Escola Superior de Agricultura “Luiz de Queiroz”
Pós-graduação em Estatística e Experimentação Agronômica

Disciplina: LCE 5700 – **Geoestatística**

Prof. Paulo Justiniano Ribeiro Junior

Disc. Iábita Fabiana Sousa – N°USP: 7041617

Resenha 2

Local influence of explanatory variables in Gaussian spatial linear models

Piracicaba - 2012

Local influence of explanatory variables in Gaussian spatial linear models

BORSSOI, J. A.; BASTIANI,F.; URIBE-OPAZO, M.A.; GALEA, M. Local influence of explanatory variables in Gaussian spatial linear models. ***Chilean Journal of Statistics***, Santiago, v. 2, n. 2, p. 29-38, 2011.

Na modelagem da variabilidade espacial de variáveis regionalizadas, o método de estimativa de máxima verossimilhança (MV) é usado para estimar os parâmetros que definem a estrutura de dependência espacial. Sendo também utilizado na interpolação de valores em áreas não amostrados pela técnica de krigagem, gerando mapas temáticos que poderiam ser utilizados para um tratamento localizado na área de estudo. A qualidade dos mapas dependem das inferências para os modelos escolhidos. Para que a técnica de krigagem der prognósticos e represente a verdadeira variabilidade do local, o processo deve ser realizado com cuidado, especialmente na presença de observações discrepantes ou influentes.

O objetivo do trabalho foi utilizar o método de influência local para modelos espaciais lineares Guassianos a fim de verificar a existência de observações que possam exercer algum tipo de influência sobre a verossimilhança de deslocamento quando existem perturbações na matriz de co - variáveis.

Considera-se um processo gaussiano estocástico $\{Z(s), s \in \mathbf{S}\}$ com $\mathbf{S} \subset \Re^d$, sendo $\Re^d$, o espaço euclidiano, d-dimensional ($d \geq 1$). Suponha que os elementos $Z(S_1), ..., Z(S_n)$ deste processo sejam registrados em localizações espaciais conhecidas s_i, para $i = 1, ..., n$, e gerados pelo seguinte modelo:

$$Z(s_i) = \mu(s_i) + \epsilon(s_i),$$

em que os termos determinísticos $\mu(s_i)$ e estocásticos $\epsilon(s_i)$, podem depender da localização espacial onde $Z(s_i)$ foi obtida. Assume-se que o erro estocástico $\epsilon(.)$ tem média zero, e que e que a variação entre pontos no espaço é determinada por alguma função de covariância $C(s_i, s_u) = Cov[\epsilon(s_i), \epsilon(s_u)]$, $i, u = 1, 2, ..., n$ e que, para algumas funções conhecidas de s como $x_1(s), ..., x_p(s)$, a média do processo estocástico é, $\mu(s_i) = \sum_{\mu=1}^{p} x_\mu(s_i)\beta_\mu$, com $\beta_1, ..., \beta_p$ parâmetros desconhecidos a serem estimados.

De forma equivalente em notação matricial tem-se:

$$Z = X\beta + \epsilon,$$

em que, $E(\epsilon) = 0$, que a matriz de covariância Σ assume-se não singular, $\boldsymbol{X}$ tem colunas com posto completo e que $\boldsymbol{Z}$ segue uma distribuição gaussiana multivariada com média $\boldsymbol{X}\beta$ e matriz de covariância Σ, isto é, $\boldsymbol{Z} \sim \mathrm{N}_n(\boldsymbol{X}\beta, \Sigma)$.

Considerando-se em uma forma particular paramétrica para a matriz de covariância,

$$\Sigma = \varphi_1 I_n + \varphi_2 \mathbf{R},$$

em que, φ_1 - efeito pepita, ou erro de variância; φ_2 - contribuição, ou variância de dispersão; $\mathbf{R}$ - matriz que é função de φ_3, $\mathbf{R} = R(\varphi_3) = [(r_{iu})]$, matriz n x n simétrica com seus elementos da diagonal $r_{ii} = 1, i = 1, ..., n$, em que φ_3 é função do alcance do modelo, e I_n é a matriz identidade n x n.

A forma paramétrica da matriz de covariância, ocorre para vários processos isotrópicos.

O estudo da detecção de *outliers* e observações influentes é um passo importante na análises de dados georreferenciados. O método de influência local sugerido por Cook (1986) avalia o efeito simultâneo de observações sobre os estimadores de MV sem removê-lo do conjunto de dados.

Em diversas situações, pode-se observar um conjunto de dados com observações discrepantes que podem ser consideradas influentes, ou seja, que podem mudar algum tipo de decisão na construção de modelos geoestatísticos.

A influência da perturbação ω, nos estimadores MV do vetor de parâmetros Ω, pode ser avaliada pelo afastamento da verossimilhança, definido por:

$$LD(\omega) = 2(l(\hat{\theta}) - l(\hat{\theta_\omega}))$$

em que, $\hat{\theta}$ - estimador de máxima verossimilhança de θ do modelo postulado, $\hat{\theta_\omega}$ - estimador de máxima verossimilhança de θ do modelo perturbado, ω - um vetor de perturbações pertencente a um espaço de pertubações Ω.

Para a aplicação da método do influência local em modelo espacial linear Gaussiano, analisaram um conjunto de dados reais, observando o parâmetro θ , por meio do *software* R, pacote geoR (RIBEIRO JR DIGGLE, 2009). Uma amostra de 47 pontos foram coletados no ano-safra 2006/2007 em uma área comercial de 57 ha, na região oeste do Paraná. Foram analisados os dados da variável Produtividade da soja e como covariáveis á resistência do solo à penetração através de camadas $0 - 0,10m(x_2)$, $0,10 - 0,20m(x_3)$ e $0,20 - 0,30m(x_4)$ de profundidade. A técnica de amostragem utilizada foi a sistemática alinhada, com distância de 50 - 75 m entre pontos. utilizaram-se os modelos exponencial, gaussiano e família Matérn, com os métodos de estimação de parâmetros máxima verossimilhança (MV). Para escolher o modelo espacial que melhor se ajusta à semivariância, utilizou-se a técnica de validação cruzada. Para investigar a existência de pontos influentes, realizou-se uma análise de diagnóstico por meio da técnica de influência local. O estudo mostrou que a presença de valores atípicos entre dados amostrados pode exercer forte influência nos mapas temáticos, alterando, assim, a dependência espacial. A aplicação de técnicas de diagnóstico de influência local deve fazer parte de toda análise geoestatística, garantindo que as informações contidas nos mapas temáticos tenham maior confiabilidade.

Observa-se por meio deste trabalho que nem sempre um valor discrepante é influente, porém um valor qualquer, mesmo que não discrepante, pode ser considerado influente. Ou seja, um valor discrepante não deve ser eliminado da amostra antes de ser realizada uma análise de diagnóstico de influência local.