

AUTOCORRELAÇÃO ESPACIAL NA ÁREA DE LOJA DE SUPERMERCADOS NOS MUNICÍPIOS PAULISTAS: mensuração por meio do Índice de Geary

Francisco Aranha

Professor do Departamento de Informática e Métodos Quantitativos da
EAESP/FGV e Consultor de Empresas.

E-mail: faranha@fgvsp.br

RESUMO

Na construção de modelos para estimação de potencial de mercado com utilização de regressão linear simples e múltipla, freqüentemente são obtidos resultados insatisfatórios devido à ocorrência de autocorrelação espacial. O Índice de Geary permite identificar a presença desse fenômeno e mensurar sua intensidade. Incorporado a um modelo para estimativa de área de loja de supermercados em municípios paulistas, o conceito de autocorrelação melhora substancialmente o desempenho do modelo.

ABSTRACT

In modeling market potential with simple and multiple regression, unsatisfactory results are frequently obtained due to the presence of spatial autocorrelation in the data. The Geary's Ratio can identify and measure the intensity of this phenomenon. The concept of autocorrelation substantially improves the performance of a model to estimate supermarket floor space in São Paulo state counties.

PALAVRAS-CHAVE

Autocorrelação espacial, modelagem, potencial de mercado, regressão, supermercados.

KEY WORDS

Spatial autocorrelation, modeling, market potential, regression, supermarkets.

INTRODUÇÃO

Uma das aplicações mais comuns de métodos quantitativos ao *marketing* diz respeito à estimativa do potencial de mercado das empresas. Na construção de modelos para essa finalidade, freqüentemente são utilizadas as técnicas de regressão linear simples e múltipla. Os resultados, no entanto, nem sempre são satisfatórios, e um dos principais empecilhos à obtenção de boas estimativas é a interferência do potencial de uma região sobre as regiões vizinhas. Neste artigo apresentamos uma medida, o Índice de Geary, que permite identificar a presença da influência das variáveis associadas a uma localização e mensurar a intensidade dessa influência sobre as realizações dessas mesmas variáveis em localizações próximas. O conceito é aplicado ao caso de área de loja de supermercados em municípios paulistas.

A RELAÇÃO ENTRE DENSIDADE DE RENDA E DENSIDADE DE LOJA DE SUPERMERCADOS EM MUNICÍPIOS PAULISTAS

Em artigo publicado recentemente (Aranha, 1998), procuramos estimar a relação entre a densidade de renda observada em um município e a quantidade de m² de loja de supermercados ali existente. O modelo proposto mostrou-se significativo, mas com baixo poder de explicação da variabilidade total da metragem de loja dos municípios.

Uma das possíveis causas identificadas para o desempenho insuficiente ($R^2 = 49\%$) do modelo foi a presença, nos dados, de autocorrelação espacial: os fatores determinantes da alta densidade de loja em um município acabam influenciando também os municípios vizinhos. Dessa forma, esses “vazamentos” do efeito das variáveis explicativas, “transbordando” de um município para o outro, reduzem as diferenças da metragem de loja que seriam observadas entre os municípios caso o “vazamento” não existisse. A manifestação aparente do fenômeno oculta parcialmente a relevância e a capacidade de explicação das variáveis utilizadas.

Formulação de um novo modelo

Na validade da hipótese acima, uma melhor compreensão dos fatores que influenciam a área de loja de supermercados existente nos municí-

pios depende da confirmação da presença da autocorrelação espacial, sua quantificação e eventual remoção. Os dados “limpos” de autocorrelação podem, então, ser utilizados na estimativa do efeito da densidade de renda sobre a densidade de loja.

O Índice de Geary permite identificar a presença da influência das variáveis associadas a uma localização e mensurar a intensidade dessa influência sobre as realizações dessas mesmas variáveis em localizações próximas.

AUTOCORRELAÇÃO ESPACIAL

A grande maioria dos fenômenos não se distribui no espaço casualmente. Por exemplo, o relevo de uma região pode ser resultado de falhas geológicas; a presença de uma espécie vegetal pode depender de certas combinações de clima e solo; a concentração de população num país pode refletir o roteiro histórico de sua colonização; a industrialização em determinados estados pode demonstrar a existência de recursos naturais. Relevo, espécies vegetais, população e indústrias apresentam-se, portanto, organizados de uma maneira que pode ser prevista, pelo menos em parte, uma vez conhecidos os seus determinantes.

Tomemos o objeto de nosso interesse, a quantidade de supermercados existente nos municípios do estado de São Paulo, medida em m² de loja.¹ Imaginemos que a disposição espacial do total de área de loja em cada município tivesse sido determinada única e exclusivamente por acaso, por meio de um sorteio: em uma urna haveria tantas bolas quanto são os municípios, cada uma registrando um valor de m² de loja; a metragem construída em cada cidade seria, então, definida pela extração de uma bola da urna.

Nesse caso, ao fazermos nosso estudo, a situação de cada município traria informação totalmente nova: antes de ler a cifra

marcada na bola, nada poderíamos afirmar sobre o município.

Imaginem, agora, uma situação diversa, em que o total de loja de cada cidade tivesse sido determinado em duas etapas. Em cada etapa seria especificada uma, de duas quantidades aditivas.

A autocorrelação espacial refere-se à redundância de informação entre duas realizações de um fenômeno quando elas ocorrem próximas uma da outra. Sua presença distorce os resultados obtidos pela aplicação de modelos estatísticos tradicionais.

Para se definir a primeira quantidade, calcularíamos a média de m^2 de loja dos municípios vizinhos a um município em foco: esse seria o valor-base do município em foco. Para definir a outra parte, uma bola seria extraída da urna de metragens e, se o valor registrado na bola fosse positivo, ele seria somado ao valor-base; caso fosse negativo, seria subtraído.

Nesse processo em duas etapas, ao fazermos um levantamento sobre um município específico, a informação obtida não seria totalmente nova. Ela seria, em parte, uma repetição do que já sabíamos em função do estudo dos municípios vizinhos.

A autocorrelação espacial refere-se a essa redundância de informação entre duas realizações de um fenômeno quando elas ocorrem próximas uma da outra. Sua presença distorce os resultados obtidos pela aplicação de modelos estatísticos tradicionais, baseados na hipótese de independência entre as realizações da variável de interesse. Uma consequência importante da presença desse fenômeno é o abafamento ou, ao contrário, a exacerbação das relações de causalidade ou associação.

A tendência de valores localizados próximos serem parecidos caracteriza uma autocorrelação positiva. Se aplicamos fertilizantes em certas áreas agrícolas, por exemplo,

áreas próximas podem beneficiar-se de parte do fertilizante, carregado pelas chuvas, resultando em uma maior produtividade também das áreas vizinhas não fertilizadas. Esse tipo de dinâmica deprime a manifestação local de diferenças na produtividade e reforça as diferenças regionais. Também oculta, frequentemente, a importância de variáveis explicativas subjacentes ao processo estudado; no caso, o efeito do fertilizante, uma vez que tanto áreas tratadas quanto não tratadas (diretamente) acabam se beneficiando da fertilização.

Existe também a situação inversa. Algumas espécies de pássaros, por exemplo, evitam formar grupos muito numerosos, procurando manter uma razoável distância entre os bandos. Assim, sabendo-se que numa determinada área há uma elevada concentração dessas aves, podemos prever com segurança que nas regiões vizinhas elas aparecerão em pequeno número.

Esse último exemplo caracteriza-se por uma autocorrelação espacial negativa, em que ocorrências bastante diferentes entre si tendem a estar próximas: muitos pássaros numa área, poucos pássaros na área vizinha. Nessa dinâmica, as diferenças locais do fenômeno são exacerbadas, ao contrário das diferenças regionais, que são enfraquecidas.

Seja positiva ou negativa, a presença de autocorrelação influencia a quantidade e a qualidade da informação contida em dados espacializados, afetando as interpretações possíveis. Os testes estatísticos tradicionais, baseados no pressuposto de independência entre as observações, resultam distorcidos quando aplicados a dados com essas características. Da mesma forma, a distribuição das estimativas de parâmetros populacionais baseadas em amostras não corresponde às formas funcionais tradicionais.

Como a autocorrelação surge pela influência das realizações vizinhas sobre a realização do fenômeno na área de interesse, sua intensidade geralmente decai com a distância, tornando-se irrelevante quando as observações estão suficientemente afastadas entre si. Em outras palavras, a partir de uma certa distância, a rega de um lote não influencia mais o crescimento de plantas em outro. Um dos aspectos que interessa conhecer sobre a autocorrelação, portanto, é a relação de sua intensidade com a distância entre observações.

Nas explicações acima, e nas que se seguirão, baseamo-nos nos trabalhos de Ripley (1981), Fotheringham e Rogerson (1994) e Griffith (1987), especialmente no trabalho deste último, inteiramente dedicado ao tema do tratamento estatístico da autocorrelação espacial.²

Correlação e autocorrelação

O conceito de “correlação” no contexto estatístico diz respeito ao relacionamento entre duas variáveis.

Quando ambas as variáveis são intervalares, geralmente se utiliza como medida da intensidade de sua relação o Coeficiente de Pearson (ρ), que mede o grau de correlação linear entre elas. Nesse caso, o coeficiente de correlação varia de -1 a 1 . Quando $|\rho|$ tende a zero, a relação entre as variáveis é fraca. Quando $|\rho|$ tende a 1 , a relação é forte. O sinal do coeficiente indica a natureza da relação: se negativo, mostra que ao crescimento de uma variável corresponde uma diminuição da outra; se positivo, sinaliza que as duas variáveis caminham na mesma direção, isto é, crescem ou decrescem juntas.³

O conceito de autocorrelação surge quando se verificam duas condições:

- as realizações de uma variável estão ordenadas de alguma maneira;
- cada realização (realização em foco) da variável relaciona-se com outra ou outras realizações da mesma variável que ocupam uma determinada posição em relação à realização em foco.

Assim, por exemplo, as cotações do preço de uma *commodity* estão ordenadas no tempo, formando uma série temporal. A cada dia, a cotação não varia livremente: em geral, ela será muito parecida com a cotação da véspera, já que as oscilações diárias, pelo menos em períodos de tranquilidade econômica, costumam ser pequenas e partem do ponto em que o pregão se encerrou na véspera. Em outras palavras, elas não são independentes.

Em geral, não só a ocorrência imediatamente anterior é relevante. No acompanhamento das vendas do comércio, por exemplo, costuma-se comparar os resultados obtidos pelo varejo no mês precedente e no mesmo mês de anos anteriores.

Dados organizados no tempo são relativamente mais fáceis de analisar do que dados organizados geograficamente. No tempo, os

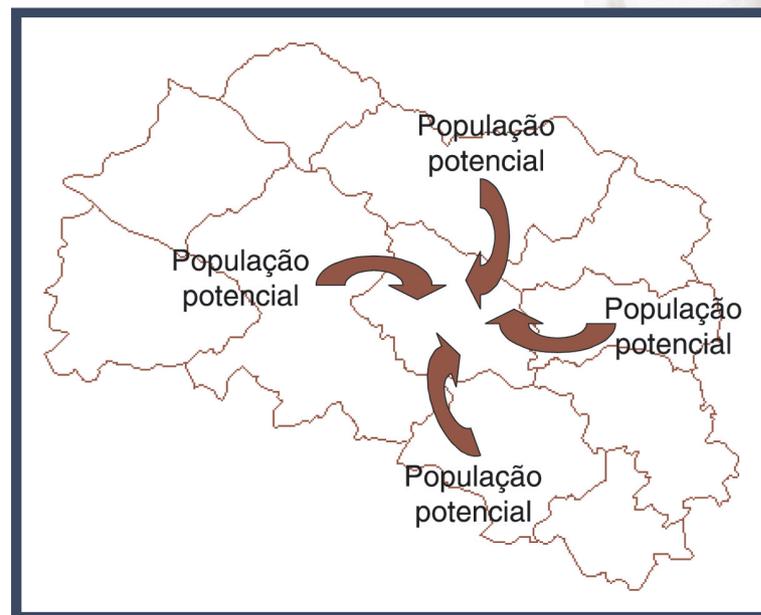
fenômenos organizam-se apenas em uma direção e em um único sentido: as cotações do próximo ano não afetam as cotações de hoje (ou de amanhã), simplesmente porque ainda não aconteceram. Já no espaço todas as direções, e todos os sentidos, podem ser relevantes. O potencial de consumo de uma cidade pode ser afetado por seus vizinhos a norte, a sul, a leste, a oeste, etc.

Em resumo, a autocorrelação espacial ocorre quando realizações (de uma única variável) organizadas no espaço influenciam-se mutuamente, e esse é o caso da concentração de supermercados. Cidades com alta concentração populacional e alta concentração de renda (portanto com muitos supermercados) tendem a estar próximas de outras cidades de alta densidade (populacional, de renda e de supermercados).

Até porque é razoável imaginar que o potencial econômico “extravasa” de um município para o outro, carregado pelas pessoas que se deslocam de uma cidade para a outra, como na nossa ilustração do fertilizante; seria raro, como regra geral, encontrar um município com densidade de loja extremamente alta ao lado de outro município com densidade extremamente baixa.

A Figura 1 representa a possibilidade de a população de municípios vizinhos gerar demanda no município de interesse.

Figura 1 – Deslocamento de potencial: a população dos municípios vizinhos pode ser atraída para um centro regional



Uma medida de autocorrelação espacial

Como as variáveis de interesse neste trabalho – densidade de renda e densidade de loja – são intervalares, uma medida da autocorrelação espacial pode ser o Índice de Geary (GR = Geary Ratio), dado pela seguinte fórmula (Griffith, 1987):

$$GR = [(n-1) \sum_{i=1}^n \sum_{j=1}^n c_{ij} (x_i - x_j)^2] / [2 \sum_{i=1}^n \sum_{i=1}^n c_{ij} \sum_{i=1}^n (x_i - \bar{x})^2]$$

em que:

- x_i e x_j são os valores da variável de interesse nas cidades i e j ;
- \bar{x} é a média da variável de interesse em todas as cidades da amostra;
- n é o número de cidades na amostra;
- c_{ij} assume o valor 1, se x_i e x_j forem vizinhos, e 0, caso contrário.

Podemos reescrever a expressão apresentada por Griffith de uma forma mais conveniente para a interpretação, como se segue:

$$GR = \frac{\sum_{i=1}^n \sum_{j=1}^n c_{ij} (x_i - x_j)^2}{2 \sum_{i=1}^n \sum_{i=1}^n c_{ij}} / \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{(n-1)} \quad (1)$$

Variabilidade de vizinhança
Variabilidade geral

Variabilidade de vizinhança

Consideremos o primeiro bloco – à esquerda do traço vertical de fração na fórmula (1) anterior –, que denominamos de “variabilidade de vizinhança”.

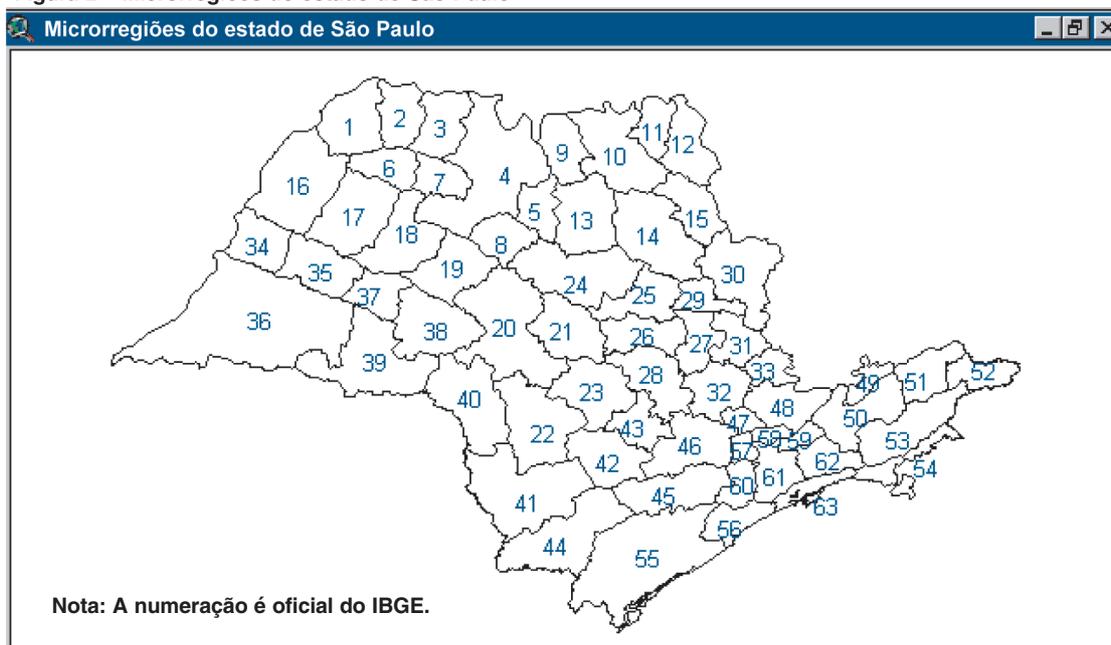
Nesse quociente, não fosse por c_{ij} , o numerador representaria a soma das diferenças ao quadrado entre cada particular x e todos os demais. No entanto, como c_{ij} assume o valor 0 quando os elementos comparados não são vizinhos, cancelando o resultado da comparação, isto é, anulando aquele particular desvio, acabam só entrando na soma os desvios entre os vizinhos, já que estes são multiplicados por 1 e, portanto, não são afetados. No denominador, $\sum_{i=1}^n \sum_{i=1}^n c_{ij}$ corresponde ao número total de parcelas somadas, e o 2 ajusta o fato de que a medida de dispersão de todos os elementos de um conjunto em relação a todos os demais resulta no dobro da dispersão de todos os elementos em relação ao meio do conjunto,⁴ isto é, resulta no dobro da variância.⁵

Em outras palavras, o valor calculado nesse bloco corresponde à variância (espalhamento) das observações definidas por c_{ij} como vizinhas entre si.

Variabilidade do conjunto total de dados

O segundo bloco, após o traço vertical de fração em (1), é, obviamente, a variância –

Figura 2 – Microrregiões do estado de São Paulo



uma medida do espalhamento quadrático médio – das observações do conjunto de dados.

Comparando as variabilidades

Se a quantidade de supermercados for definida espacialmente por acaso em cada município, isto é, se a metragem de loja em uma cidade não tiver nenhuma relação com a metragem das cidades vizinhas, a variabilidade de metragem no conjunto nos dados e no subconjunto dos vizinhos são duas estimativas da mesma grandeza, a variabilidade de metragem entre as cidades em geral. Assim, seu quociente deve ser próximo de 1.

Se, contudo, cidades próximas tiverem uma tendência a serem parecidas, a variância de vizinhança será menor do que a variância geral, e o GR será menor que 1.

Ao contrário, se cidades com grandes metragens ficarem circundadas por cidades com pouca metragem (ou vice-versa), a variabilidade entre os vizinhos será maior que a variabilidade geral, e o GR será maior que 1.

Tipicamente, o GR assume valores entre 0 e 2.

Microrregiões

Por limitações de capacidade de processamento do *software* que utilizamos, em vez de trabalhar com os municípios isoladamente, os dados disponíveis foram agrupados segundo as microrregiões definidas pelo IBGE.

Foram consideradas vizinhas as microrregiões que compartilhavam uma fronteira. As relações de vizinhança foram representadas numa matriz de conectividade.

Nessa matriz estão representados os valores de c_{ij} , dispostos em uma tabela na qual as linhas representam as microrregiões, e as colunas, as microrregiões vizinhas. Conforme observamos anteriormente, quando ocorre a vizinhança, a célula assume o valor 1 e, caso contrário, assume o valor 0.

Assim, por exemplo, a primeira linha da matriz

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	
1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46			
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
47	48	49	50	51	52	53	54	55	56	57	58													
0	0	0	0	0	0	0	0	0	0	0	0													

nos informa que os vizinhos da microrregião 1 (linha 1) são as microrregiões 2, 6 e 16

(correspondendo a colunas assinaladas com 1), o que pode ser facilmente conferido no mapa anterior (Figura 2).

Uma alternativa à representação tabular é o gráfico apresentado a seguir (Figura 3),⁶ de mais fácil visualização.

GR da distribuição de metragem de loja

Com o auxílio de um programa escrito em MatLab (Hanselman e Littlefield, 1997),⁷ calculamos o GR para a metragem de loja, considerando a matriz de conectividade definida acima, obtendo-se:

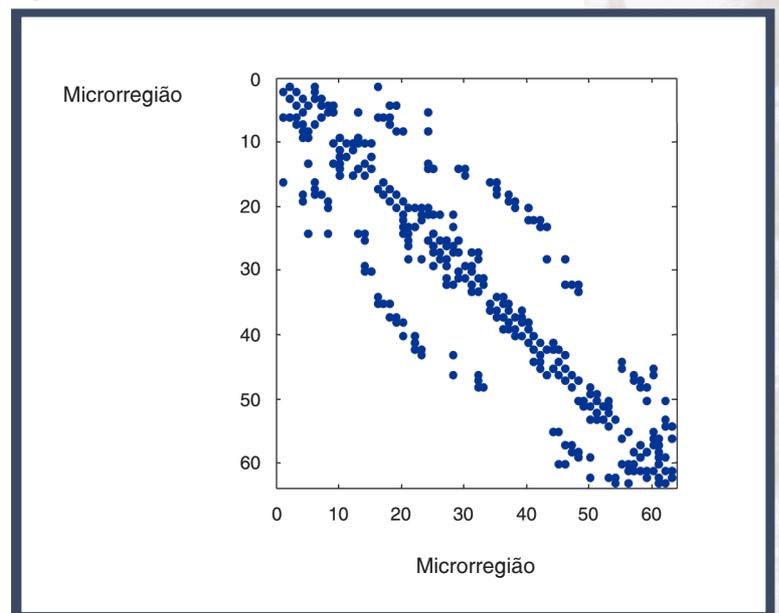
- variância da metragem de loja entre os vizinhos = 0,30
- variância da metragem de loja entre todas as microrregiões = 0,53
- GR (Índice de Geary) = 0,57

Assim, podemos perceber que os vizinhos são bem mais parecidos entre si do que o conjunto de todas as microrregiões. A variabilidade dos vizinhos é quase a metade da variabilidade geral (GR de 0,57).⁸

Remoção da autocorrelação espacial da variável dependente

Consideremos que a metragem de loja de cada microrregião é definida pela média da metragem de loja das microrregiões vizinhas mais um valor explicado por outros motivos, por hora chamados, em conjunto, de “acaso”. Esse modelo pode ser enunciado

Figura 3 – Matriz de conectividade



algebricamente como:

$$x_i = \rho \left(\frac{\sum_{j=1}^n c_{ij} x_j}{\sum_{j=1}^n c_{ij}} \right) + u_i \quad (2)$$

em que:

- x_i representa os m² de loja da cidade i
- $\rho \left(\frac{\sum_{j=1}^n c_{ij} x_j}{\sum_{j=1}^n c_{ij}} \right)$ representa a média de m² dos vizinhos da cidade i , ponderada pelo coeficiente de autocorrelação
- u_i representa a parcela de m² da cidade i não explicada pelos m² dos seus vizinhos

Calculando o valor de ρ por regressão linear simples, chegamos aos resultados⁹ apresentados na Tabela 1.

Esse resultado mostra que, com um nível

altíssimo de significância, um modelo de autocorrelação espacial explica cerca de 40% da variabilidade dos dados. Além disso, mostra que o valor esperado da densidade de m² de loja de um município é igual a 94% do valor da densidade média de seus vizinhos. A diferença entre o valor esperado e o observado, chamada de resíduo do modelo de autocorrelação, é o valor que precisa ainda ser explicado por outras variáveis independentes. Corresponde aos dados originais sem a parcela que pode ser explicada pela autocorrelação.

Remoção da autocorrelação da variável independente

Utilizando a mesma estratégia empregada para a densidade de loja, obtivemos o seguinte modelo de autocorrelação para a densidade de renda, conforme apresentado na Tabela 2.

Esses resultados indicam um modelo altamente significativo, capaz de explicar cerca de 60% da variabilidade total da densidade de renda com base na média dos vizinhos, com um coeficiente de autocorrelação de quase 1.

MODELO FINAL: METRAGEM DE LOJA COMO FUNÇÃO DA DENSIDADE DE RENDA AUTOCORRELACIONADA E DA RENDA AUTÔNOMA

Após o exame de várias alternativas,¹⁰ formulamos um modelo em que a densidade de m² de loja é explicada pela densidade da renda do município desmembrada em duas partes, cada uma funcionando como uma variável independente: a densidade de renda autocorrelacionada com os vizinhos e o resíduo, isto é, a parcela não explicada por autocorrelação. Ajustado por regressão múltipla sobre os dados convertidos em escala logarítmica, o modelo está descrito na Tabela 3, a seguir.

Ele corresponde à seguinte equação:

$\log(\text{dens m}^2 \text{ loja}) = -5.88 + 1,02 (\text{dens. de renda autocorrelacionada}) + 1,20 (\text{resíduo da dens. de renda após remoção da autocorrelação})$

Tabela 1 – Resumo dos resultados

Estatística de regressão	
R múltiplo	0.64
R-quadrado	0.41
R-quadrado ajustado	0.40
Erro-padrão	0.56
Observações	63

ANOVA					
	g/	SQ	MQ	F	F de significação
Regressão	1	13.65	13.65	43.80	1.04616E-08
Resíduo	62	19.33	0.31		
Total	63	32.98			

	Coeficientes	Erro-padrão	Stat t	Valor-P
Média dos vizinhos	0.94	0.08	11.57	3.98E-17

Tabela 2 – Resumo dos resultados

Estatística de regressão	
R múltiplo	0.76
R-quadrado	0.58
R-quadrado ajustado	0.57
Erro-padrão	0.39
Observações	63

ANOVA					
	g/	SQ	MQ	F	F de significação
Regressão	1	13.21	13.21	87.20	2.2899E-13
Resíduo	62	9.40	0.15		
Total	63	22.61			

	Coeficientes	Erro-padrão	Stat t	Valor-P
Média dos vizinhos	0.99	0.01	130.37	2.4E-77

Interpretação

O modelo, que é significativo e tem resíduos bem comportados, mostra-nos que a densidade de supermercado de uma microrregião depende:

- em parte, da densidade de renda das microrregiões vizinhas;
- em parte, do componente (autônomo) da densidade de renda da própria microrregião, isto é, da parcela da sua densidade de renda que é independente da densidade de renda dos vizinhos. Esta última tem um peso maior na determinação da densidade de área de loja na própria microrregião.

Retornando os valores das estimativas para a escala original, temos uma correlação de 0,944 entre os valores previstos e observados segundo o modelo para a metragem de loja das microrregiões, portanto uma capacidade de explicação de quase 90% da variabilidade total dos dados.

CONCLUSÕES

- Havia autocorrelação espacial nos nossos dados de densidade de loja e de densidade de renda.

Tabela 3 – Resumo dos resultados

Estatística de regressão	
R múltiplo	0.91
R-quadrado	0.84
R-quadrado ajustado	0.83
Erro-padrão	0.30
Observações	63

ANOVA					
	g/	SQ	MQ	F	F de significação
Regressão	2	27.55	13.77	152.20	3.14049E-24
Resíduo	60	5.43	0.09		
Total	62	32.98			

	Coefficientes	Erro-padrão	Stat t	Valor-P
Interseção	(5.88)	0.54	(10.80)	0.00
Dens. e auto.	1.02	0.08	12.06	0.00
Dens. e res.	1.20	0.10	12.20	0.00

- Essa autocorrelação fazia parecer menor a influência da densidade de renda sobre a densidade de loja (R^2 de 49% sem isolamento da autocorrelação espacial, contra R^2 de 90% quando a autocorrelação é explicitada e levada em consideração separadamente).
- A remoção da autocorrelação espacial permitiu elaborar um modelo com alta capacidade de explicação da densidade de loja, agora passível de utilização em exercícios de previsão. ○

REFERÊNCIAS BIBLIOGRÁFICAS

ARANHA, Francisco. Indicadores de potencial de consumo dos municípios paulistas. *RAE – Revista de Administração de Empresas*, v.38, n.4, p.18-25, out./dez. 1998.

FOTHERINGHAM, Stewart, ROGERSON, Peter (Eds.). *Spatial analysis and GIS*. London: Taylor & Francis, 1994.

GRIFFITH, Daniel A. *Spatial autocorrelation*. Washington: Association of American Geographers, 1987.

HANSELMAN, Duane, LITTLEFIELD, Bruce. *MatLab: the language of technical computing*. Upper Saddle River: Prentice-Hall, 1997.

RIPLEY, Brian D. *Spatial statistics*. NY: Wiley, Wiley Series in Probability and Statistics, 1981.

NOTAS

Este artigo baseia-se em pesquisa financiada pelo NPP – Núcleo de Pesquisas e Publicações da EAESP/FGV intitulada “Autocorrelação espacial na distribuição de área de loja de supermercados em municípios do Estado de São Paulo”.

1. Detalhes sobre o levantamento realizado para a obtenção desses dados podem ser encontrados em ARANHA, Francisco. *Potencial de consumo dos municípios paulistas*. São Paulo: NPP/EAESP/FGV, 1998. (Relatório de Pesquisa n. 2). No projeto de pesquisa seguinte, que deu origem a este artigo (*Autocorrelação espacial na distribuição de área de loja de supermercados em municípios do Estado de São Paulo*, São Paulo: NPP/EAESP/FGV, Relatório de Pesquisa em editoração), os m² de loja foram totalizados por microrregião do estado.

2. Um dos desafios para a leitura da literatura sobre autocorrelação espacial é a exigência de uma matemática avançada, superior àquela de que dispõem, em geral, os profissionais com formação em Administração de Empresas. Esse trabalho de Griffith tem o grande mérito de tentar fazer uma ponte entre as abordagens de estatística básica e as mais avançadas.

3. Para maiores detalhes sobre o Coeficiente de Correlação de Pearson, por favor, refira-se a qualquer texto de estatística básica. Entre eles, sugerimos, por exemplo, STEVENSON, William J. *Estatística aplicada à Administração*. São Paulo: Harbra, 1981, p. 370-5.

4. Para uma análise detalhada desse índice, veja ARANHA, Francisco. *Autocorrelação espacial na distribuição de área de loja de supermercados em municípios do Estado de São Paulo*. São Paulo: NPP/EAESP/FGV. (Relatório de Pesquisa em editoração).

5. Para maiores esclarecimentos sobre a variância, por favor, consulte um texto de estatística básica.

6. Nos gráficos, os “0”s da tabela foram omitidos, e os “1”s, representados por pontos.

7. O programa foi reproduzido no relatório de pesquisa ARANHA, Francisco. *Autocorrelação espacial na distribuição de área de loja de supermercados em municípios do Estado de São Paulo*. São Paulo: NPP/EAESP/FGV. (Relatório de Pesquisa em editoração).

8. Com base em uma pequena simulação da distribuição da estatística, esse resultado é significativo ao nível de 5%.

9. Cálculos realizados com o programa Excel.

10. Uma discussão mais detalhada pode ser encontrada em ARANHA, Francisco. *Autocorrelação espacial na distribuição de área de loja de supermercados em municípios do Estado de São Paulo*. São Paulo: NPP/EAESP/FGV. (Relatório de Pesquisa em editoração).