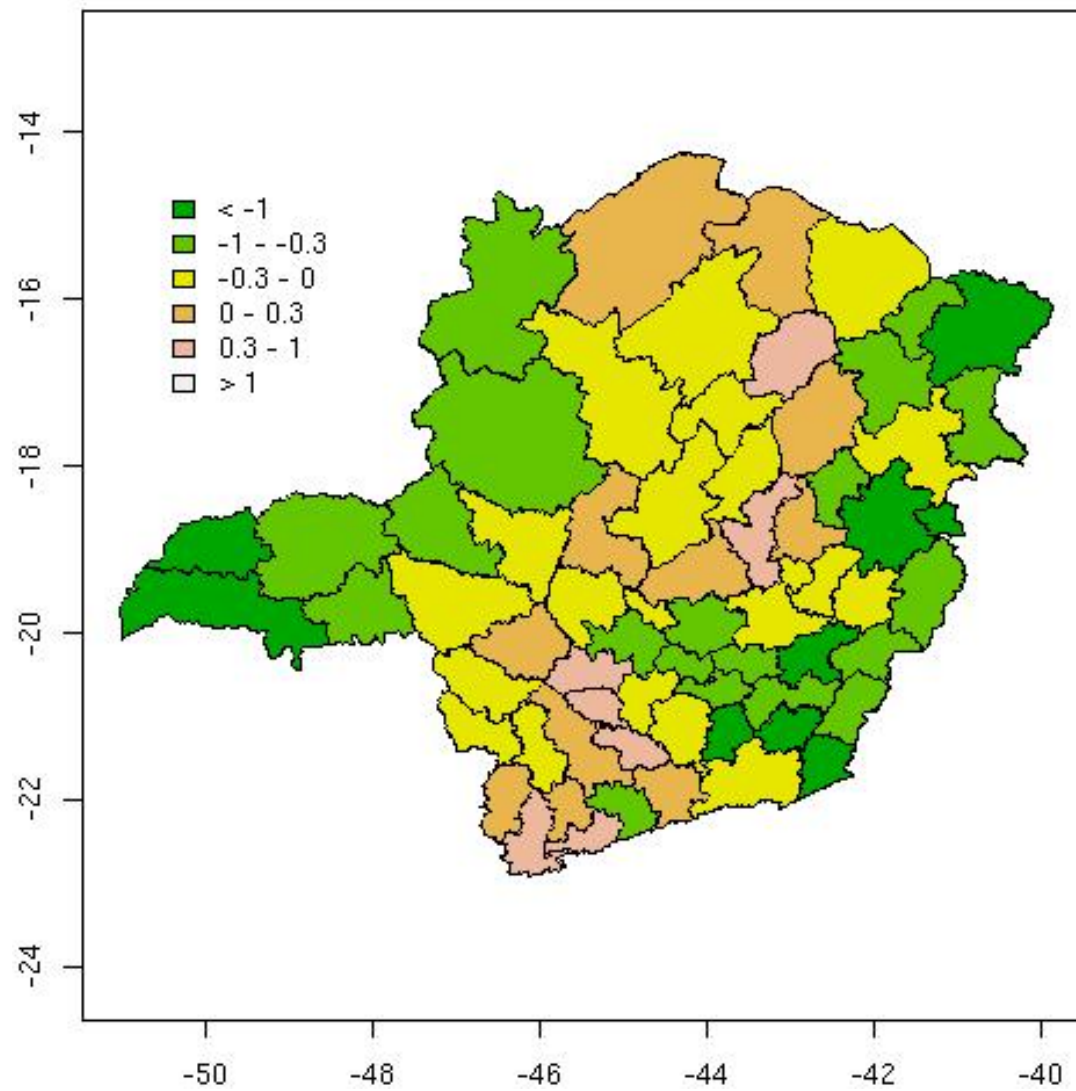
Exemplos: ICAR



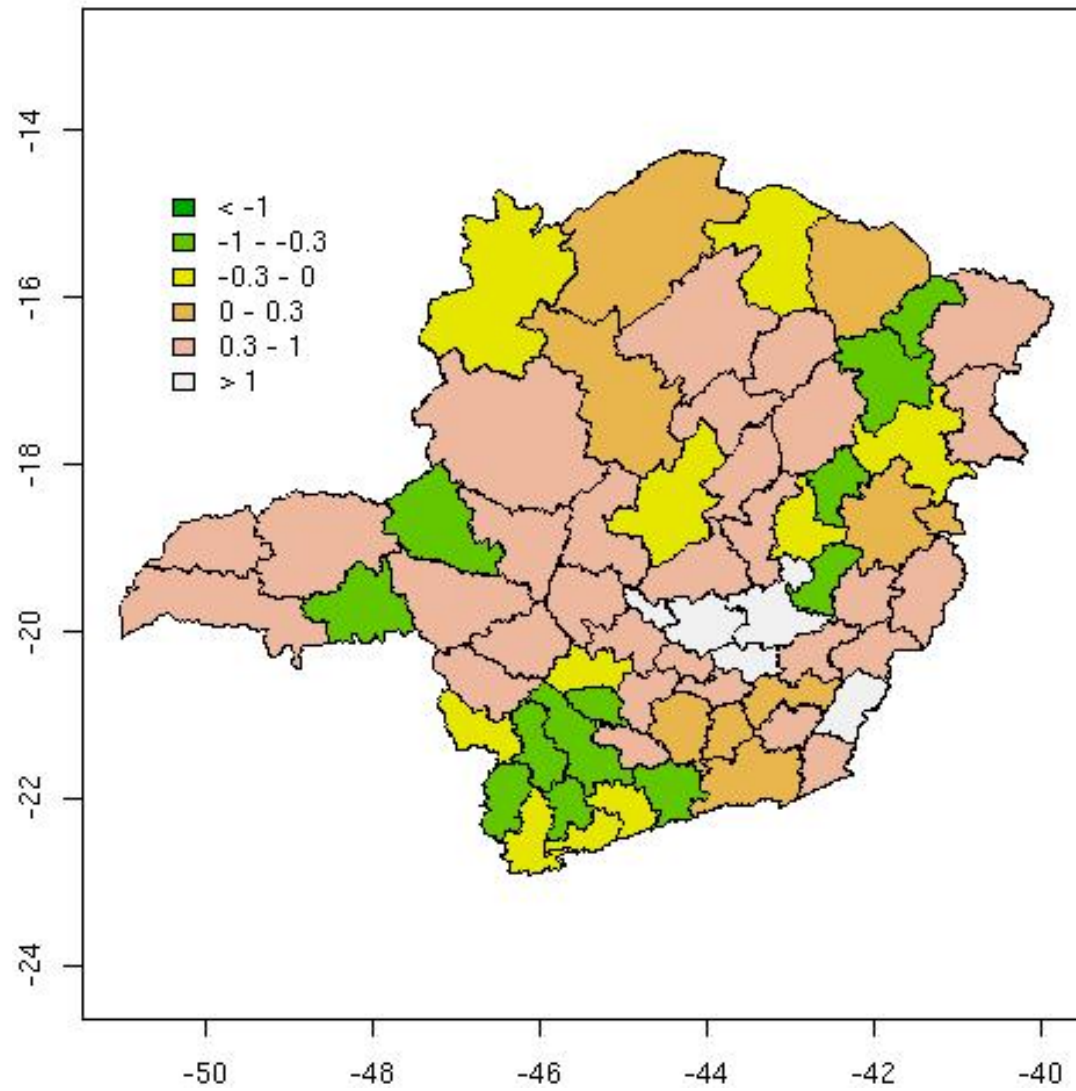
Exemplos: ICAR



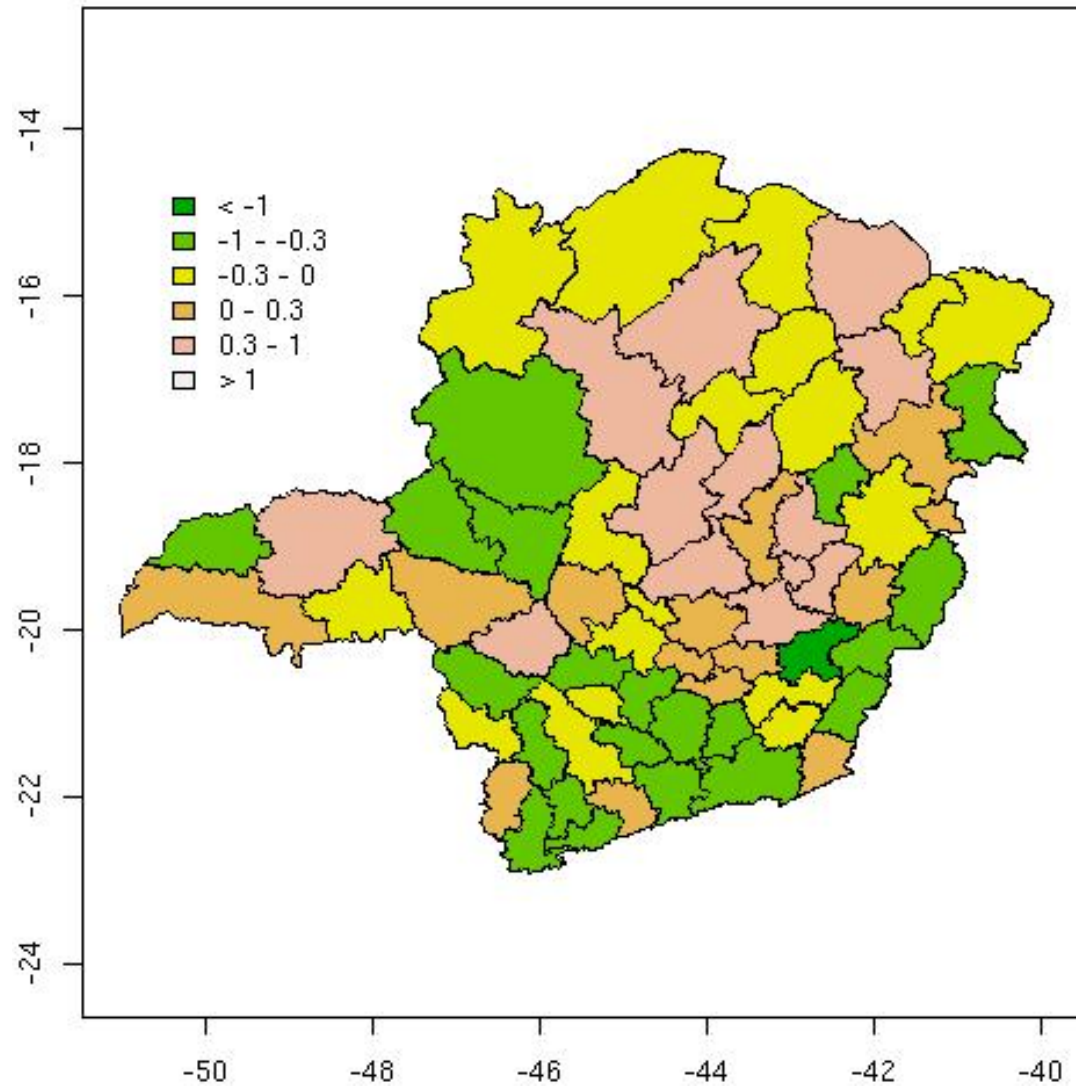
Exemplos: ICAR



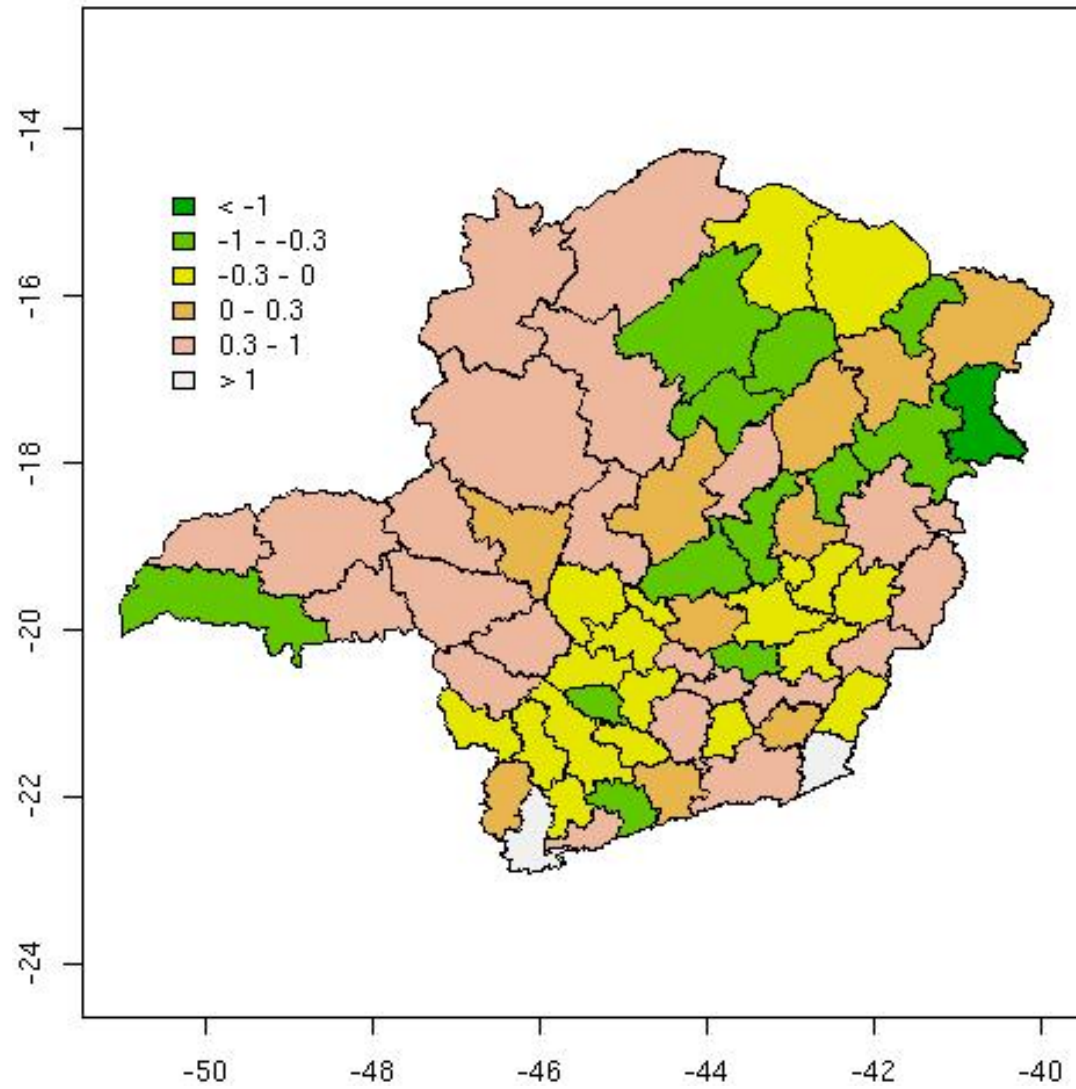
Exemplos: ICAR



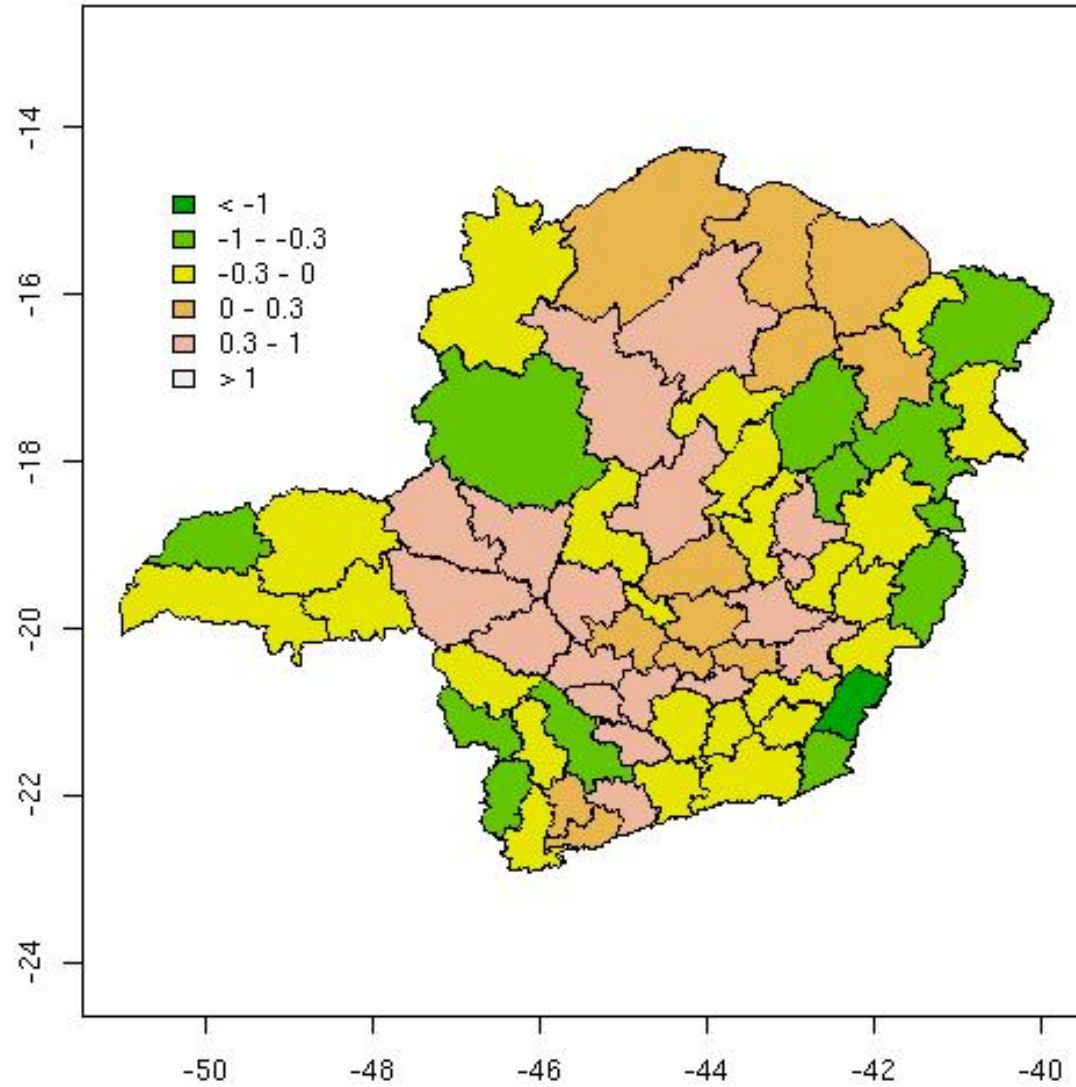
E este? (1)



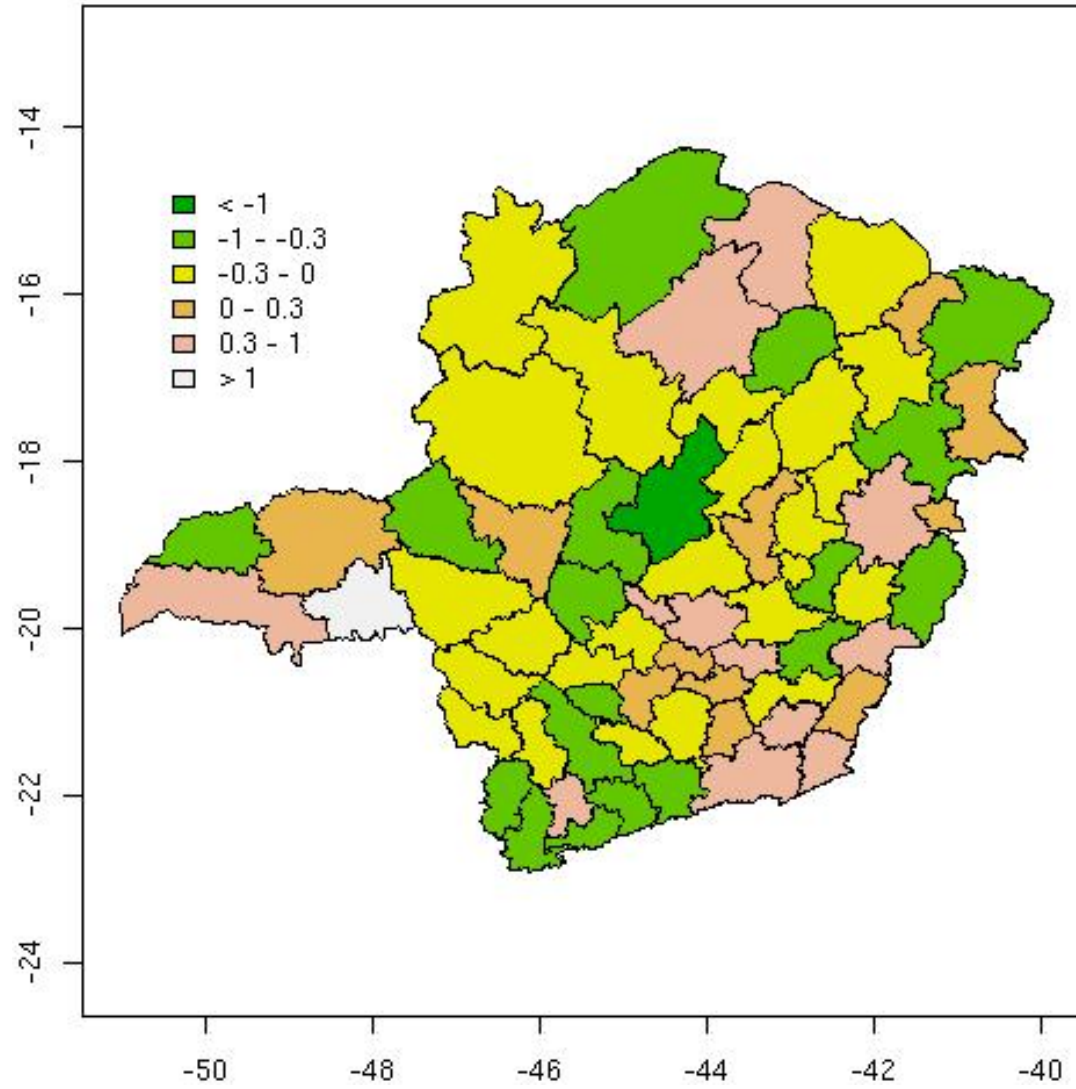
E este? (2)



E este? (3)



E este? (4)



Por onde começar?

A primeira coisa a se fazer análise espacial é verificar se é mesmo necessário fazer uma análise espacial!

Se o mecanismo gerador dos dados parece gerar padrões espaciais espúrios, qual o sentido de analisar espacialmente estes dados?

Estaremos procurando explicar o acaso, o eventual.

Análogo a achar que, ao ver que certas nuvens organizam-se em formas de animais, achar que existe um mecanismo que as organiza dessa forma e sair em busca desse mecanismo.

Existe a tendência humana a enxergar padrões onde não existe nenhum.

Hipóteses vagamente definidas

A tarefa número 1 então é fazer um teste de hipóteses.

Consideramos duas situações possíveis para o mecanismo gerador dos dados que observamos no mapa.

A hipótese nula diz que os dados y_1, \dots, y_n foram jogados no mapa de forma totalmente casual, de forma "cega" em relação à localização espacial das áreas.

A hipótese alternativa diz que existe um mecanismo alocando os dados às áreas que estimula áreas próximas a terem valores mais altos que o valor médio global ou a terem valores mais baixos que o valor médio global.

OBS: A hipótese alternativa também pode incluir mecanismos diferentes do acima: um valor ALTO numa região estimular valores BAIXOS no seu redor. Este caso é menos comum na prática.

Decisões a partir de teste de hipóteses

A idéia de um teste de hipóteses é contrastar os dados com as situações típicas sob as duas hipóteses em consideração e decidir por uma delas.

Se concluirmos que os dados podem muito bem ter sido gerados sob a hipótese nula, então não faz sentido prosseguir com uma análise espacial.

Os aglomerados que vemos no mapa são perfeitamente explicáveis por um mecanismo casual de alocação de cores ao mapa.

Aceitamos a hipótese nula até evidência adicional futura vir a mudar nossa conclusão.

Como decidir?

Para decidir entre hipóteses precisamos de um pouco mais de estrutura matemática pois temos de considerar o que é um padrão espacial USUAL ou TÍPICO sob a hipótese nula.

Idéia básica (solução estatística):

usar um índice I que mede o grau de correlação espacial entre os valores do mapa

Isto é, o índice I vai medir quão parecidas são as áreas que estão localizadas próximas umas das outras.

Obter a distribuição desse índice sob a hipótese nula, ou seja, como o índice se comporta na ausência de padrão.

contrastar o comportamento do índice sob a hipótese de ausência de padrão com o valor do índice para os dados observados

na prática pode ser feito de duas formas: (i) usando a distribuição amostral do índice; (ii) por simulação

Como decidir? (cont.)

Se o valor do índice for tal que ele poderia muito bem ter acontecido se a hipótese nula fosse verdadeira, não poderemos descartar a hipótese nula como explicação para o padrão espacial.

Para isto, calculamos se o valor do índice I poderia acontecer facilmente se a hipótese nula fosse verdadeira. Isto é, vamos calcular o P-VALOR associado com o índice medido.

Se o P-VALOR for alto (maior que 5%), NÃO descartamos a hipótese nula e interrompemos a análise ESPACIAL (prossequimos com uma análise NÃO-ESPACIAL).

Se o P-VALOR for baixo (menor que 5%), DESCARTAMOS a hipótese nula e PROSSEGUIMOS com uma análise espacial.

Índices de auto-correlação espacial

Alguns índices para dados de área:

O mais famoso é o índice de Moran.

Outro índice é o de Geary, e existem outros

Geary é menos usado que o de Moran.

Os dois índices dependem da definição de uma matriz de vizinhança W .

os índices são casos particulares da medida de Mantel

Estes índices cumprem papel análogo a funções G , F e K de processos pontuais, ou variogramas (e envelopes) em dados geoestatísticos em detectar evidência da presença de padrão espacial

Índice de Moran

Como calcular:

$$I = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Pode ser escrito (decomposto) para cada unidade de área como:

$$I_i = \frac{(y_i - \bar{y}) \sum_{j=1}^n w_{ij} (y_j - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

I é o índice *global* e I_i é o índice *local*

Matriz de Vizinhaça W

Matriz W de dimensão $n \times n$.

Elemento W_{ij} representa o peso ou o grau de conectividade ou de proximidade espacial entre as áreas i e j .

Diagonal nula: $W_{ii} = 0$

Escolha de W_{ij} é arbitrária (dentro de certas condições).

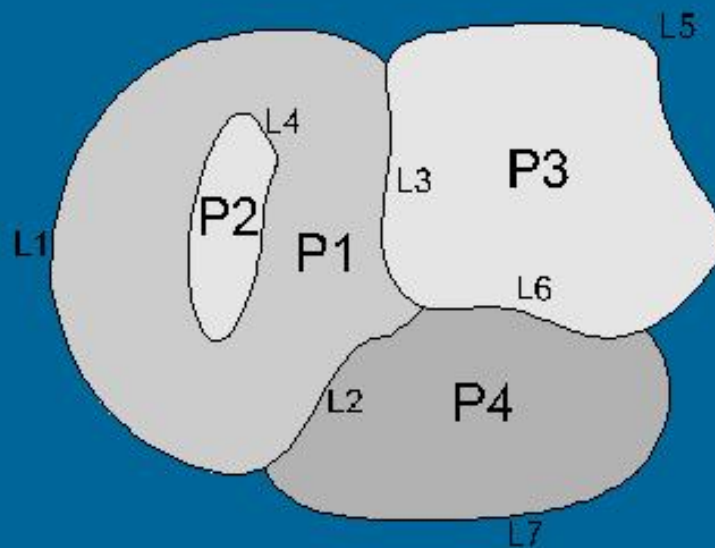
Especificação de W deveria ser feita levando em conta o problema específico sob análise.

Para mapas com muitas áreas, a matriz W é construída usando um Sistema de Informação Geográfico.

Armazenamento da matriz é feito usando uma codificação para matrizes esparsas (com muitos zeros) já que, tipicamente, mais de 95% dos elementos da matriz são iguais a zero.

De curso do INPE

Matriz de Proximidade



$$W = \begin{bmatrix} w_{11} & w_{12} & w_{13} & w_{14} \\ w_{21} & w_{22} & w_{23} & w_{24} \\ w_{31} & w_{32} & w_{33} & w_{34} \\ w_{41} & w_{42} & w_{43} & w_{44} \end{bmatrix}$$

w_{ij} : “distância” do objeto i ao objeto j .

Exemplos para W

Por definição, $W_{ii} = 0$, SEMPRE

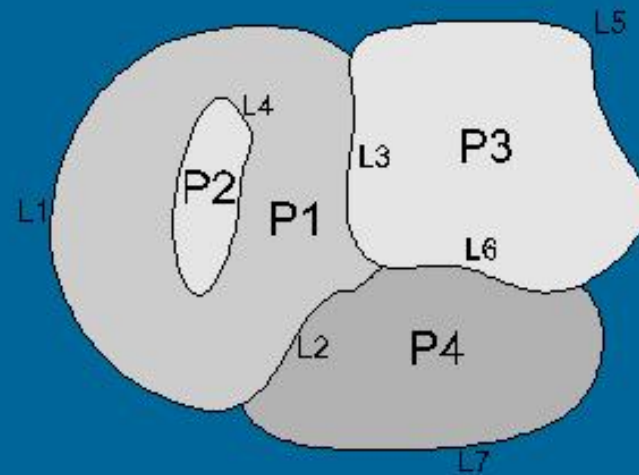
Contiguidade espacial:

$$W_{ij} = \begin{cases} 1, & \text{se áreas } i \neq j \text{ compartilham fronteiras} \\ 0, & \text{caso contrário} \end{cases}$$

W é simétrica: $W = W^t$

Exemplos de medidas

- Existência de fronteira comum.



$w_{14} = 1$ P1 faz fronteira com P4

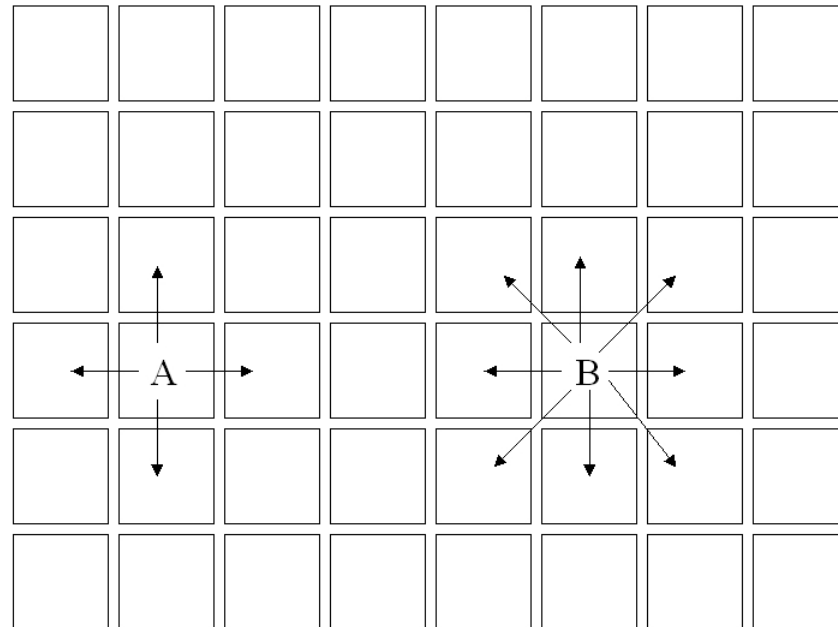
$w_{24} = 0$ P2 não tem fronteira com P4

Vizinhança por Contiguidade

Considere um reticulado regular

Dois tipos simples de vizinhança por contiguidade

A = movimento do peao, B = movimento do rei



Exemplos para W

Para cada área, associe uma posição no plano como sendo a "localização" da área.

Por exemplo, o centróide da área ou as coordenadas da sede de um município.

Seja d_{ij} a distância entre as "posições" das áreas i e j .

Então

$$W_{ij} = \begin{cases} g(d_{ij}), & \text{se } d_{ij} < \delta \\ 0, & \text{caso contrário} \end{cases}$$

Exemplos para W

Opções típicas para $g(d_{ij})$ se $d_{ij} < \delta$:

Step-function: $g(d_{ij}) = 1$

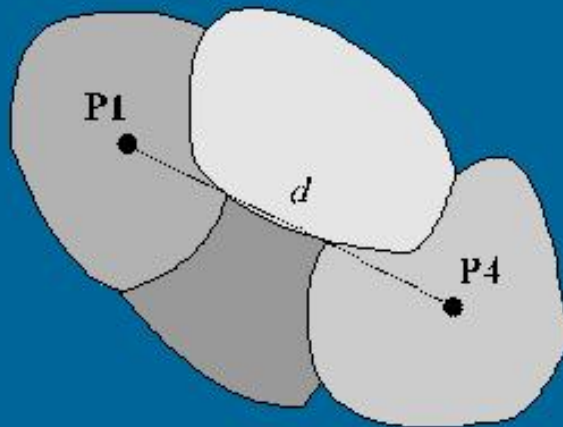
Inverso da distância: $g(d_{ij}) = 1/d_{ij}^\alpha$

Deacimento exponencial: $g(d_{ij}) = e^{-\alpha d_{ij}}$

Tipicamente, o parâmetro α é conhecido e, em geral, fixado a priori como sendo 1 ou 2.

Exemplos de medidas

- distância linear entre centróides dos objetos.



$$w_{14} = 0 \quad \text{para } d > \text{limiar}$$

$$w_{14} = 1 \quad \text{para } d \leq \text{limiar}$$

- inverso da distância linear.

$$w_{14} = \frac{1}{d}$$

Ainda mais um exemplo para W

l_{ij} = comprimento da fronteira comum entre as áreas i e j

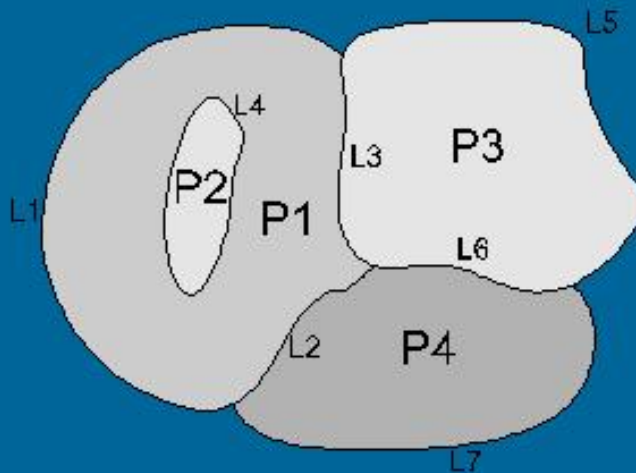
l_i = o perímetro da área i

$$W_{ij} = l_{ij}/l_i$$

$W_{ij} \neq W_{ji}$, em geral.

Exemplos de medidas

- proporção da fronteira pelo perímetro.



$$w_{14} = \frac{L_2}{L_1 + L_2 + L_3 + L_4}$$

- $w_{ij} \neq w_{ji}$

- “média ponderada!”

Padronizando W

Às vezes, matriz W é linha-padronizada: linhas da matriz W somam 1 e podem ser vistos realmente como pesos entre 0 e 1 e somando 1.

Redefina a matriz de vizinhança como sendo W^* com

$$W_{ij}^* = W_{ij}/W_{i+}$$

onde $W_{i+} = \sum_k W_{ik}$.

Isto é, $W^* \mathbf{1} = \mathbf{1}$.

Esta padronização garante que $0 \leq W_{ij} \leq 1$

Permite também que parâmetros espaciais de diferentes modelos sejam comparáveis entre si: parâmetros espaciais ρ deverão ser $< 1/\lambda_{\max} = 1$. No entanto, não é simples interpretar estes parâmetros (ver Melanie Wall, 2004).

Padronização pode tornar W^* não simétrica mesmo quando W é simétrica.

Variáveis filtradas ou defasadas espacialmente

Considere a matriz $n \times n$ de vizinhança espacial W com $W_{ii} = 0$

Vamos usar \mathbf{Y} defasada espacialmente

$$W\mathbf{Y} = (\bar{Y}_1, \dots, \bar{Y}_n)^t:$$

$$\begin{aligned}\bar{Y}_1 &= W_{11}Y_1 + W_{12}Y_2 + W_{13}Y_3 + \dots + W_{1n}Y_n \\ \bar{Y}_2 &= W_{21}Y_1 + W_{22}Y_2 + W_{23}Y_3 + \dots + W_{2n}Y_n \\ \bar{Y}_3 &= W_{31}Y_1 + W_{32}Y_2 + W_{33}Y_3 + \dots + W_{3n}Y_n \\ \dots & \qquad \qquad \qquad \dots \\ \bar{Y}_n &= W_{n1}Y_1 + W_{n2}Y_2 + W_{n3}Y_3 + \dots + W_{nn}Y_n\end{aligned}$$

\bar{Y}_i é a média ponderada das observações nas áreas vizinhas à área i .

Nesta média \bar{Y}_i , o peso dado à área j é W_{ij} .

Se i e j não são vizinhos, o peso W_{ij} é igual a zero.

Quando W é padronizada?

Na discussão sobre o índice de Moran, a matriz W pode ser linha-padronizada ou não-padronizada.

Os resultados que vamos apresentar valem para os dois casos, linha-padronizada ou não-padronizada.

Quando estivermos lidando com uma matriz linha-padronizada nós deixaremos isto explícito.

Do curso do INPE

Índice Global de Moran

$$I = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

onde:

- n corresponde ao número de áreas,
- y_i é o valor do atributo considerado na área i ,
- \bar{y} representa o valor médio do atributo na região de estudo,
- w_{ij} são os pesos atribuídos conforme a conexão entre as áreas i e j .

Do curso do INPE

Indicadores Globais Moran

$$I = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

- Qual o significado do índice global de Moran (I) ?
- Como interpretar a equação acima ?
- Qual sua significância ou validade estatística ? Como avaliar ?

Do curso do INPE

Índice Globais de Moran

$$I = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

- É análogo ao coeficiente de correlação convencional, porque têm em seu numerador um termo que é produto de momento.
- Como um coeficiente de correlação, os valores de I também variam de -1 a +1, quantificando o grau de autocorrelação espacial existente.
 - ⇒ -1 autocorrelação espacial negativa ou inversa.
 - ⇒ 0 significa aleatoriedade
 - ⇒ +1 significa autocorrelação espacial positiva ou direta.

Do curso do INPE

Indicadores Globais de Autocorrelação Espacial

- Consideremos o exemplo que segue:



Matriz de Proximidade

	A	B	C	D
A	0	1	1	0
B	1	0	1	1
C	1	1	0	1
D	0	1	1	0

$$\text{Média } \bar{y} = \frac{20 + 15 + 24 + 5}{4} = 16$$

$$\text{Variância } \sigma^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n} = \frac{(20 - 16)^2 + (15 - 16)^2 + (24 - 16)^2 + (5 - 16)^2}{4} = 50,5$$

$$\text{Desvio Padrão } \sigma = \sqrt{\sigma^2} = \sqrt{50,5} = 7,1063$$

Do curso do INPE

Indicadores Globais de Autocorrelação Espacial

$$I = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})^2}$$



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

- A equação de I pode ser simplificada [padronizando y] e alterando W , de forma que a soma dos elementos de cada linha seja igual a 1.

$$z_i = \frac{(y_i - \bar{y})}{s}$$

	A	B	C	D
A	0	1	1	0
B	1	0	1	1
C	1	1	0	1
D	0	1	1	0



	A	B	C	D
A	0	1/2	1/2	0
B	1/3	0	1/3	1/3
C	1/3	1/3	0	1/3
D	0	1/2	1/2	0

Do curso do INPE

Indicadores Globais de Autocorrelação Espacial



	A	B	C	D
A	0	1/2	1/2	0
B	1/3	0	1/3	1/3
C	1/3	1/3	0	1/3
D	0	1/2	1/2	0

$$\bar{y} = 16,0$$

$$s = 7,1063$$

$$z_i = \frac{(y_i - \bar{y})}{\sigma}$$

$$z_A = 0,5628$$

$$z_B = -0,1407$$

$$z_C = 1,1257$$

$$z_D = -1,5479$$

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

$$\begin{matrix}
 w_{ij} & * & z_i z_j & = & M_{ij} \\
 \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix} & * & \begin{bmatrix} 0,3167 & -0,0792 & 0,6335 & -0,8711 \\ -0,0792 & 0,0197 & -0,1583 & 0,2177 \\ 0,6335 & -0,1583 & 1,2672 & -1,7424 \\ -0,8711 & 0,2177 & -1,7424 & 2,3959 \end{bmatrix} & = & \begin{bmatrix} 0 & -0,0396 & 0,3167 & 0 \\ -0,0264 & 0 & -0,0527 & 0,0725 \\ 0,2111 & -0,0527 & 0 & -0,5808 \\ -0,00 & 0,1088 & -0,8712 & 0 \end{bmatrix}
 \end{matrix}$$

$$\sum_{i=1}^n \sum_{j=1}^n w_{ij} z_i z_j = \sum_{i=1}^n \sum_{j=1}^n M_{ij} = -0,9143$$

$$\sum_{i=1}^n z_i^2 = 4$$

$$I = \frac{-0,9143}{4} = -0,2286$$

Significância do Índice de Moran

Avaliação da significância do índice de Moran (I).

Para estimar a significância de I , será preciso associar a este uma distribuição estatística, para tanto, duas abordagens são possíveis:

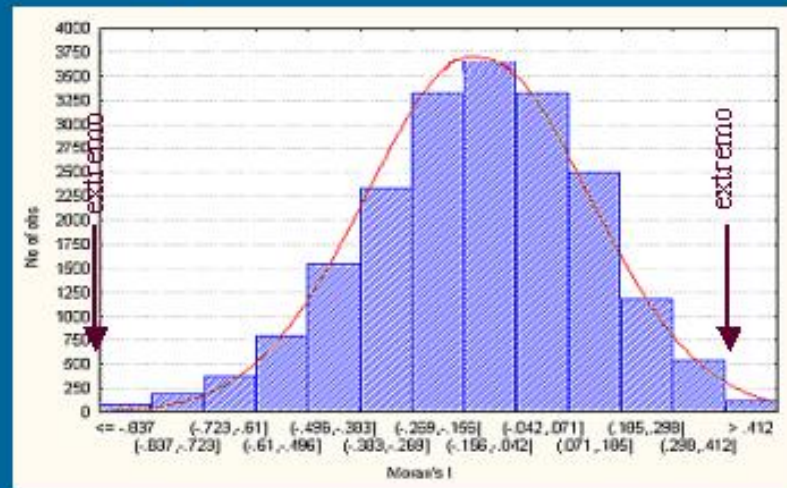
- Teste de pseudo-significância (experimento aleatório).
- Distribuição aproximada (hipótese da normalidade).

Do curso do INPE

Indicadores Globais de Autocorrelação Espacial

- A validade estatística do índice de Moran (I) sob o teste de pseudo-significância.

Distribuição simulada



- Se o índice I efetivamente medido corresponder a um “**extremo**” da distribuição simulada, então trata-se de evento com significância estatística.

I de Moran: Validade Estatística

- Se n é grande e as observações possuem distribuição NORMAL, então o índice I tem uma distribuição amostral que é aproximadamente normal, dada por:

$$E(I) = \frac{1}{(n-1)}$$

$$\sigma^2 = \frac{n^2(n-1)S_1 - n(n-1)S_2 - 2S_0^2}{(n+1)(n-1)^2 S_0^2}$$

Índice Moran Normalizado

$$I_N = \frac{I - E(I)}{\sigma}$$

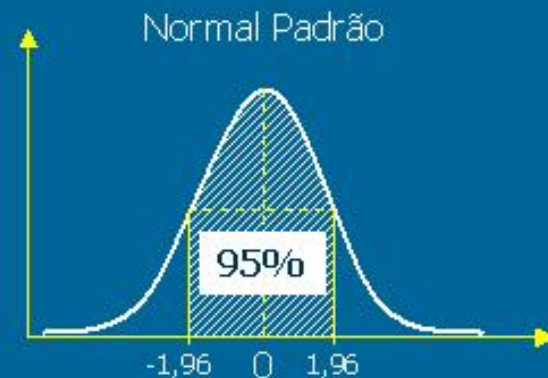
onde:

n = número de regiões,

$$S_0 = \sum \sum w_{ij} \text{ para } i \neq j$$

$$S_1 = \sum \sum (w_{ij} + w_{ji})^2 \text{ para } i \neq j$$

$$S_2 = \sum (\sum w_{ij} + \sum w_{ji})^2 \text{ para } i \neq j$$



Acontece que nem sempre y será normal...

Nem sempre as observações y serão normais e a aproximação anterior NÃO pode ser usada .

Por exemplo, y pode ser binária, com apenas dois valores: 0 ou 1.

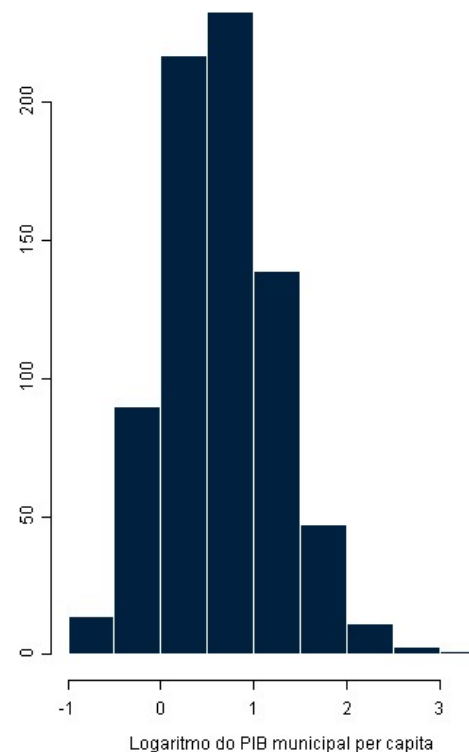
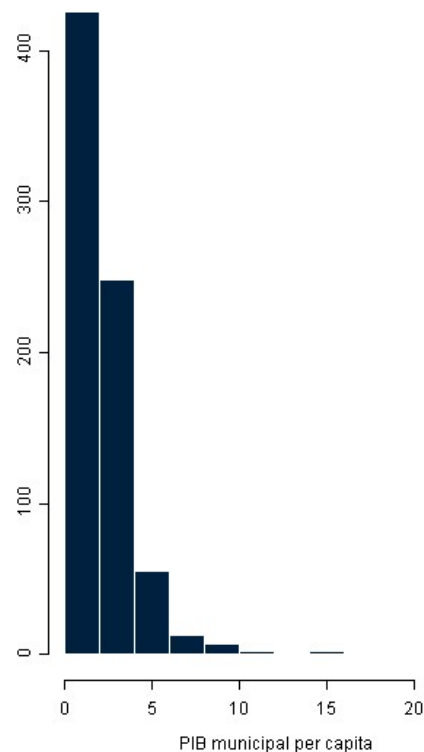
Uma variável binária poderia estar apenas indicando se a área i possui ou não possui certo atributo.

Pode ser uma informação mais fácil de ser obtida do que mensurar exatamente o valor de certa variável:.

Por exemplo, área i possui área verde ou não; área i teve pelo menos um incêndio florestal no último ano.

E y pode ser assimétrica...

Muitas vezes, as observações y poderão ter distribuição assimétrica (e portanto, não-normais) como no PIB per capita em municípios mineiros (ou seu log).



Teste de permutação

Nossa hipótese nula é a que as variáveis aleatórias Y_1, \dots, Y_n são permutáveis (exchangeable, em inglês).

Isto é, se não existe autocorrelação espacial então toda possível alocação das variáveis Y_1, \dots, Y_n às áreas do mapa seria igualmente provável.

Ou seja, se $(1), \dots, (n)$ indica uma permutação qualquer dos índices $1, \dots, n$, então a distribuição conjunta de Y_1, \dots, Y_n e a de $Y_{(1)}, \dots, Y_{(n)}$ são idênticas.

vetor permutável não implica ter componentes i.i.d. mas o contrário é válido: se Y_1, \dots, Y_n são i.i.d., então elas são permutáveis.

Teste de permutação - como fazer?

Calcule o índice I de Moran com os dados observados no mapa obtendo I^1 .

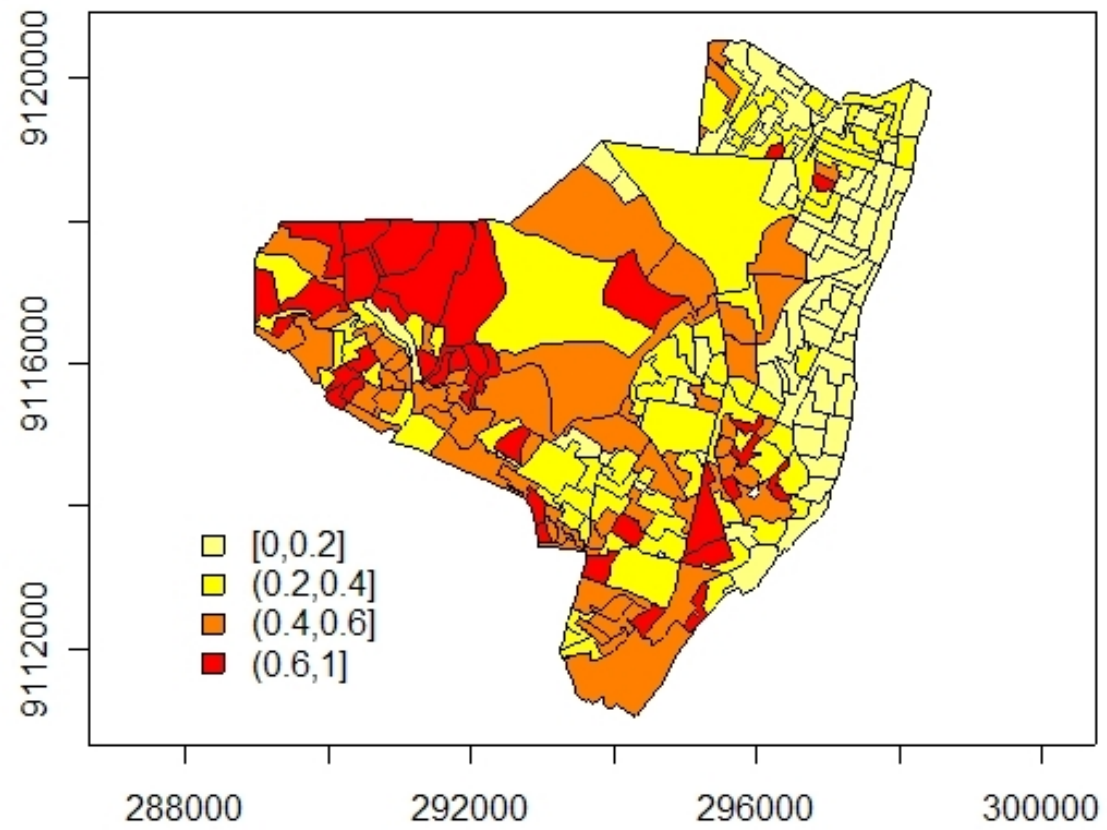
Gere um PSEUDO-MAPA permutando aleatoriamente os valores de y_i entre as áreas

Calcule o índice I no pseudo-mapa obtendo um valor para o índice de Moran $I^{(1)}$

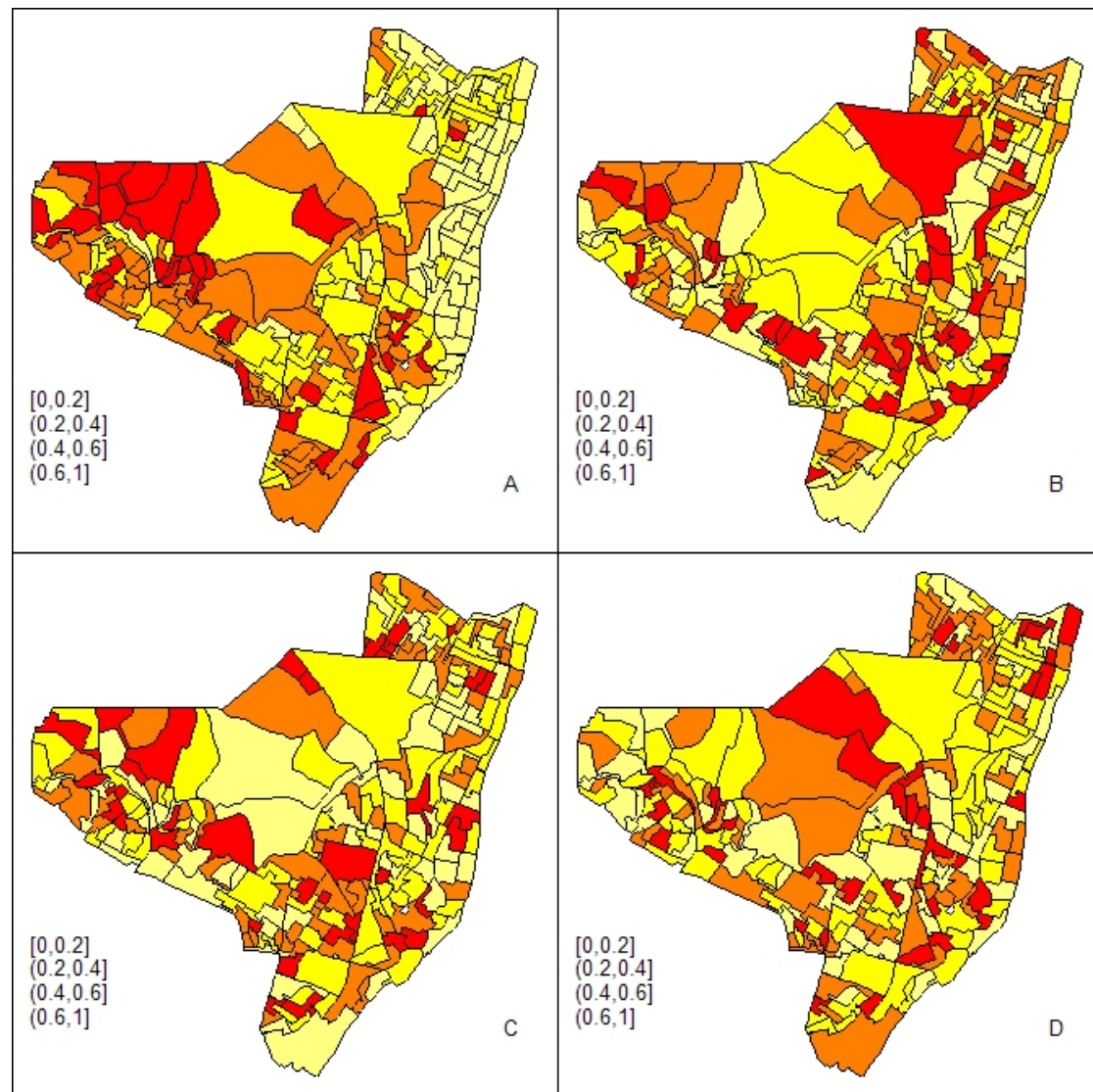
Gere outros 998 PSEUDO-MAPAS adicionais calculando sempre o índice de Moran em cada um deles.

Mapa original

Índice de Carência social



Mapa original e 3 pseudo-mapas



Teste de permutação: p-valor

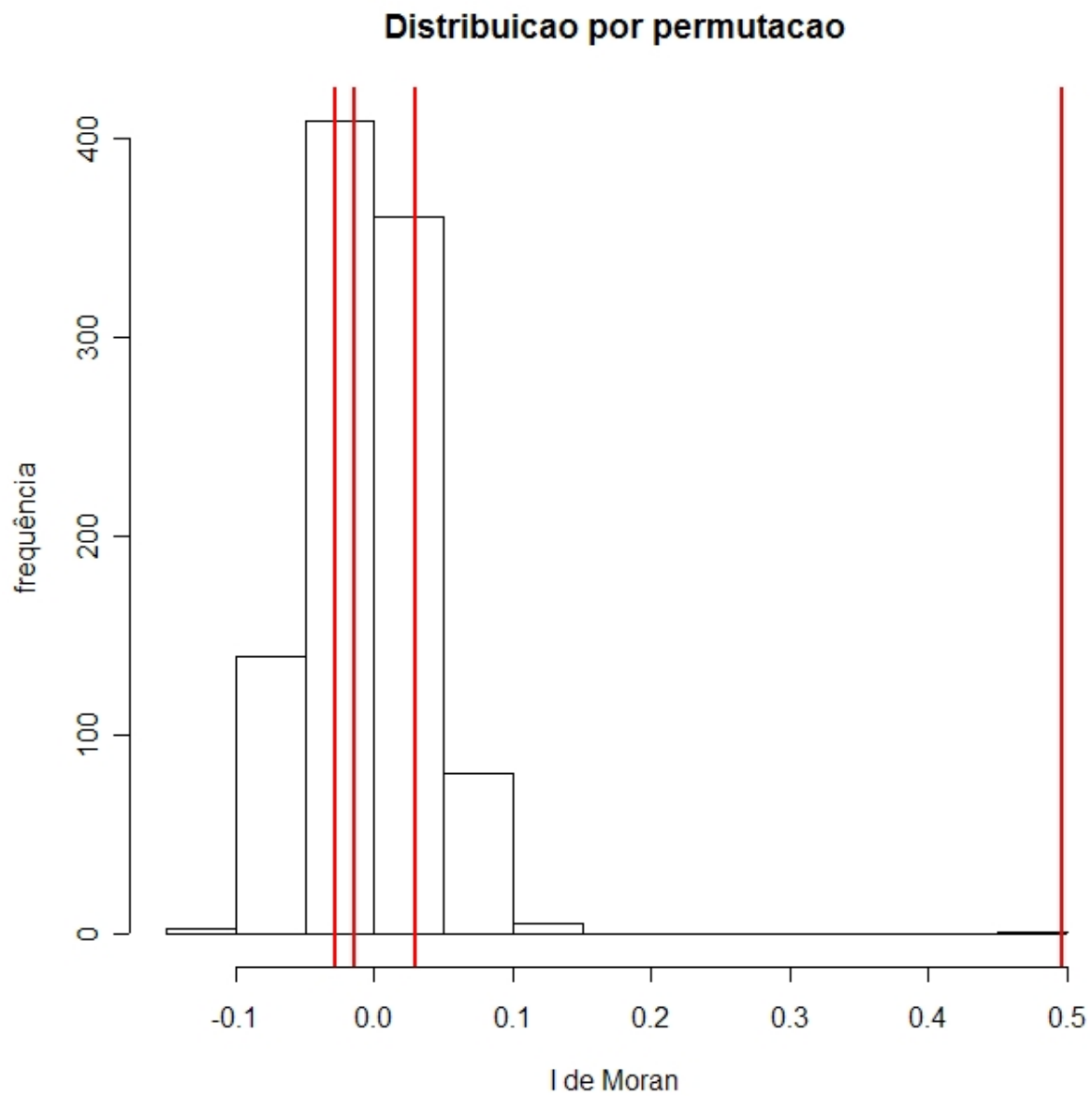
Temos o índice de Moran $I^{(1)}$ com os dados originais e mais 999 valores de I calculados com os pseudo-mapas.

Os 999 valores dos pseudo-mapas dão uma boa idéia do que são os valores usuais para I quando a hipótese nula é verdadeira.

Se o índice $I^{(1)}$ dos dados originais for similar aos 999 ele então poderia aparecer facilmente se os dados fossem alocados de forma completamente casual ao mapa.

É muito útil fazer um histograma com os valores gerados para situar quão comum ou quão raro (extremo) é o valor original $I^{(1)}$ do índice.

Moran de mapa original e 3 pseudo-mapas



Teste de permutação: p-valor

O p-valor do teste é a proporção dos índices dentre os 1000 valores que são maiores ou iguais ao valor original $I^{(1)}$ do índice.

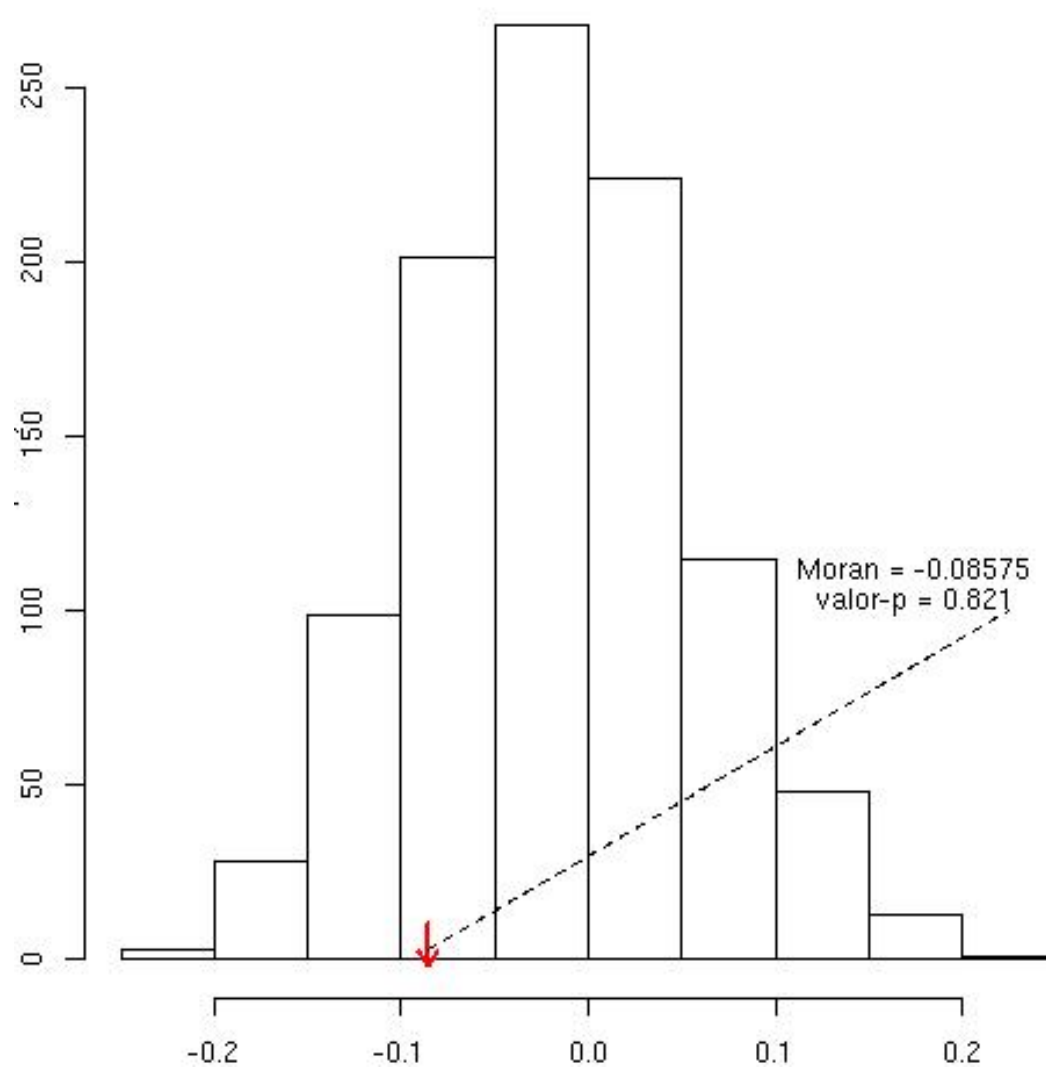
Isto é, o p-valor é igual a

$$\text{p-valor} = \frac{\text{número de } I\text{'s que são } \geq I^{(1)}}{1000}$$

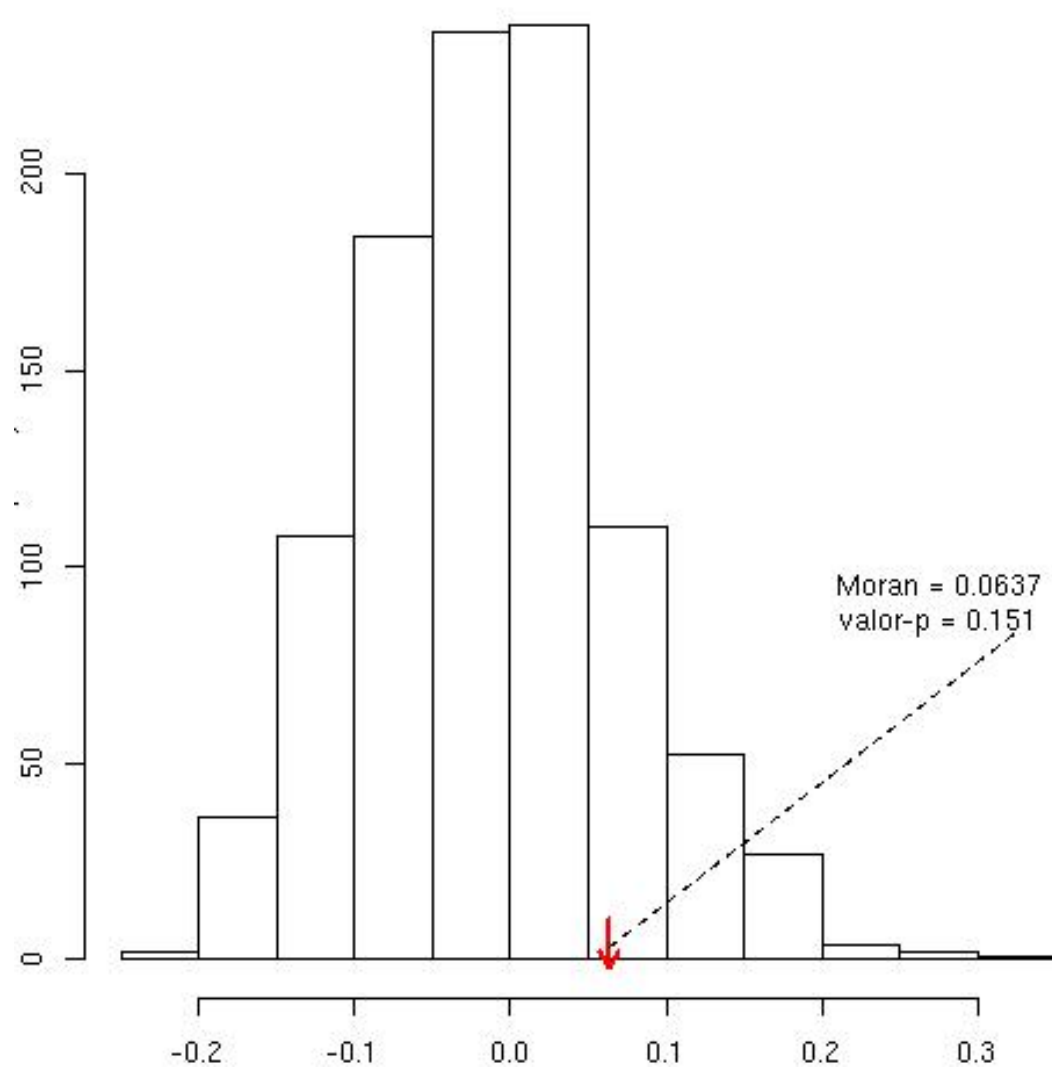
Observe que calculamos a proporção incluindo o próprio valor original do índice. Assim, o p-valor é, no mínimo, $1/1000$ pois $I^{(1)} \geq I^{(1)}$.

Rejeite a hipótese nula num teste de nível 5% se $p\text{-val} < 0.05$.

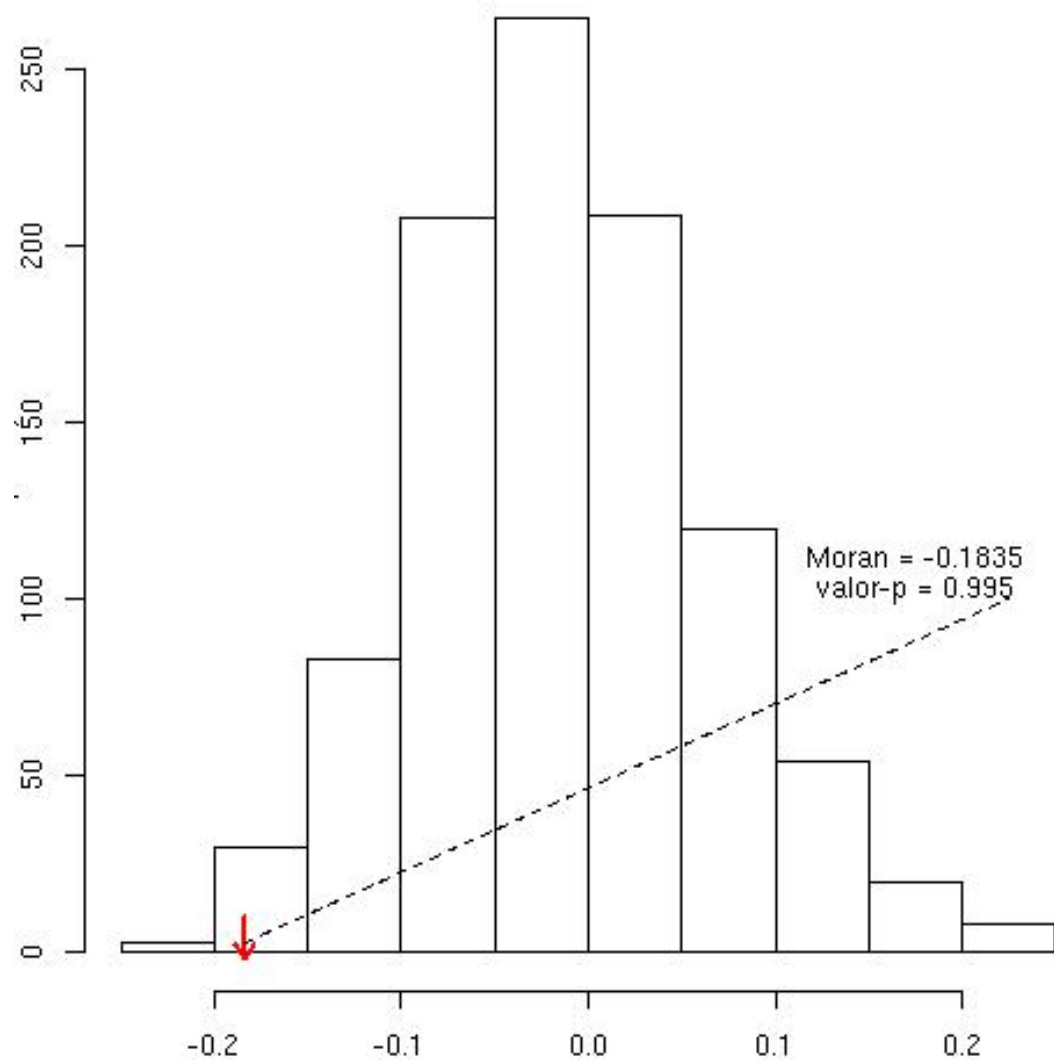
Voltando aos mapas de MG: aleatório



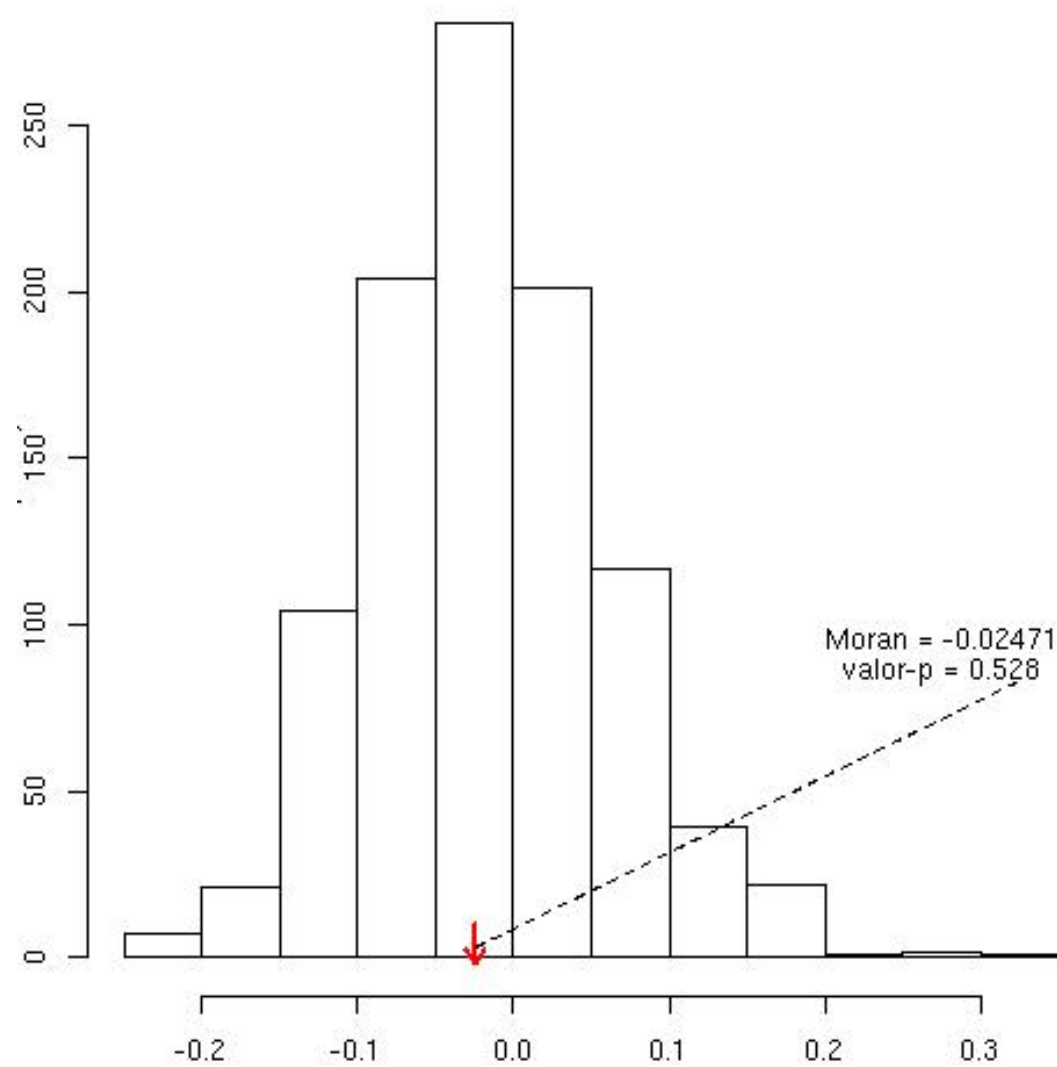
Voltando aos mapas de MG: aleatório



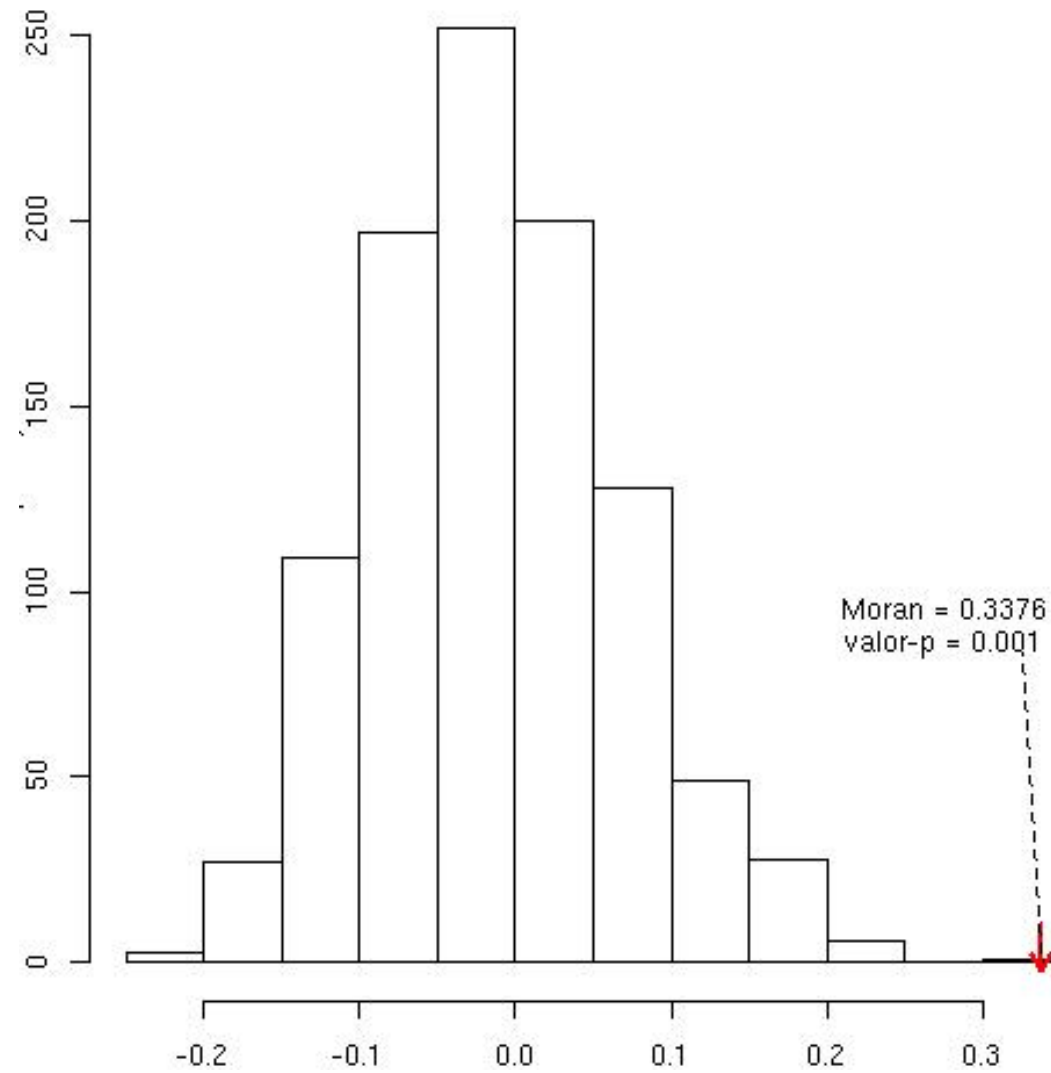
Voltando aos mapas de MG: aleatório



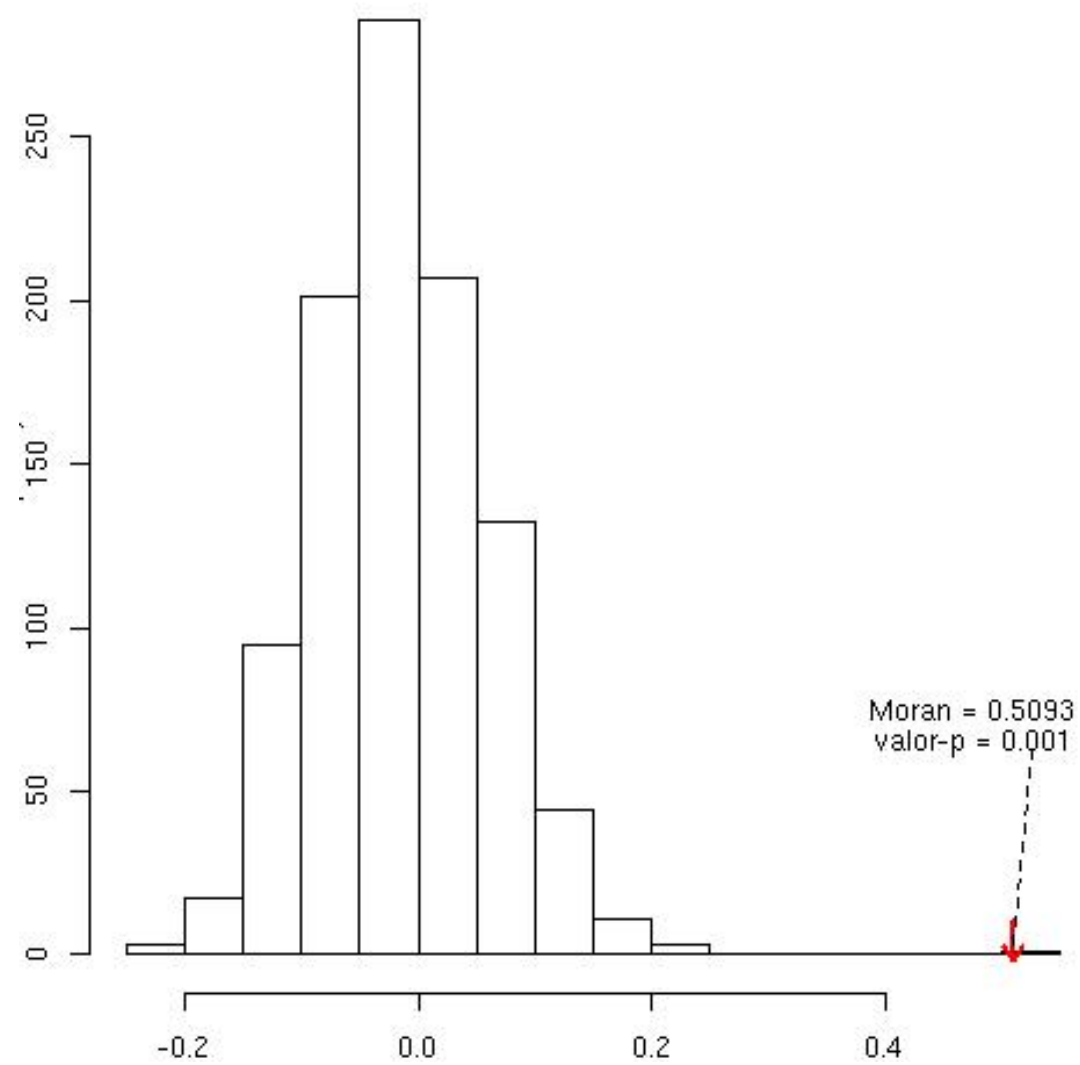
Voltando aos mapas de MG: aleatório



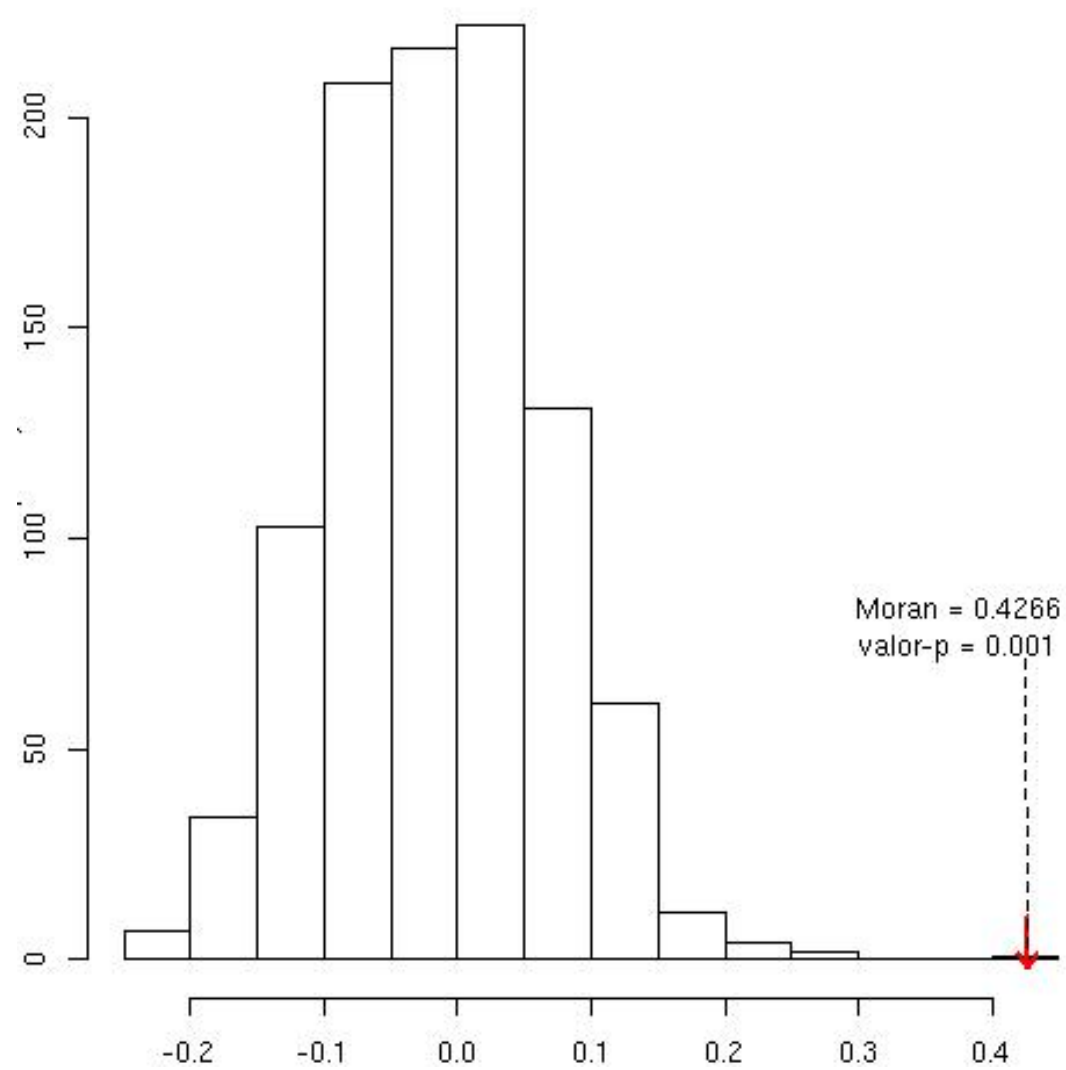
Voltando aos mapas de MG: ICAR



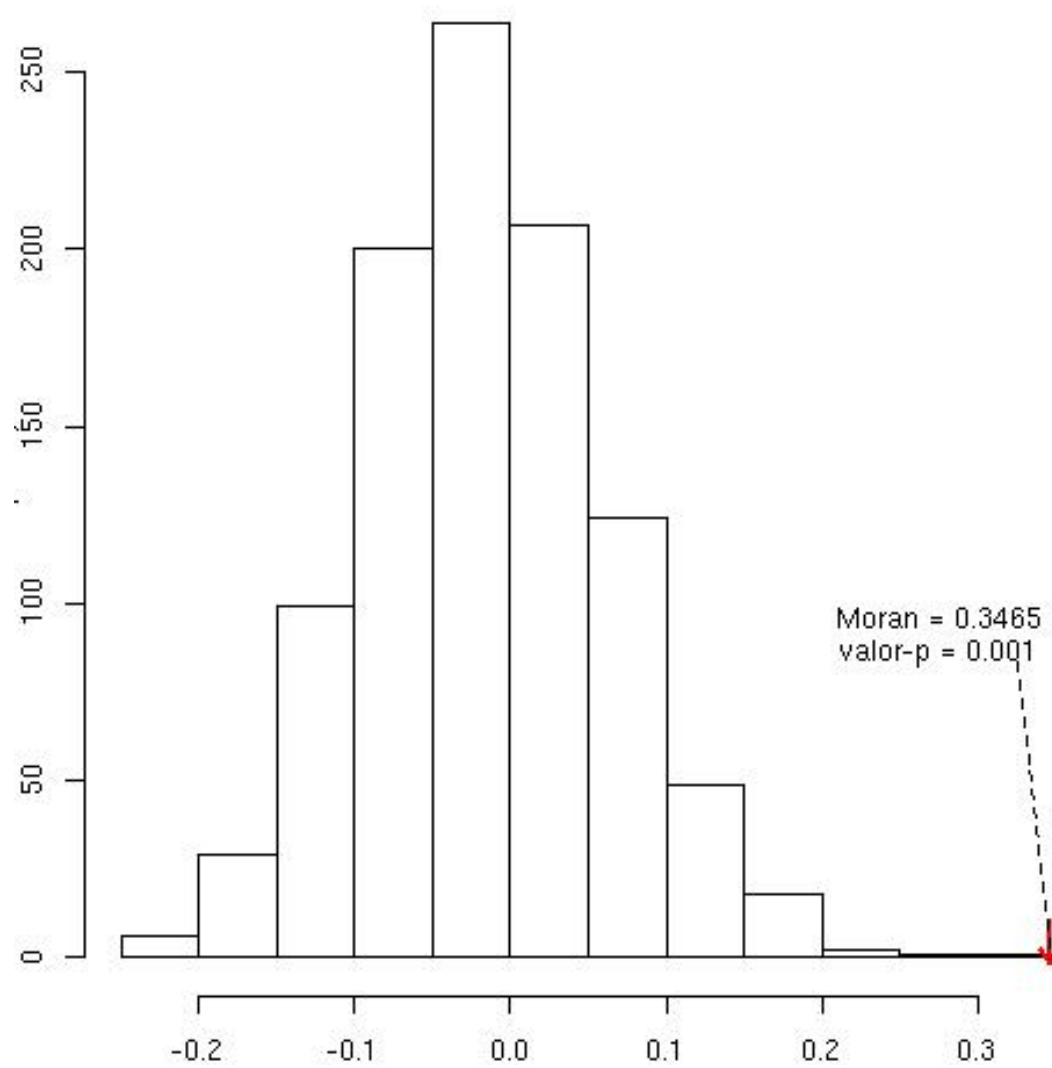
Voltando aos mapas de MG: ICAR



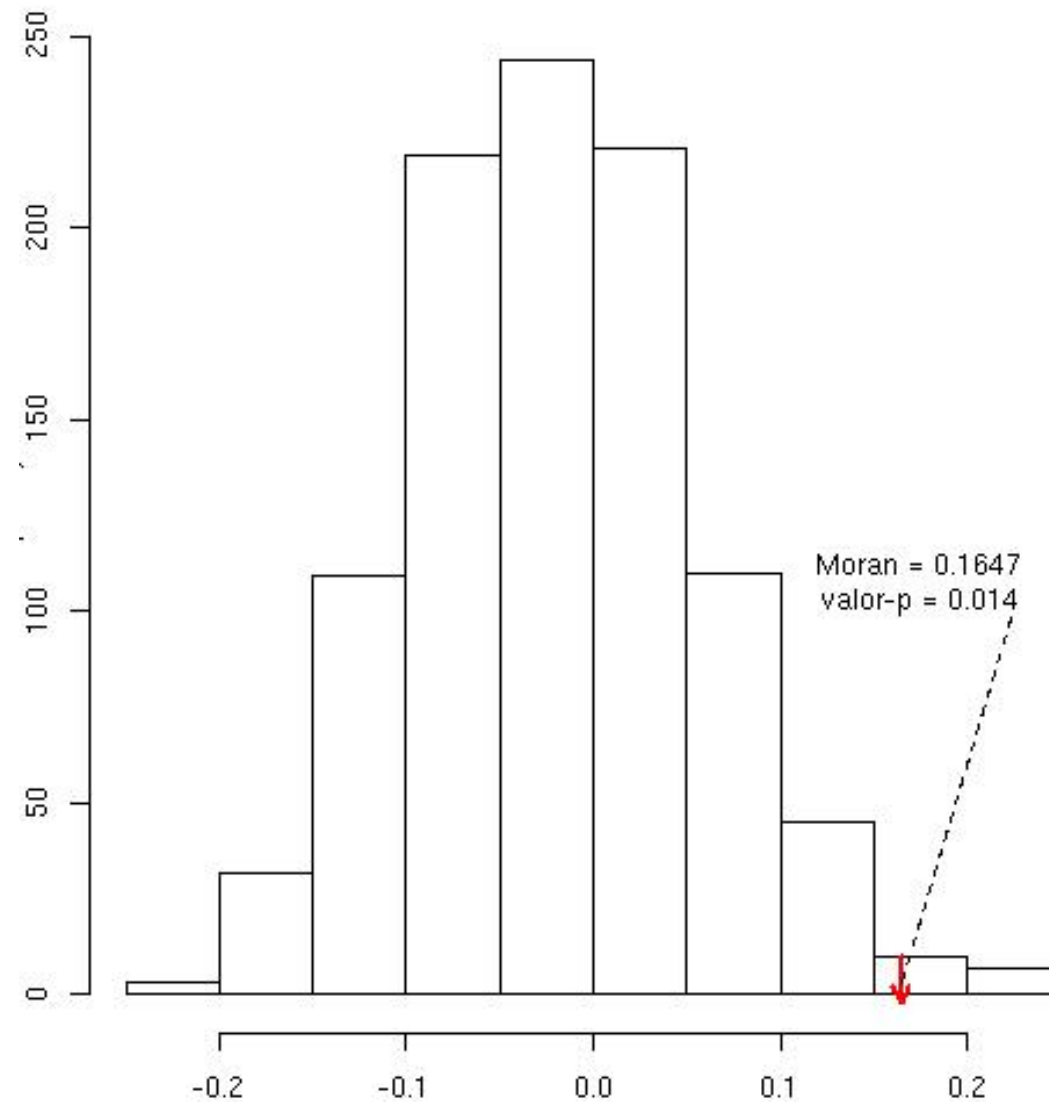
Voltando aos mapas de MG: ICAR



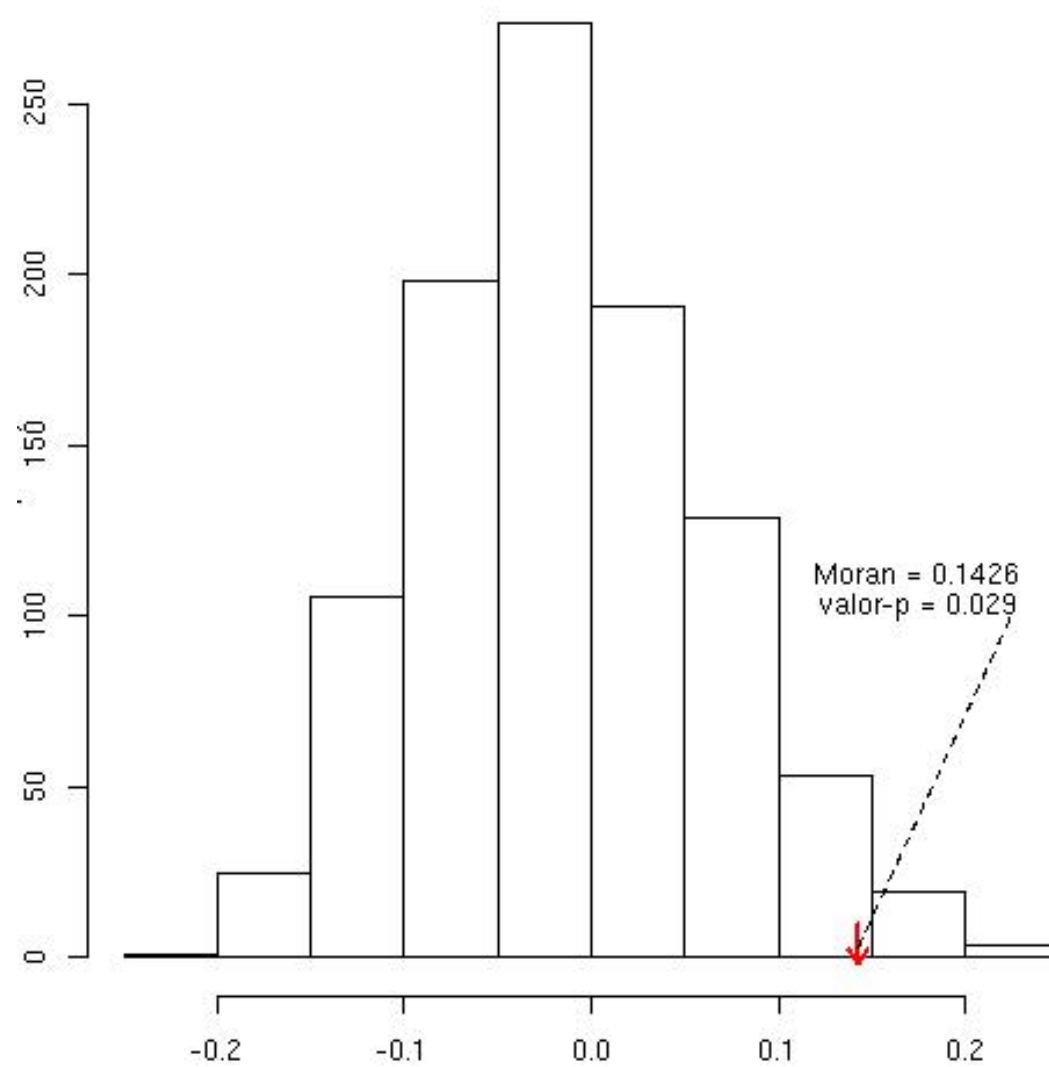
Voltando aos mapas de MG: ICAR



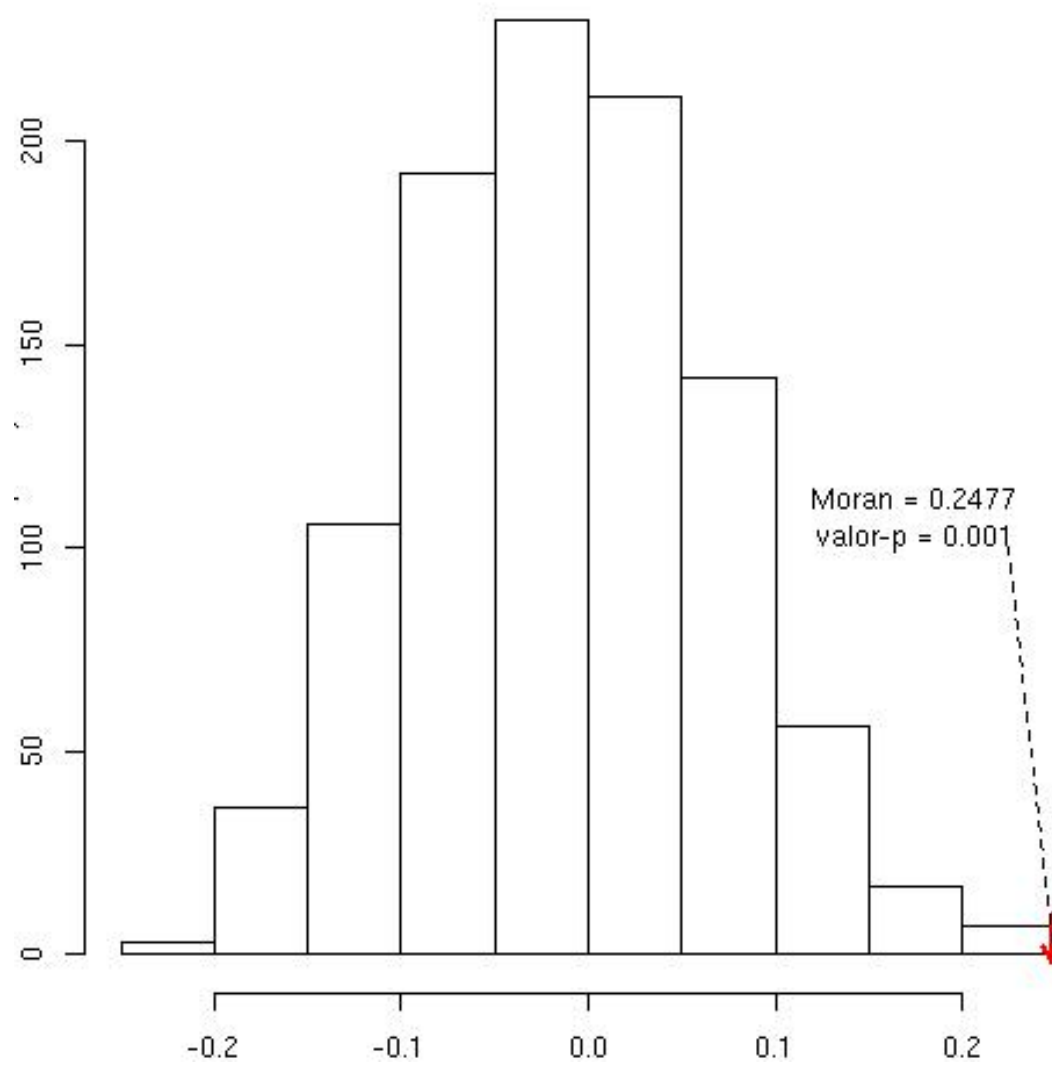
Voltando aos mapas de MG: *CAR* com $\rho = 0.7$



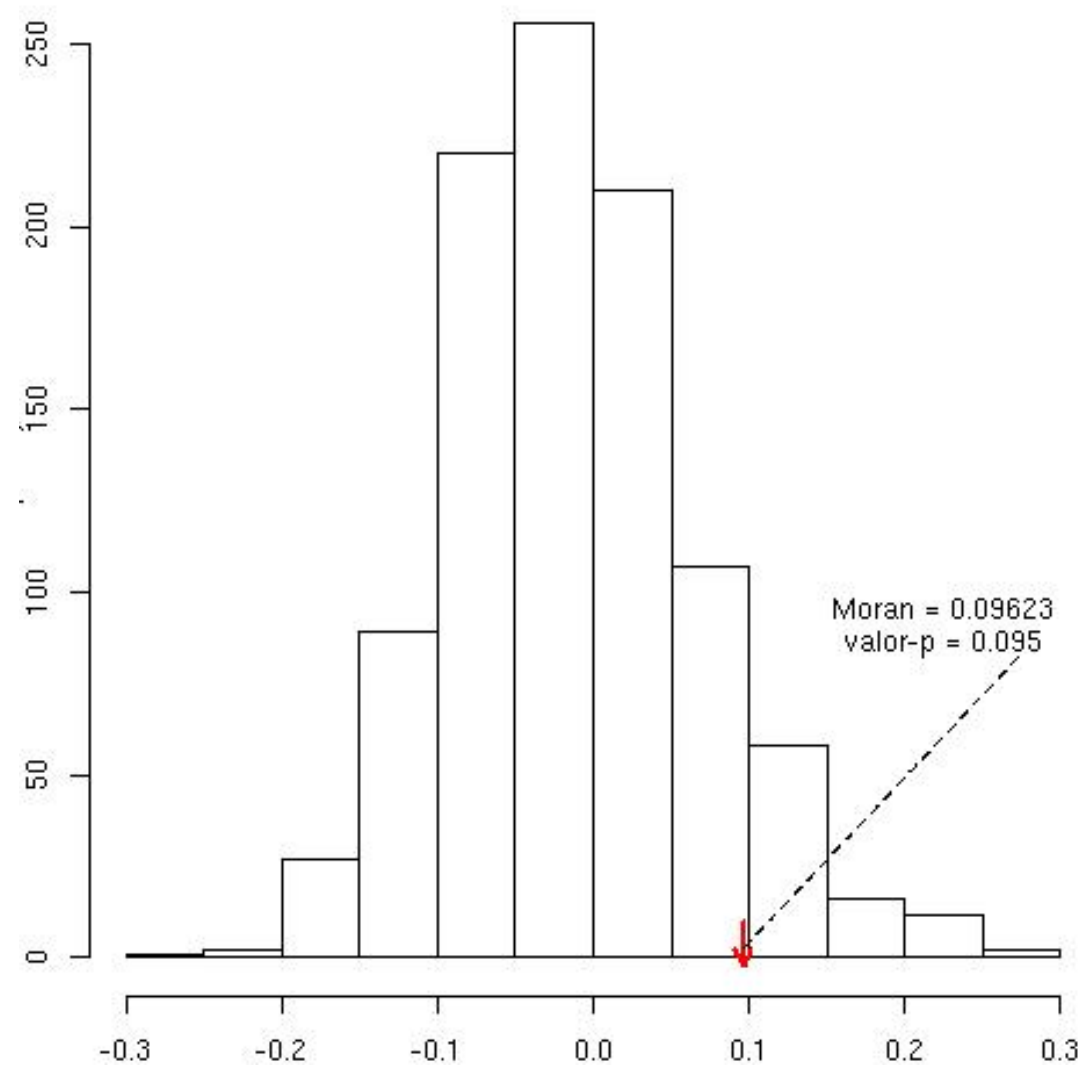
Voltando aos mapas de MG: *CAR* com $\rho = 0.7$



Voltando aos mapas de MG: *CAR* com $\rho = 0.7$



Voltando aos mapas de MG: *CAR* com $\rho = 0.7$



Testes com variáveis binárias

Quando a variável for binária (valores de y_i só podem ser 0 ou 1), podemos calcular o número de conexões entre áreas vizinhas.

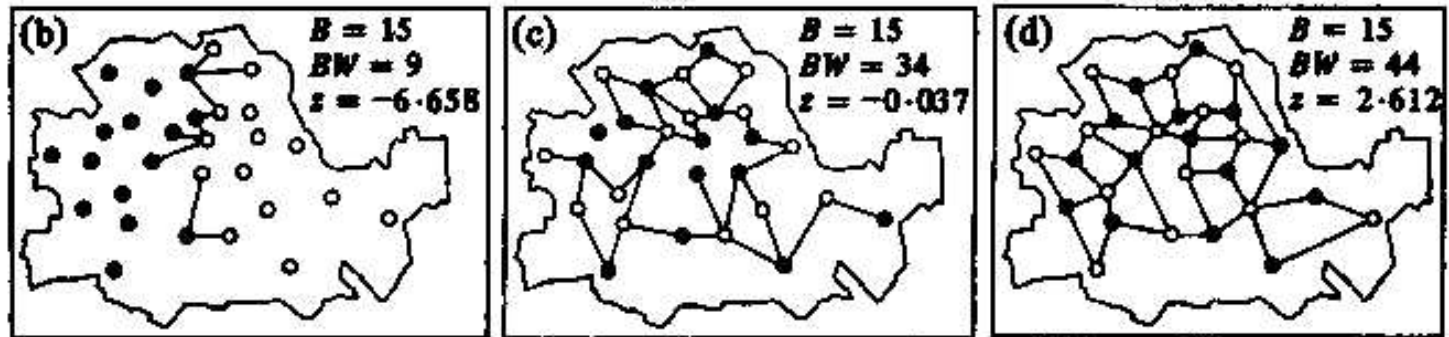
O índice usual é o número de conexões do tipo 0-1.

Quando os dados são binários, o teste de permutação baseado no índice de conexões é equivalente ao teste baseado em Moran.

Podemos também calcular o número de conexões do tipo 0-0 ou do tipo 1-1 (mas estas alternativas são piores que o teste baseado em conexões do tipo 0-1)

P-valor é calculado do mesmo modo que antes, após permutar várias vezes (999) os valores binários de y_i .

Exemplo



...