

Problemas com dados espaciais

Patrícia de Siqueira Ramos

UNIFAL-MG, *campus* Varginha

27 de Maio de 2019

Além da dependência espacial e da heterogeneidade espacial (os efeitos espaciais), existem outros problemas que podem afetar a análise desse tipo de dados:

- falácia ecológica
- unidade de área modificável (MAUP) - sensibilidade dos resultados em relação à dimensão e à configuração da área de estudo
- efeito de beirada
- influência de *outliers* espaciais

1 Falácia ecológica

- perigo de se tirarem conclusões sobre o indivíduo baseadas em dados agregados (regionalmente ou por área geográfica)
- inferência ecológica: deduzir o comportamento do indivíduo a partir da análise de dados agregados
- falácia ecológica: erros decorrentes dessa inferência, levando a resultados diferentes dos que seriam obtidos se dados individuais fossem usados
- conclusões sobre o comportamento individual devem se basear em dados individuais e conclusões sobre um agregado de indivíduos devem ser fundamentadas em dados agregados

1 Falácia ecológica

- Mas algumas razões fazem com que tenhamos que inferir, mesmo que parcialmente, comportamentos individuais a partir de dados agregados:
 - 1 - alguns comportamentos individuais são influenciados pelo comportamento do grupo de indivíduos (vandalismo, desempenho escolar de um estudante)
 - 2 - alguns estudos SÓ podem ser realizados por área (dados eleitorais - disponibilidade de dados individuais de voto é uma impossibilidade por causa do direito do voto secreto)

2 Problema da unidade de área modificável (MAUP)

- os resultados de uma análise de dados dependem da definição do critério usado para a agregação espacial dos dados
- para introduzir o problema, suponha que, para uma mesma área de estudo, os dados podem estar agregados em escalas diferentes
- na figura a seguir, os dados da área de estudo foram agregados de 16 regiões (A) para 8 (B e D) e para 4 (C e E)
- os valores de B e C são as médias dos valores que aparecem em A

Ex.: Problema de agregação e zoneamento

A

x=5 y=3	x=4 y=5	x=5 y=6	x=4 y=2
x=2 y=3	x=7 y=6	x=3 y=2	x=3 y=4
x=2 y=1	x=5 y=5	x=3 y=4	x=5 y=3
x=1 y=2	x=6 y=4	x=3 y=1	x=4 y=4

B

4,5 4,0	4,5 4,0
4,5 4,5	3,0 3,0
3,5 3,0	4,0 3,5
3,5 3,0	3,5 2,5

C

4,5 4,3	3,8 3,5
3,5 3,0	3,8 3,0

D

3,5 3,0	5,5 5,5	4,0 4,0	3,5 3,0
1,5 1,5	5,5 4,5	3,0 2,5	4,5 3,5

E

2,5 2,5	5,5 5,0	3,5 3,3	4,0 3,3
------------	------------	------------	------------

F

4,7 4,7	4,0 2,0
3,6 3,1	4,0 3,7

	n	\bar{x}	\bar{y}	s_x	s_y	r_{xy}	CV_x	CV_y
A	16	3,88	3,44	1,59	1,59	0,66	40,94	46,26
B	8	3,88	3,44	0,58	0,68	0,88	15,03	19,73
C	4	3,88	3,44	0,42	0,61	0,94	10,88	17,79
D	8	3,88	3,44	1,33	1,24	0,95	34,31	36,00
E	4	3,88	3,44	1,25	1,05	0,98	32,26	29,88
F	4	4,07	3,38	0,46	1,13	0,67	11,22	33,47

Neste problema, há sensibilidade dos resultados devido a diferentes escalas espaciais:

- De A para C, as médias não se alteram, mas os CVs caem e o coeficiente de correlação aumenta
- Assim, os resultados mudam à medida que há um aumento no número de unidades espaciais num determinado agrupamento

Problema de zoneamento (ou partição)

- Este problema é uma sofisticação do primeiro: mesmo mantendo constante a escala espacial, há diferentes formas de arranjar as unidades espaciais (zoneamentos ou partições)
- No exemplo, em C, E e F, mesmo mantendo a escala espacial em 4 regiões, o zoneamento é diferente
- A média, o desvio padrão, a correlação de x e y e o CV das áreas E e F são diferentes entre si
- Assim, há sensibilidade dos resultados em função de diferentes combinações de unidades espaciais contíguas em zonas, dada uma mesma escala

- Estudar o zoneamento é importante em algumas situações, por exemplo, com um sistema eleitoral baseado em voto distrital, a definição dos distritos é decisiva
- Se for adotado um determinado zoneamento, um partido pode vencer as eleições e se for adotado outro, ele pode perder

- MAUP = problema de escala + problema de zoneamento
- O MAUP se manifesta na situação em que os resultados da análise são sensíveis à forma como os dados espaciais são organizados (nível de escala e zoneamento)
- Uma escolha inadequada da escala geográfica pode gerar dependência espacial artificial
- O MAUP limita a possibilidade de replicação de um modelo a uma outra região de estudo, se a escala e o zoneamento forem distintos daqueles da aplicação inicial

- Se os dados fossem distribuídos aleatoriamente no espaço, o problema de escala não apareceria
- O problema de escala vincula-se ao efeito da dependência espacial
- Se o espaço fosse absolutamente homogêneo, qualquer combinação de unidades espaciais em zonas forneceria os mesmos resultados (não haveria o problema de zoneamento)

3 Efeito de beirada

- Observações próximas da fronteira estão correlacionadas espacialmente com regiões que não fazem parte da área de estudo original
- Além disso, as regiões na beirada da área de estudo costumam ter menos vizinhos que as regiões localizadas no interior

3 Efeito de beirada

- Exemplo de falta de influência do efeito de fronteira: uma epidemia em uma ilha isolada

3 Efeito de beirada

- Exemplo de falta de influência do efeito de fronteira: uma epidemia em uma ilha isolada
- Exemplo de influência: produção de cana-de-açúcar no estado de São Paulo
- Soluções possíveis:
 - “dobrar o mapa”
 - estender a área de estudo para além da fronteira

4 *Outliers* e pontos de alavancagem

- Observações do tipo HL ou LH podem ser originadas por erros de medida, mas podem ser legítimas também
- Exemplo: cidades que são capitais, um condomínio fechado em torno de bairros com elevada criminalidade