



UNIVERSIDADE FEDERAL DE JUIZ DE FORA - *CAMPUS* Governador Valadares
ECONOMETRIA ESPACIAL
Prof. Vinícius de Azevedo Couto Firme

Econometria Espacial:

Capítulo 4 – Análise Exploratória de Dados Espaciais



UNIVERSIDADE FEDERAL DE JUIZ DE FORA - *CAMPUS* Governador Valadares

ECONOMETRIA ESPACIAL

Prof. Vinícius de Azevedo Couto Firme

Estrutura da Apresentação:

1. Introdução
2. Autocorrelação Espacial Global Univariada
3. Autocorrelação Espacial Global Bivariada
4. Autocorrelação Espacial Local Univariada
5. Autocorrelação Espacial Local Bivariada
6. Detecção de *Outliers*
7. Heterogeneidade



1. Introdução

- A imbricação entre autocorrelação e heterogeneidade espaciais dificulta a especificação correta de modelos (ambos os efeitos devem ser controlados).
- As técnicas de Análise Exploratória de Dados Espaciais (AEDE) têm o propósito de auxiliar o processo de especificação dos modelos.
- De modo geral a AEDE permite:
 - a) Identificar padrões de distribuição espacial (autocorrelação espacial - ACS);
 - b) Identificar localidades atípicas (outliers espaciais);
 - c) Identificar associações espaciais (clusters espaciais) e sugerir a adoção de regimes espaciais ou outra forma de controle da instabilidade espacial.

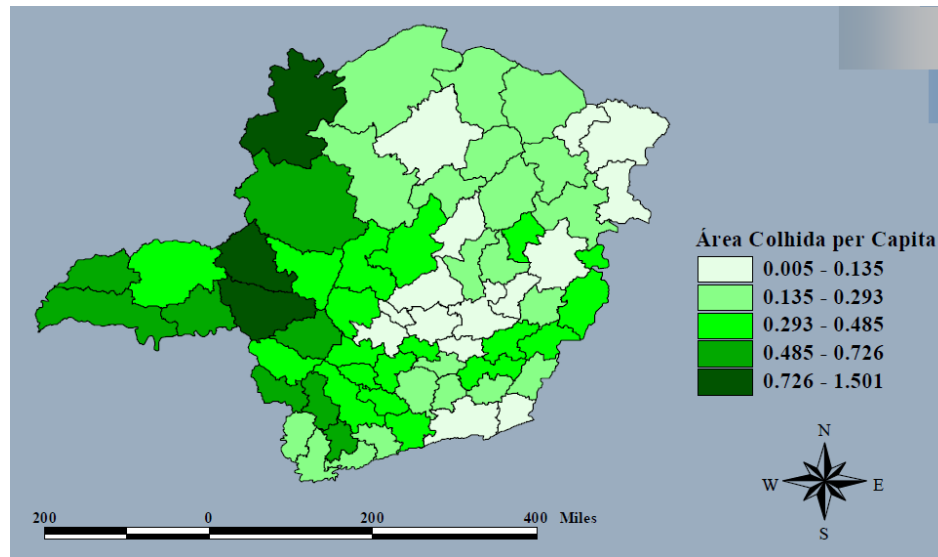
Dica 1: Toda regressão deve ser precedida de uma análise exploratória dos dados.

Dica 2: Deve-se trabalhar com variáveis espacialmente densas ou intensas (divididas pela população ou área) a fim de evitar *outliers* e correlações espúrias.

2. Autocorrelação Espacial Global Univariada

- Os dados do Mapa 1 estão distribuídos aleatoriamente ou há ACS?

Mapa 1. Distribuição da área colhida per Capita de Minas Gerais



- Como a análise visual é, frequentemente, viesada, sugere-se o uso de estatísticas para responder esta pergunta.



2. Autocorrelação Espacial Global Univariada

- Os coeficientes de autocorrelação (espaciais ou não) se baseiam em alguma medida de autocovariância e de variância dos dados.
- As medidas de autocovariância mais usadas são:
 - a) Produto-cruzado: $(y_i * y_j)$
 - b) Quadrado da diferença: $(y_i - y_j)^2$
 - c) Módulo da Diferença: $|y_i - y_j|$
- A fim de se evitar a influência da escala das variáveis, deve-se dividir a autocovariância pela variância das variáveis.
- Por fim, como se trata de uma abordagem espacial, deve-se considerar uma matriz de pesos espaciais (W) no cálculo da ACS.



2. Autocorrelação Espacial Global Univariada

2.2) Estatística I de Moran

- Trata-se do primeiro teste criado para verificar ACS (mais usado na literatura!). Definida por Moran (1948), esta estatística varia de -1 a 1. Valores próximos de 1 indicam ACS (+) e valores próximos de -1 indicam ACS(-).

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

Matricialmente:

$$I = \frac{n}{S_0} \frac{z' W z}{z' z} \quad (2)$$

Onde: n é o número total de regiões; z é a variável de interesse padronizada/normalizada; W é a matriz de pesos espaciais; $S_0 = \sum_i \sum_j w_{ij}$; i e j representam duas regiões quaisquer.

Nota 1: Normalização $\rightarrow z_i = (y_i - \bar{y})/\sigma$, onde $\sigma = \sqrt{\sum_i (y_i - \bar{y})^2 / n}$

Nota 2: Se W for normalizada na linha, $S_0 = n$, logo: $I = (z' W z)/(z' z)$



2. Autocorrelação Espacial Global Univariada

2.2) Estatística I de Moran

$$E(I) = -[1/(n - 1)] \quad (3)$$

- Se o valor do I de Moran estimado for:
 - a) $I > E(I)$ há ACS positiva (similaridade espacial \rightarrow prevalece perfil AA x BB);
 - b) $I < E(I)$ há ACS negativa (dissimilaridade espacial \rightarrow prevalece perfil AB x BA);
 - c) $I \cong E(I)$ aceita-se H_0 e assume-se aleatoriedade espacial.

Nota: Geralmente assume-se que $E(I) = 0$, pois $E(I) \rightarrow 0$ quando $n \rightarrow \infty$.

Exercício: Calcular I e $E(I)$ com base nas informações abaixo usando Eq. 1, 2 e 3:

$$Y = \begin{bmatrix} 10 \\ 5 \\ 20 \end{bmatrix}; W = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix};$$

Nota: $\bar{Y} = 11.67$ e $\sigma_Y = 6.24$.

Resp.: $I = 0,857$ ou $I = -0,893$? Com base em $E(I)$, houve ACS(+) ou ACS(-)?



2. Autocorrelação Espacial Global Univariada

2.2) Estatística I de Moran

- A significância estatística do I de Moran pode ser calculada de 2 formas:

a) Assumindo que $Z \sim normal(0,1)$

$$\text{Neste caso: } Z(I) = [I - E(I)] / DP(I) \quad (4)$$

Onde $Z(I)$ representa o valor crítico para um conjunto de variáveis normalizadas (usar tabela de distribuição normal); $E(I)$ e $DP(I)$ representam o valor esperado e o desvio padrão teóricos associado ao I de Moran.

b) Usando critério de permutações

- i) Os dados são embaralhados no mapa de forma aleatória t vezes a fim de gerar t mapas distintos do real (r). ii) Estima-se o I de Moran (I) para cada t mapa e cria-se uma distribuição para os I coeficientes gerados. iii) Compara-se o I gerado com base em r com a distribuição gerada com base nos t mapas.



2. Autocorrelação Espacial Global Univariada

2.2) Estatística I de Moran

Tabela 1. Coeficiente de I de Moran baseado no Mapa 1:

| Convenção | Normalidade | | Permutação | |
|-----------|-------------|---------------|------------|---------------|
| | I | Probabilidade | I | Probabilidade |
| Rainha | 0,359 | 0,000 | 0,359 | 0,000 |
| Torre | 0,361 | 0,000 | 0,361 | 0,000 |

- Diagrama de Dispersão de Moran: permite visualizar a ACS.
- Formalmente, o diagrama apresenta o coeficiente de angulação (β) oriundo da seguinte estimação:

$$Wz = \alpha + \beta z + \varepsilon \quad (5)$$

$$\text{Note que } \hat{\beta} = I = \frac{n}{S_0} \frac{z'Wz}{z'z}$$

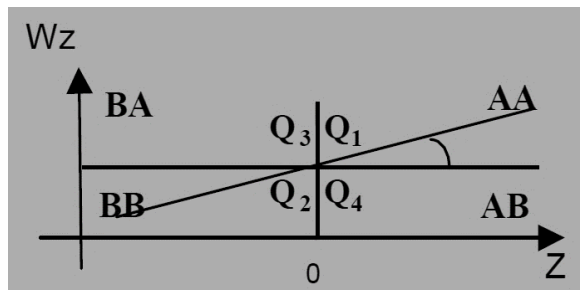
Nota: lembre-se que z é a variável y padronizada, com média=0 e variância=1.

2. Autocorrelação Espacial Global Univariada

2.2) Estatística I de Moran

Figura 1.

Diagrama de Dispersão Teórico

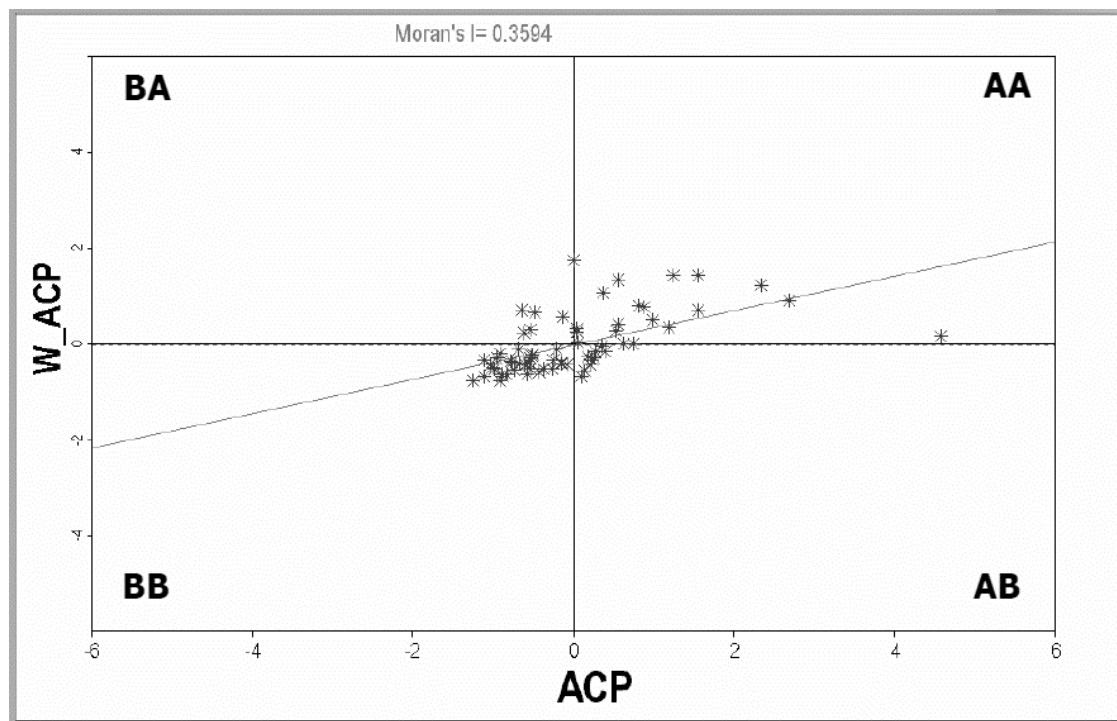


Nota: possibilita verificar outliers.

Questão: o que garante que a reta sempre passe no eixo (0,0)?

Figura 2.

Diagrama de Dispersão do Mapa 1





2. Autocorrelação Espacial Global Univariada

2.2) Estatística c de Geary: criada em 1954, esta estatística baseia-se na autocovariância medida via quadrado da diferença. Varia de 0 a 2. Valores próximos de 0 indicam ACS (+) e valores próximos de 2 indicam ACS(-).

$$c = \frac{n-1}{2 \sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij} (y_i - y_j)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (6)$$

Nota: **a)** se $c < E(c)$, há ACS (+); **b)** se $c > E(c)$, há ACS (-); **c)** se $c \cong E(c)$, não há ACS (aceita-se H_0); **d)** Se $n \rightarrow \infty, E(c) \cong 1$.

▪ A significância estatística do c de Geary pode ser calculada de 2 formas:

a) Assumindo que $Z \sim normal(0,1)$

Neste caso: $Z(c) = [c - E(c)] / DP(c)$ (7)

b) Usando critério de permutações

Nota: ambos os critérios “a” e “b” são análogos aos do I de Moran.



2. Autocorrelação Espacial Global Univariada

2.2) Estatística c de Geary

Tabela 2. Coeficiente de c de Geary baseado no Mapa 1:

| <i>Convenção</i> | <i>Normal</i> | | <i>Permutação</i> | |
|------------------|---------------|----------------------|-------------------|----------------------|
| | <i>c</i> | <i>Probabilidade</i> | <i>c</i> | <i>Probabilidade</i> |
| Rainha | 0,566 | 0,000 | 0,566 | 0,000 |
| Torre | 0,569 | 0,000 | 0,569 | 0,000 |

- Em resumo, o c de Geary e o I de Moran fornecem 3 informações:
 - a) O nível de significância dos testes informa se há ACS ou não;
 - b) O coeficiente obtido nos testes indica se a ACS é (+) ou (-);
 - c) A magnitude dos coeficientes revela a força da ACS.

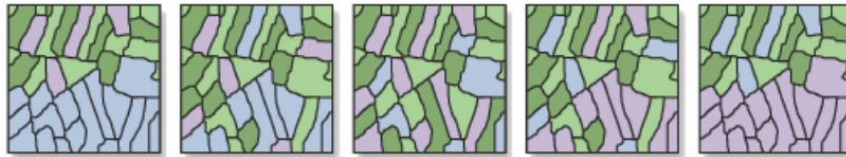
2. Autocorrelação Espacial Global Univariada

2.3) Estatística G de Getis-Ord

- Proposta por Getis-Ord (1992), trata-se de uma estatística de aglomeração espacial (e não de ACS, como é o caso do I de Moran e do C de Geary).

➤ G de Getis-Ord

Measures the degree of clustering for either high values or low values

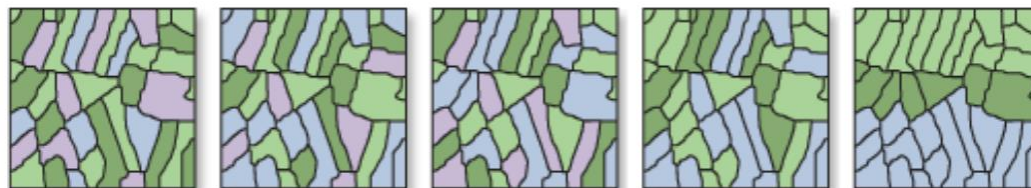


Lows Cluster ←————→ Highs Cluster
 Prevalece BB Aleatório Prevalece AA

Nota: Quando há agrupamentos do tipo AA (*hot spot*) e do tipo BB (*cool spot*) no mesmo mapa, um tende a anular o efeito do outro e a estatística G pode se tornar não significativa.

➤ I de Moran e C de Geary

Measures spatial autocorrelation based on feature locations & attribute values.



Dispersed ←————→ Clustered
 AB-BA Aleatório AA-BB



2. Autocorrelação Espacial Global Univariada

2.3) Estatística G de Getis-Ord

$$G(d) = \frac{\sum_i \sum_j w_{ij}(d) y_i y_j}{\sum_i \sum_j y_i y_j}, \quad \text{para } j \neq i \quad (8)$$

Onde: $w_{ij}(d)$ é uma matriz espacial binária baseada numa distância pré-fixada.

• Para verificar a significância do $G \rightarrow Z(G) = [G - E(G)]/DP(G) \quad (9)$

a) se $G > E(G)$, há aglomeração do tipo AA (*hot spot*);

b) se $G < E(G)$, há aglomeração do tipo BB (*cool spot*);

c) se $G \cong E(G)$, dados podem ser aleatórios ou apresentar ACS (-).

Problemas (versão 1992): **1)** só capta ACS (+); **2)** só aceita valores de $y > 0$; **3)** só aceita matrizes binárias (não pode normalizar nem usar distância inversa).

Nota: Versão de G proposta por Ord-Gets (1995) resolve problemas 2 e 3.

Tabela 3. Coeficiente de G de Getis-Ord baseado no Mapa 1:

| G | Desvio-padrão | Z | P-valor |
|-------|---------------|-------|---------|
| 0,088 | 0,006 | 1,887 | 0,059 |

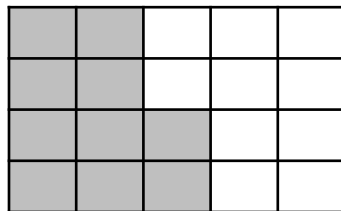
Nota: $G > 0$ indica que Perfil AA prevalece sobre o BB a 10% de significância.

2. Autocorrelação Espacial Global Univariada

2.4) Joint Count ou Contagem de Junções

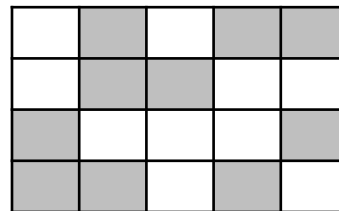
- Mede a ACS para variáveis binárias (útil em Logit/Probit Espacial).
- Considere o seguinte caso onde $\boxed{N} \rightarrow y_i = 1$; $\boxed{B} \rightarrow y_i = 0$. Logo:

a) Concentrado (AA-BB)



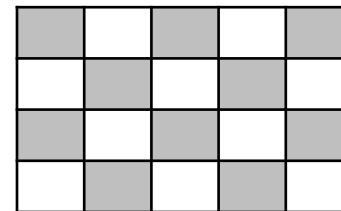
NN=13; BB=13; NB=5

b) Aleatório



NN=5; BB=6; NB=20

c) Disperso (AB-BA)



NN=0; BB=0; NB=31

- Como o total de células (n) é igual ao número de células escuras (n_N) + o número de células brancas (n_B), a probabilidade encontrarmos N e B é:

$$\mathbf{a)} p_N = n_N/n ; \mathbf{b)} p_B = n_B/n \quad (10)$$

$$\mathbf{▪} \text{ Logo: } \mathbf{a)} Prob.^{NN} = p_N * p_N; \mathbf{b)} Prob.^{BB} = p_B * p_B; \mathbf{c)} Prob.^{NB} = 2(p_N * p_B) \quad (11)$$



2. Autocorrelação Espacial Global Univariada

2.4) *Joint Count* ou Contagem de Junções

- Com base nas probabilidades das junções NN, BB e NB (ou BN), tem-se que:

$$\mathbf{a) } E(NN) = p_N^2 L; \mathbf{b) } E(BB) = p_B^2 L; \mathbf{c) } E(NB) = 2p_N p_B L \quad (12)$$

Onde L representa o número total de junções no mapa.

- A significância desta estatística implica calcular:

$$\begin{aligned} Z(NN) &= [NN - E(NN)] / DP(NN) \\ Z(BB) &= [BB - E(BB)] / DP(BB) \\ Z(NB) &= [NB - E(NB)] / DP(NB) \end{aligned} \quad (13)$$

Nota: a derivação do DP de uma *Joint Count* está disponível em Cliff e Ord (1981).

**Tabela 4. Contagem do NN para vitórias do PT (N=1) versus PSDB (B=0)
(2º Turno Eleição presidencial 2002)**

| Variável | NN | E(NN) | Desvio Padrão | Z |
|----------|------|---------|---------------|-------|
| PT | 7419 | 6105.22 | 53.99 | 24.33 |

Nota: como $NN > E(NN)$, conclui-se que os dados estão agrupados (AA-BB).



3. Autocorrelação Espacial Global Bivariada

- Permite verificar se uma variável qualquer y está associada à outra variável em regiões vizinhas (W_x).
- Nota: A análise univariada considerava apenas a relação entre y e W_y .
- Neste caso, o **I de Moran bivariado** seria (variáveis z_1 e z_2):

$$I^{z_1, z_2} = \frac{n}{S_0} \frac{z_1' W z_2}{z_1' z_1} \quad (14)$$

Onde: z_1 é a densidade rodoviária pavimentada (RPP); z_2 é a área colhida *per capita* (ACP).

Tabela 5. I de Moran Bivariado com base em z_1 e z_2 :

| <i>Estatística I</i> | <i>E(I)</i> | <i>Desvio-padrão</i> | <i>Probabilidade</i> |
|----------------------|-------------|----------------------|----------------------|
| 0,180 | -0,015 | 0,061 | 0,01 |

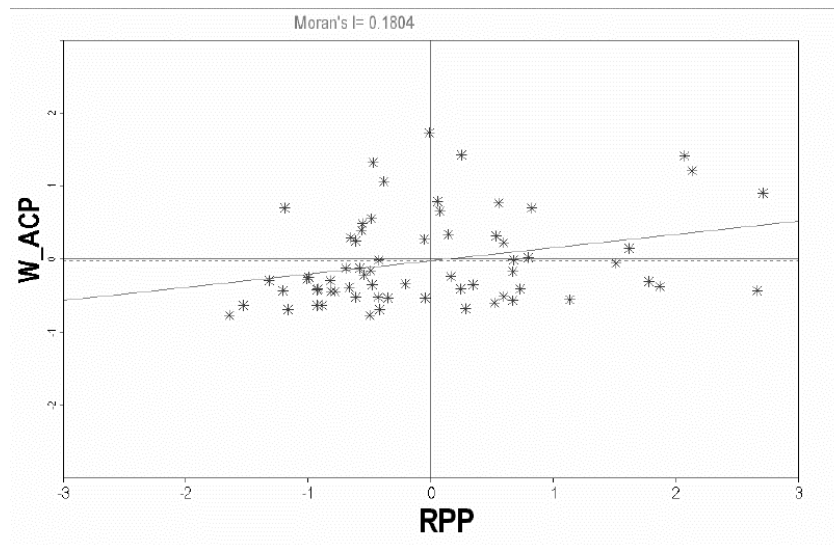
3. Autocorrelação Espacial Global Bivariada

- O Diagrama de Dispersão de Moran Bivariado representa a seguinte estimação:

$$Wz_2 = \alpha + \beta z_1 + \varepsilon \quad (15)$$

$$\text{Note que } \hat{\beta} = I^{z_1, z_2} = \frac{n}{S_0} \frac{z_1' W z_2}{z_1' z_1} \quad (16)$$

Figura 3. Diagrama de Dispersão de Moran para z_1 (RPP) e z_2 (ACP)





4. Autocorrelação Espacial Local Univariada

- As estatísticas de ACS global (média geral) podem esconder padrões locais de associação espacial.

Caso 1: $I \cong 0$ (não significativo) pode esconder associações fracas AA-BB e AB-BA.

Caso 2: $I > 0$ (significativo) pode esconder associações fracas AB-BA (Mapa 1 e Fig. 2).

Solução: uso de estatísticas de associação locais.

4.1) G_i local: proposto por Getis e Ord (1992), o G_i permite detectar, para cada região i , “bolsões” locais do tipo AA (*hot spot*) e BB (*cool spot*).

$$G_i(d) = \frac{\sum_j w_{ij}(d)y_j}{\sum_j y_j}, \quad \text{para } j \neq i \quad (17)$$

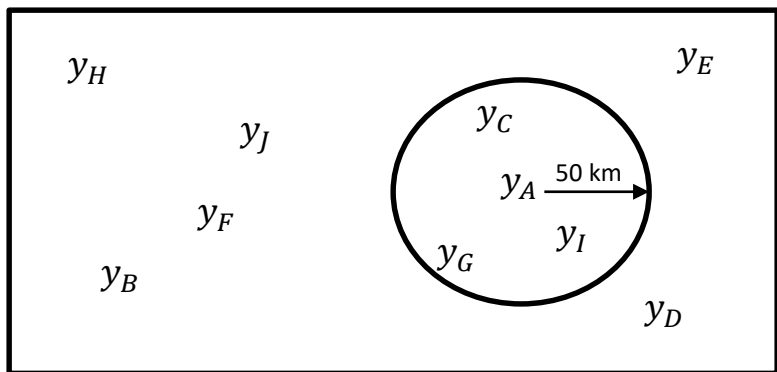
Onde: $w_{ij}(d)$ é uma matriz binária baseada na distância em relação à i ; $\sum_j w_{ij}(d)y_j$ é o somatório dos vizinhos de i (dada a distância “ d ”); $\sum_j y_j$ é o somatório de todas as regiões da amostra (independente da distância), exceto i .

4. Autocorrelação Espacial Local Univariada

- Uma variação de G_i , denominada G_i^* , considera a região i no denominador:

$$G_i^*(d) = \frac{\sum_j w_{ij}(d)y_j}{\sum_j y_j}, \quad \text{para qualquer } j \quad (17)$$

Figura 4. Exemplo de Matriz usada em G_i e G_i^*



- Com base na Fig. 4, tem-se que:

$$G_A = \frac{y_C + y_G + y_I}{y_B + \dots + y_J} \quad \text{ou} \quad G_A^* = \frac{y_C + y_G + y_I}{y_A + y_B + \dots + y_J} \quad (18)$$



4. Autocorrelação Espacial Local Univariada

- G_i e G_i^* permanecem com as mesmas desvantagens de G (global).
- Análise análoga ao G , isto é: $G_i > E(G_i) \rightarrow \text{hot spot}$; $G_i < E(G_i) \rightarrow \text{cool spot}$.

$$E(G_i) = W_i / (n - 1), \text{ onde } W_i = \sum_j w_{ij}(d) \quad (19)$$

- A inferências estatística (significância) é realizada com base em:

$$Z(G_i) = [G - E(G_i)] / DP(G_i) \quad \text{ou} \quad Z(G_i^*) = [G - E(G_i^*)] / DP(G_i^*) \quad (20)$$

Onde o cálculo da variância (lembrar que $\sqrt{VAR(G_i)} = DP(G_i)$) envolve:

$$Var(G_i) = \frac{W_i(n-1-W_i)}{(n-1)^2(n-1)} \left[\frac{s(i)}{\bar{y}(i)} \right]^2 \quad (21)$$

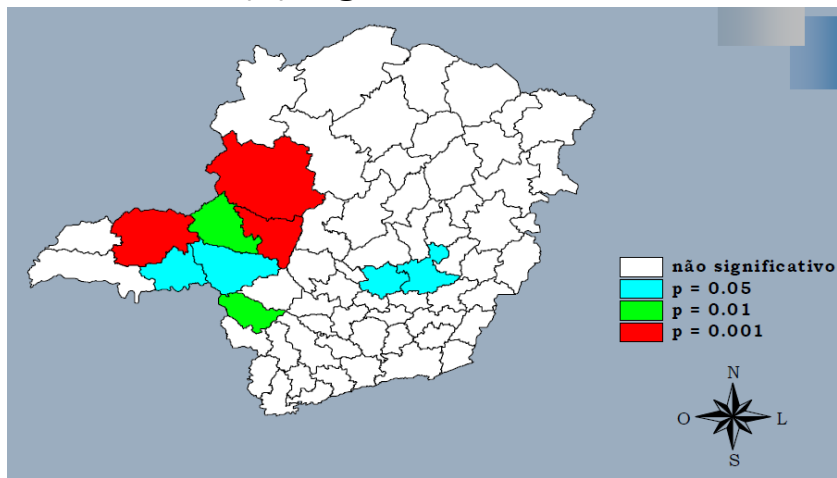
$$\text{Sendo: } \bar{y}(i) = \frac{\sum_j y_j}{(n-1)} \quad \text{e} \quad s^2(i) = \frac{\sum_j y_j^2}{(n-1)} - [\bar{y}(i)]^2 \quad (22)$$

4. Autocorrelação Espacial Local Univariada

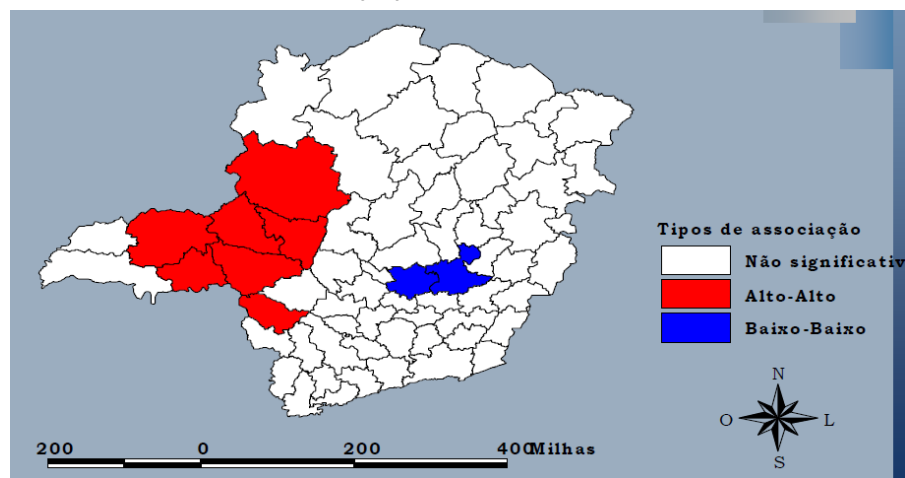
- Posteriormente, Getis e Ord (1995) revisaram a fórmula de G_i e G_i^* a fim de permitir o uso de matrizes não binárias e que y apresente valores negativos, positivos e nulos.

Mapa 2. Significância e *Cluster* do G_i de Getis-Ord para a Área Colhida per Capita.

(a) Significância



(b) *Cluster*





4. Autocorrelação Espacial Local Univariada

- Segundo Getis e Ord (1995), G_i e G_i^* são afetados pela ACS global (resultados mais confiáveis na ausência de ACS).

Nota: ACS global aumenta a variância de G_i e G_i^* , afetando inferência via Eq. 20.

- Na presença de ACS, usa-se o **O_i de Ord e Getis (2001)**: útil para testar *hot* e *cool spots* locais.
- O teste verifica se há diferença estatística entre a média dos vizinhos de i (\bar{y}_d) e a média dos não vizinhos (\bar{y}_o) – ver Fig. 4.

$$O_i(d) = \bar{y}_d - \bar{y}_o \quad (23)$$

Nota: Se $\bar{y}_d > \bar{y}_o \Rightarrow$ *Hot Spot*; Se $\bar{y}_d < \bar{y}_o \Rightarrow$ *Cool Spot*.

- A inferência de O_i é realizada com base em:

$$Z(O_i) = [O_i - E(O_i)]/DP(O_i) \quad (24)$$



4. Autocorrelação Espacial Local Univariada

4.2) I de Moran Local (I_i): proposto por Anselin (1995), trata-se do primeiro “*Local Indicator of Spatial Association*” – LISA.

- Uma estatística LISA deve satisfazer 2 critérios:
 - a) Gerar indicadores individuais (para cada região i) de *clusters* espaciais com significância estatística.
 - b) O somatório dos indicadores locais (I_i) deve ser proporcional ao global (I).

Nota: G_i , G_i^* e O_i não satisfazem o segundo critério.

- Assim como o I de Moran Global, o I_i capta associações positivas e negativas (G_i , G_i^* e O_i só captam ACS positivas). Formalmente:

$$I_i = z_i \sum_{j=1}^J w_{ij} z_j \quad (25)$$

Nota: I_i mede a associação entre a variável de interesse da região i (z_i) e a média de seus j vizinhos (z_j).



4. Autocorrelação Espacial Local Univariada

- Prova de que o critério “b” de um LISA é atendido:

$$\begin{aligned}\sum_i I_i &= \sum_i z_i \sum_j w_{ij} z_j \\ \sum_i I_i &= \sum_i \sum_j w_{ij} z_i z_j\end{aligned}\tag{25}$$

$$\text{Como: } I = \frac{n}{s_0} \frac{\sum_i \sum_j w_{ij} z_i z_j}{\sum_i z_i^2}\tag{26}$$

$$\text{Tem-se que: } I = \frac{n}{s_0} \frac{\sum_i I_i}{\sum_i z_i^2}\tag{27}$$

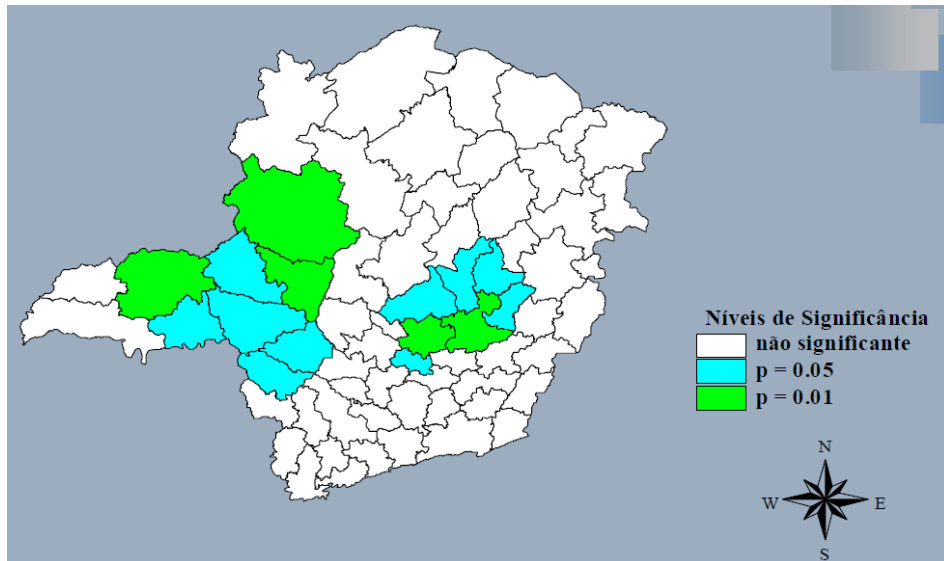
$$\text{Caso } W \text{ seja normalizada: } \sum_i I_i = kI, \text{ onde } k = \sum_i z_i^2\tag{28}$$

- Portanto, o somatório dos índices locais (I_i) é proporcional o índice global (I), sendo k o fator de proporcionalidade.

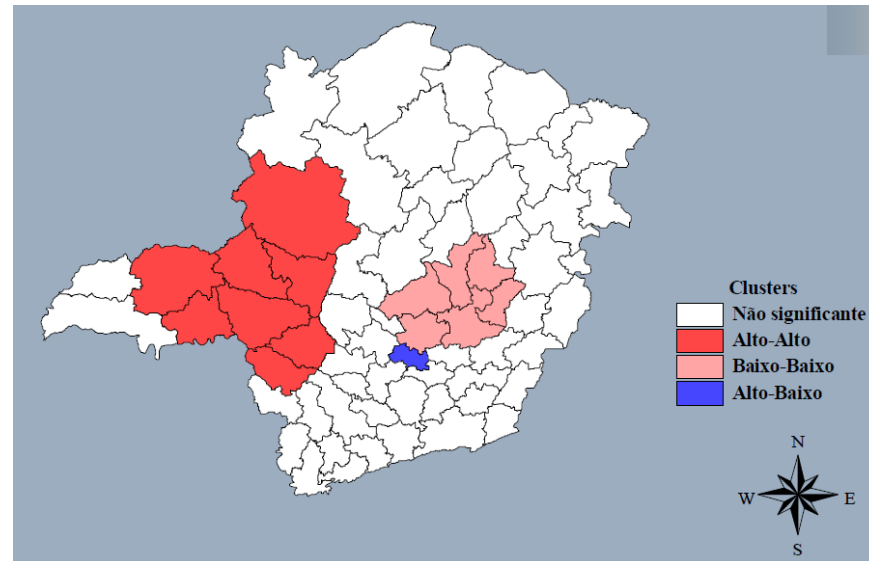
4. Autocorrelação Espacial Local Univariada

Mapa 2. Significância e *Cluster* do I_i de Moran para a Área Colhida per Capita.

(a) Significância



(b) *Cluster*



Nota: valores não significativos (na cor branca) indicam que o valor da variável na região i está próxima da média e, portanto, não são consideradas Alto ou Baixo.



4. Autocorrelação Espacial Local Univariada

4.3) c de Geary Local (c_i): outra estatística LISA apresentada por Anselin (1995).

$$c_i = \sum_j w_{ij} (z_i - z_j)^2 \quad (29)$$

Nota: semelhante ao I_i , porém c_i considera o quadrado da diferença.

Geralmente usa-se: $c_i = \frac{\sum_j w_{ij} (z_i - z_j)^2}{2} \quad (30)$

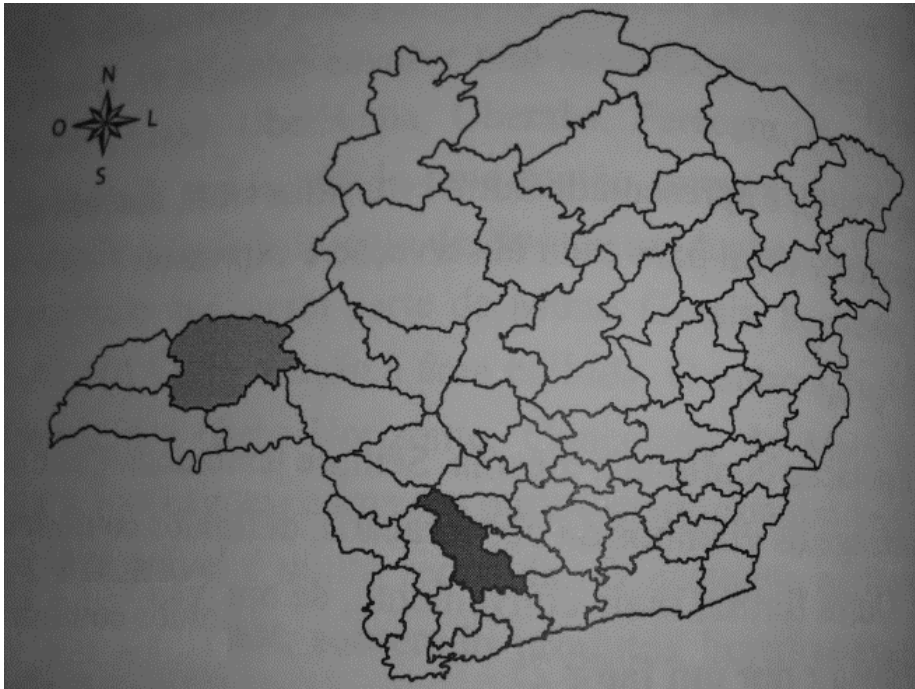
Nota: Eq. 30 garante que $E(c_i) = 1$. Logo, análise de c_i é análoga ao c global.

- $0 \leq c_i < 1 \rightarrow ACS(+)$;
- $2 \geq c_i > 1 \rightarrow ACS(-)$
- $c_i \cong 1 \rightarrow$ não há ACS

Nota: análise via I_i é mais usual (c_i não fornece as 4 ACS : AA, BB, AB e BA).

4. Autocorrelação Espacial Local Univariada

Mapa 3: *Cluster* do c_i de Geary para a Área Colhida per Capita.



Nota:

1 cluster com ACS(-): Uberlândia
(cinza claro)

1 cluster com ACS(+): Varginha
(cinza escuro)

4. Autocorrelação Espacial Local Univariada

4.4) Análise de Robustez: testa estabilidade dos I_i locais frente à escolha de W .

- Consiste em verificar se mudanças em W alteram os resultados de I_i .
- O ideal é que a diagonal principal da matriz de robustez seja próxima de 100.

Tabela 6. Análise da Robustez do I de Moran Local (I_i).

| 2003 | | k=10 para k=15 | | | | |
|------|-------|----------------|--------|--------|--------|--|
| | NS | AA | BB | AB | BA | |
| NS | 82,76 | 3,45 | 0,00 | 5,17 | 8,62 | |
| AA | 0,00 | 100,00 | 0,00 | 0,00 | 0,00 | |
| BB | 0,00 | 0,00 | 100,00 | 0,00 | 0,00 | |
| AB | 0,00 | 0,00 | 0,00 | 100,00 | 0,00 | |
| BA | 0,00 | 0,00 | 0,00 | 0,00 | 100,00 | |

| 2003 | | k=10 para k=20 | | | | |
|------|-------|----------------|--------|--------|--------|--|
| | NS | AA | BB | AB | BA | |
| NS | 70,00 | 5,17 | 1,20 | 8,10 | 15,52 | |
| AA | 0,00 | 100,00 | 0,00 | 0,00 | 0,00 | |
| BB | 0,00 | 0,00 | 100,00 | 0,00 | 0,00 | |
| AB | 0,00 | 0,00 | 0,00 | 100,00 | 0,00 | |
| BA | 0,00 | 0,00 | 0,00 | 0,00 | 100,00 | |

Nota: a alteração de $k=10$ para $k=15$ e de $k=10$ para $k=20$ afetou apenas as regiões NS. Àquelas que, inicialmente, haviam sido agrupadas em AA, BB, AB e BA mantiveram-se estáveis.

5. Autocorrelação Espacial Local Bivariada

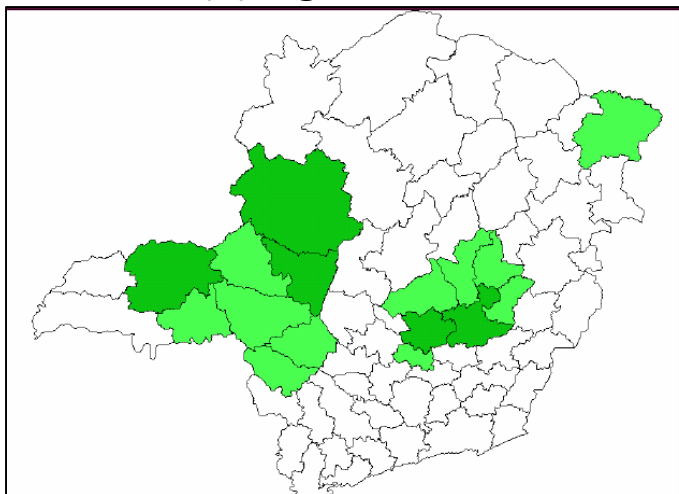
- O I de Moran local bivariado ($I_i^{z_1 z_2}$) indica o grau de associação linear entre a variável z_1 da região i (z_{1i}) e a média de outra variável (z_2) das j regiões vizinhas à i ($W_{z_2 i}$). Formalmente:

$$I_i^{z_1 z_2} = z_{1i} W_{z_2 i} \quad (31)$$

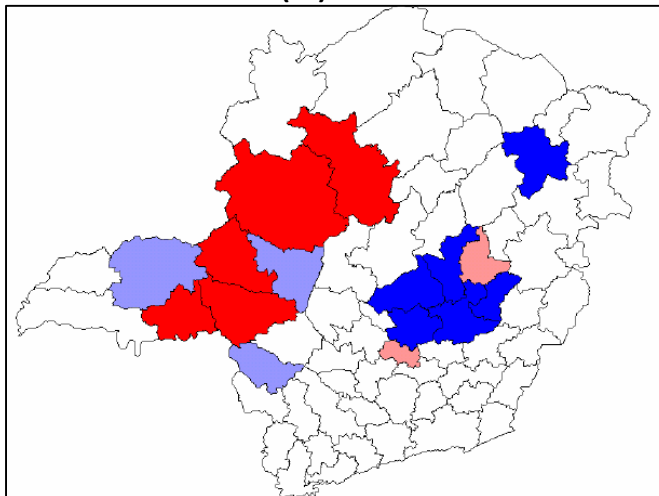
Onde: z_{1i} e z_{2i} são versões normalizadas de y_i e x_i .

Mapa 4. Significância e *Cluster* do $I_i^{z_1 z_2}$ bivariado para ACP e RPP.

(a) Significância



(b) *Cluster*



Nota: autor deve ter usado matrizes distintas em (a) e (b). Mapeamento deveria ser rigorosamente igual.



6. Detecção de *Outliers*

- *Outliers* são informações discrepantes, que destoam das demais.
- No contexto da Econometria Espacial, pode-se destacar os *Outliers Globais* e os *Outliers Espaciais*.

6.1) *Outlier Global*: trata-se de uma observação com valor muito acima ou muito abaixo das demais.

- Os *Outliers Globais* podem ser detectados via ***Box Plot*** e visualizados via *Box Map* ou *Cartograma*.
- A técnica do *Box Plot* define como *Outliers Globais* as informações que estão fora de um determinado intervalo interquartilício (IIQ), onde: $IIQ = Q3 - Q1$.
- Geralmente, define-se um critério (*Hinge*) igual a 1,5 ou 3 vezes o *IIQ*.

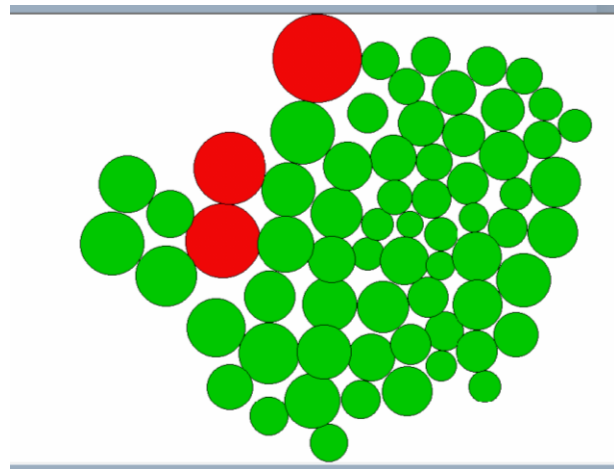
Nota: assim como a σ^2 e o σ , o *IIQ* também mede o grau de dispersão dos dados. Contudo, na análise via *IIQ* os dados são ordenados de forma crescente.

6. Detecção de *Outliers*

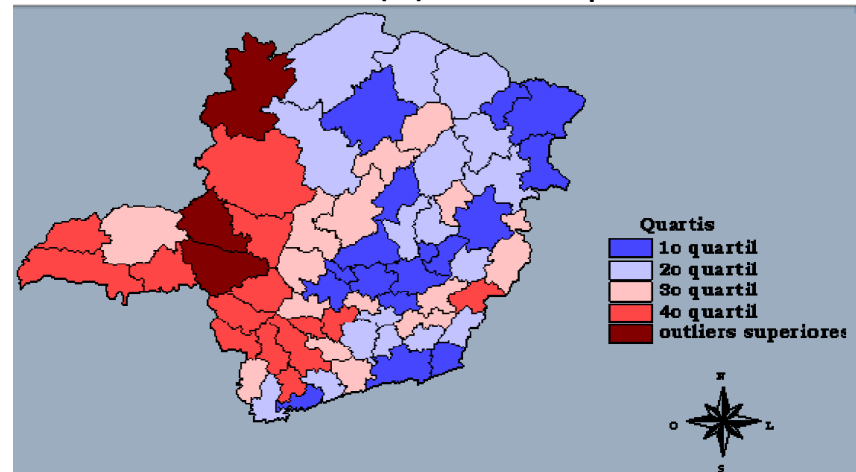
- Logo, se $Q3 = 75$ e $Q1 = 25$, tem-se que $IIQ = 50$. Neste caso:
 - a) $(25 - 75) = -50 > \text{Outlier} > 150 = (75 + 75)$ (*Hinge* = 1,5)
 - b) $(25 - 150) = -125 > \text{Outlier} > 225 = (75 + 150)$ (*Hinge* = 3,0)

Mapa 5. *Outliers Globais* com base no *Box Plot* para a variável ACP:

(a) Cartograma



(b) Box Map



Nota: O Box Map (b) detecta *outliers globais* superiores e inferiores.



6. Detecção de *Outliers*

6.1) *Outlier Espacial*: trata-se de uma observação que não segue o mesmo padrão de dependência espacial das demais.

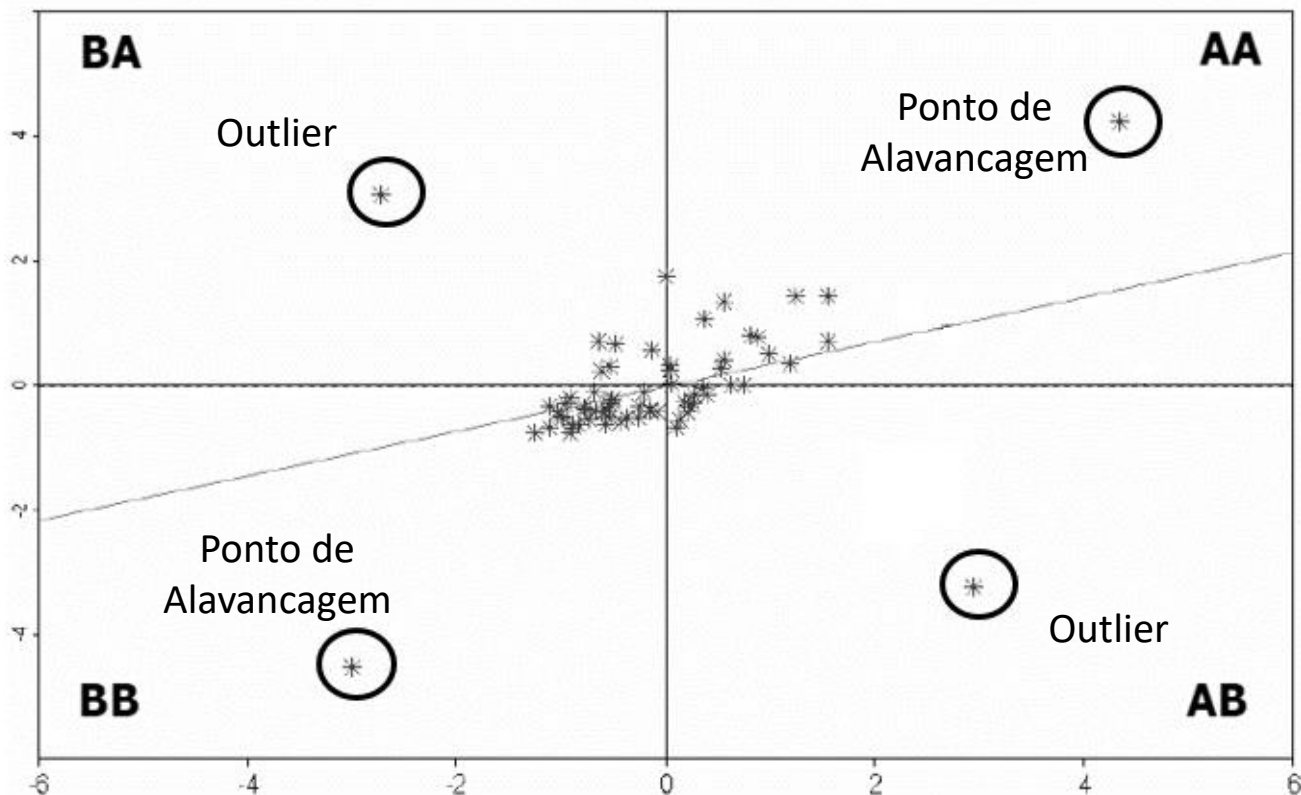
- Os *Outliers Espaciais* podem ser detectados via ***Diagrama de Dispersão de Moran***.
- Geralmente, as informações distantes do centro (0,0) em mais de 2 desvios padrões (2σ) são consideradas *outliers espaciais*.
- O diagrama de dispersão permite verificar, ainda, a existência de *pontos de alavancagem* (mesmo padrão de ACS, porém excessivo).

Nota: Os pontos de alavancagem também podem ser detectados via *distância de Cook*, que consiste em regressar a Eq. 5 com e sem as observações “suspeitas” de alavancagem.

- Tanto os *outliers espaciais* e quanto os *pontos de alavancagem* podem afetar o nível de dependência espacial.

6. Detecção de *Outliers*

- **Figura 5.** Detecção de *Outliers* Espaciais e Pontos de Alavancagem via Diagrama de Dispersão de Moran.





7. Heterogeneidade

- A heterogeneidade espacial pode se refletir nos parâmetros de intercepto, de inclinação e no termo de erro (heterocedasticidade).
- A fim de detectar estes problemas foram desenvolvidos alguns testes:
 - a) Técnica Linking (**Mapas 6 e 7**) e Map Movie (**Mapas 8**): permitem definir *clusters* espaciais (regimes).
 - b) Diagrama condicional: permite verificar se a relação entre 2 variáveis quaisquer varia conforme a direção adotada (Ex. Norte, Sul, Leste, Oeste). Neste caso, haveria heterogeneidade entre as relações espaciais (**Figura 6**).

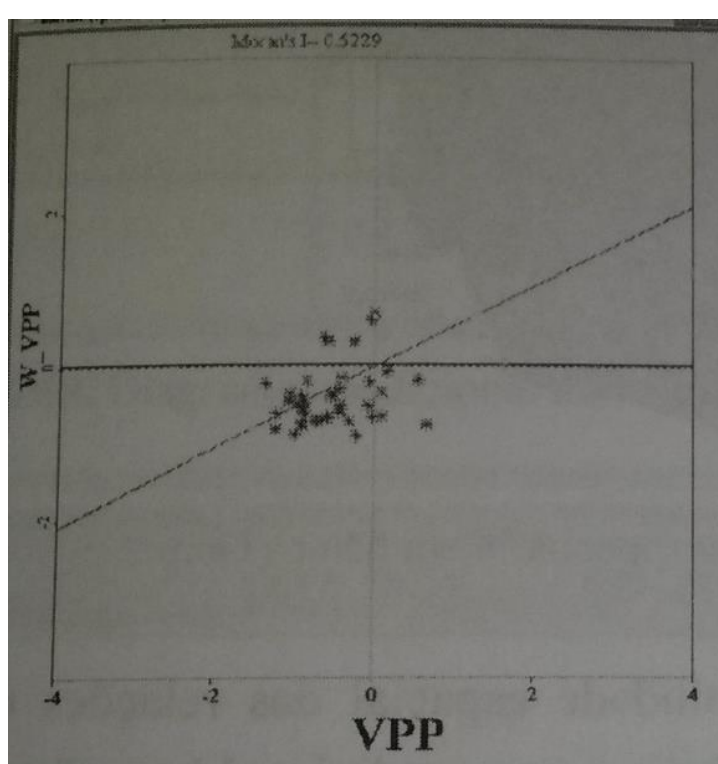
Nota: “a” e “b” são úteis quando o problema está no intercepto ou na inclinação.

- c) Box Splot dos Resíduos de uma regressão: detecta *outliers* nos resíduos (**Fig. 7**).

Nota: o controle dos *outliers* nos resíduos geralmente elimina a não normalidade e atenua a heterocedasticidade.

7. Heterogeneidade

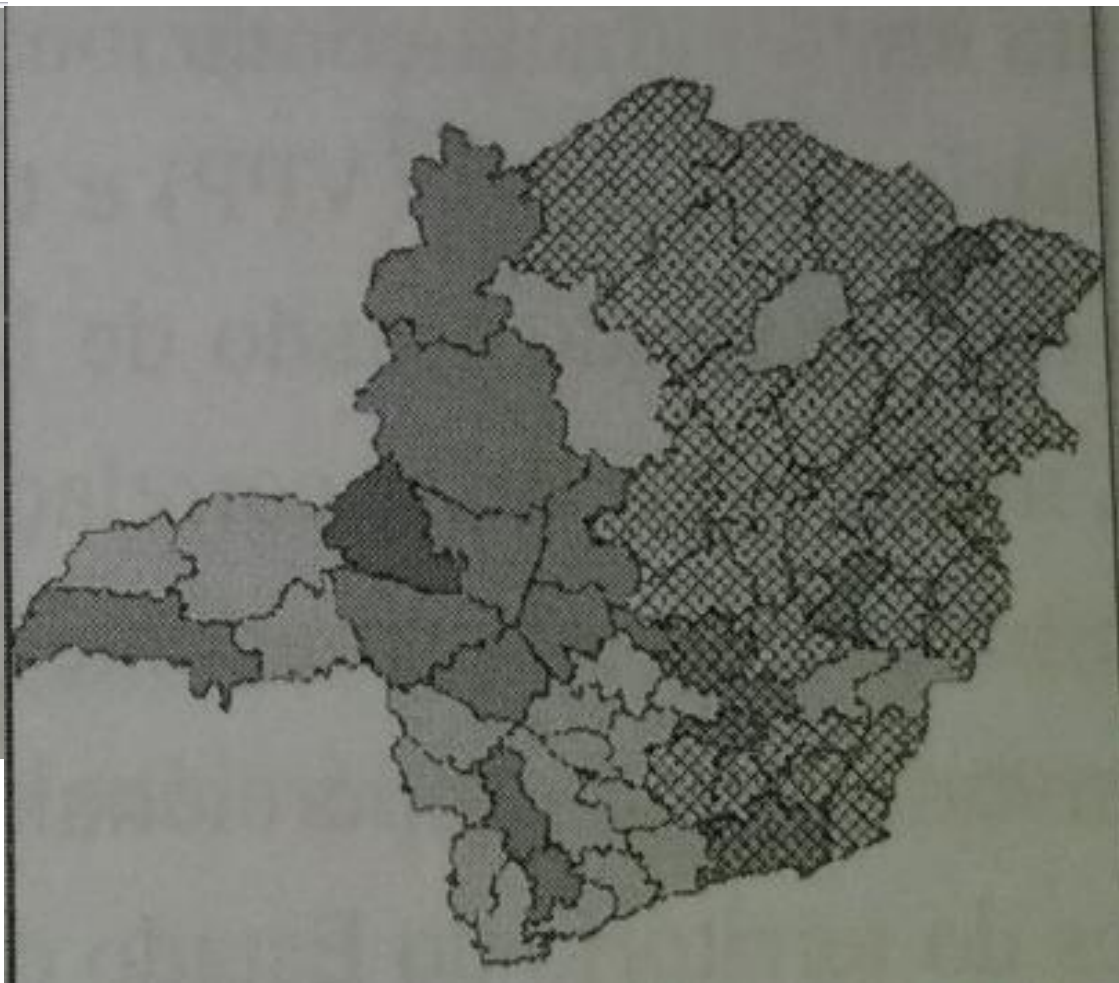
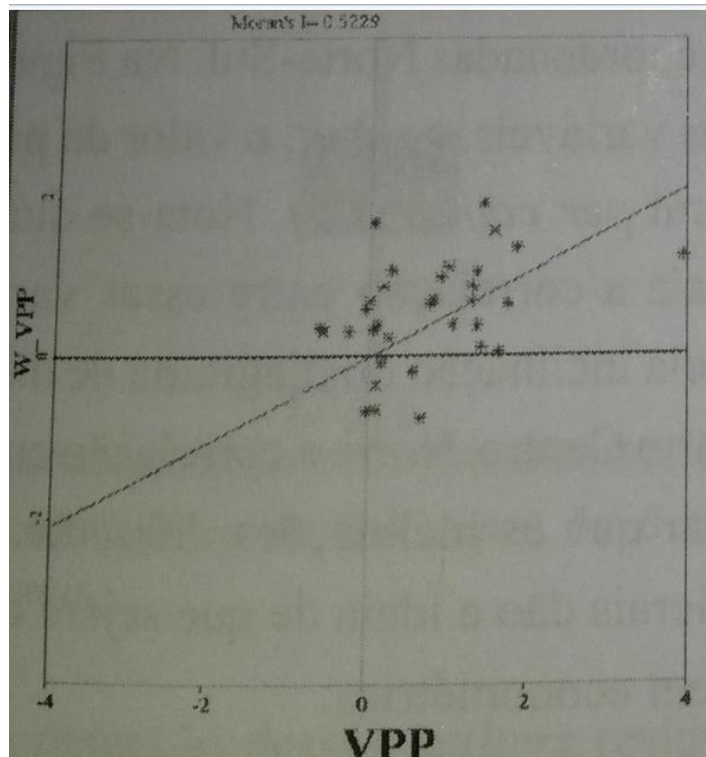
Mapa 6. Técnica *Linking* para detectar regimes do tipo AA.



Nota: AA = quadriculados

7. Heterogeneidade

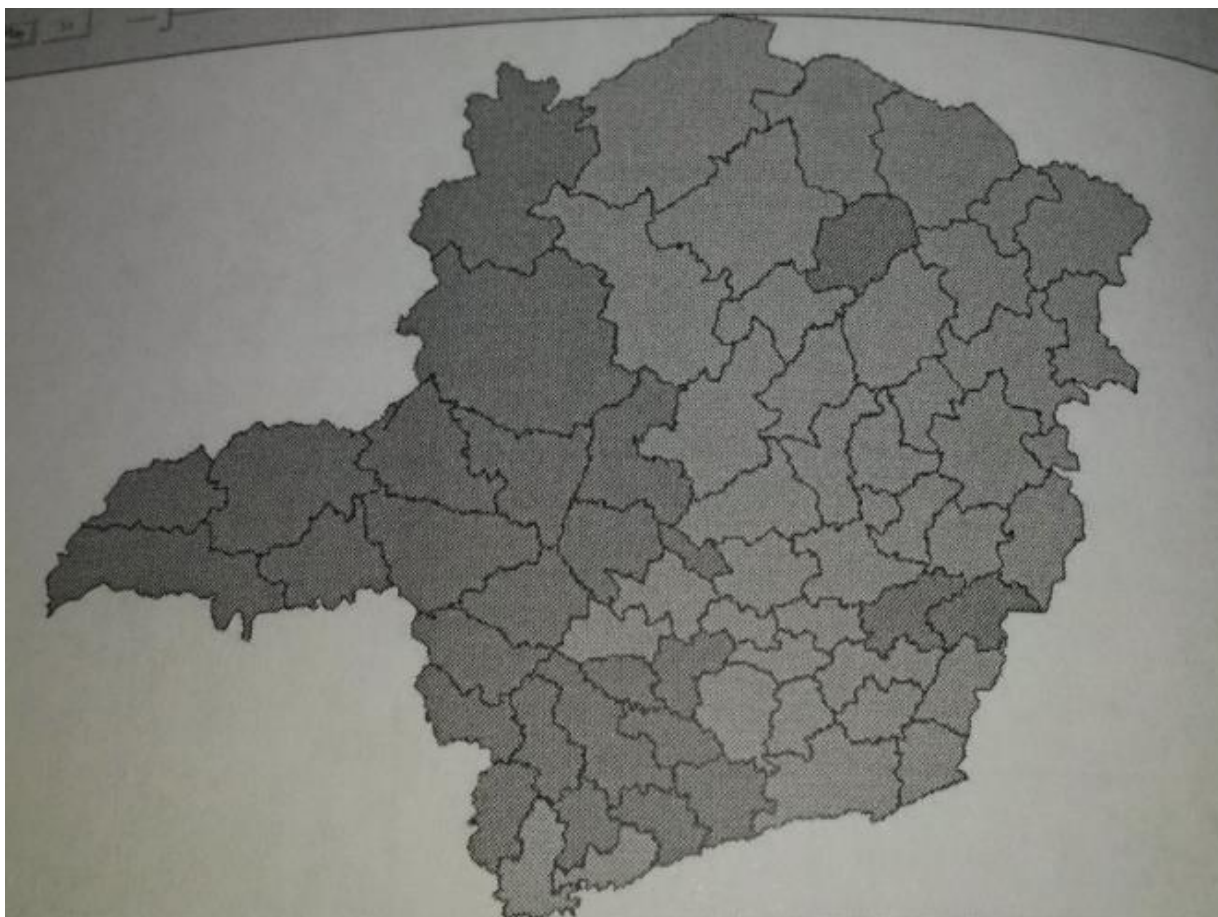
- **Mapa 7.** Técnica *Linking* para detectar regimes do tipo BB.



Nota: BB = quadriculados

7. Heterogeneidade

Mapa 8. *Map Movie* para detectar regimes espaciais.



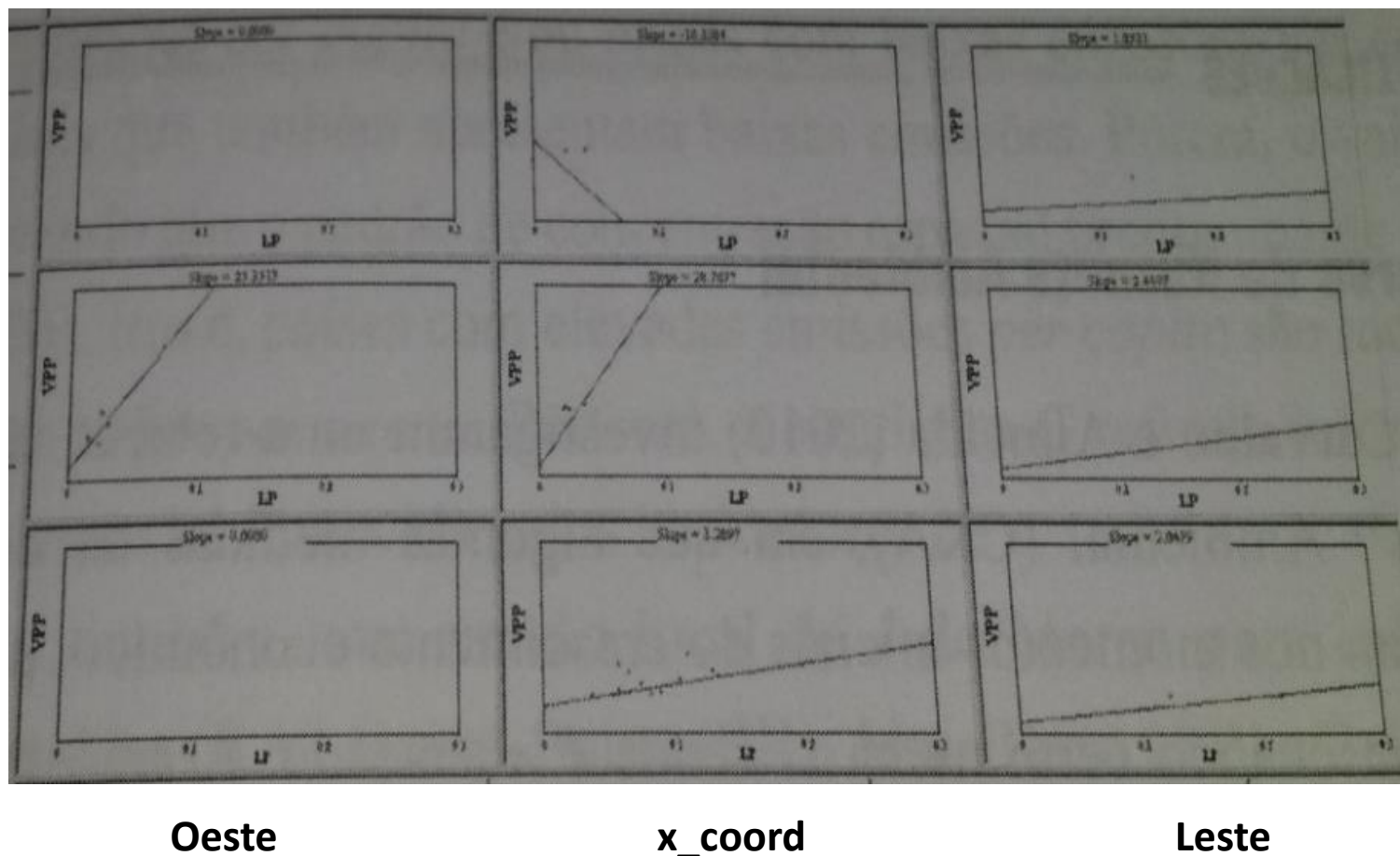
Nota:

Alto – cinza escuro

Baixo – cinza claro

7. Heterogeneidade

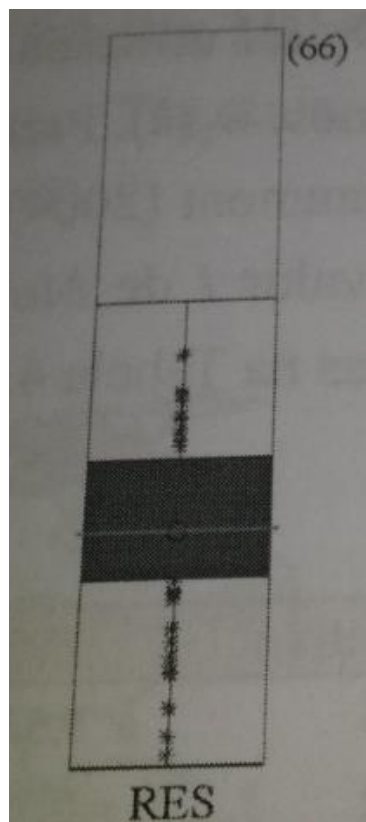
- **Figura 6.** Diagrama Condicional entre Produção Agropecuária per capita (VPP) e trabalho rural *per capita* (LP).



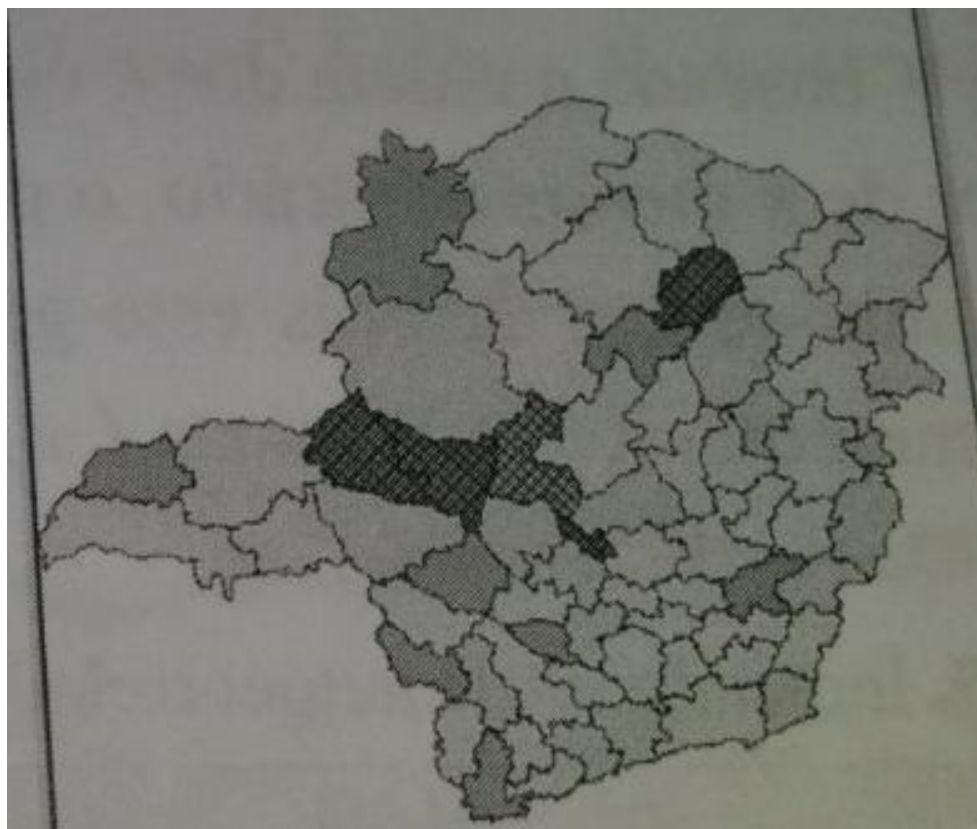
7. Heterogeneidade

Figura 7. *Box Plot* e *Box Map* dos Resíduos (RES)

a) *Box Plot*



b) *Box Map*



Nota:

a) 66 Observações

b) *Outlier* = Quadriculado



UNIVERSIDADE FEDERAL DE JUIZ DE FORA - *CAMPUS* Governador Valadares
ECONOMETRIA ESPACIAL
Prof. Vinícius de Azevedo Couto Firme

Referência

1. ALMEIDA, E. *Econometria Espacial Aplicada*. 1ª ed. Alínea, 2012.