

RESENHA

Nome: João Vitor Teodoro

Disciplina: Geoestatística

REFERÊNCIA BIBLIOGRÁFICA

Camargo, E. C. G.; Druck, S.; Monteiro, A.M.V.; Freitas, C. C.; Câmara, G. **Mapeamento do risco de homicídio com base na co-krigeagem binomial e simulação: um estudo de caso para São Paulo, Brasil.** Cad. Saúde Pública, Rio de Janeiro, 24(7): 1493-1508, Julho, 2008.

SÍNTESE DO ARTIGO

Conforme informações do DATASUS (Departamento de informática do SUS), o homicídio é a principal causa de mortes violentas. Na cidade de São Paulo este fator é sentido pela população de forma crescente, porém, sua distribuição pela cidade não ocorre de forma homogênea.

A análise geoestatística dos dados de homicídio fornece informações para avaliação de risco de homicídio e sua distribuição espacial, possibilitando estabelecer estratégias de controle e prevenção.

Métodos como bayesianos empíricos ou da aplicação dos indicadores de Moran & Geary, consideram o fenômeno distribuído homogeneamente no interior da região e, variação apenas nas fronteiras. Já os procedimentos de simulação condicionada permitem delimitar níveis de risco probabilísticos espacialmente, melhorando a representação e entendimento.

Os métodos tradicionais geoestatísticos não consideram a heterocedasticidade, ou seja, que o valor da variância assume valores diferentes em cada lugar dependente do tamanho populacional.

Foi considerada uma região dividida em N áreas, em que, para cada área há o número de ocorrências $l(u_i)$, população em risco $n(u_i)$ e vetor de coordenadas espaciais $u_i = (x_i, y_i)$, que é a centróide da área em questão. Define-se a variável aleatória $Z(u_i)$ como taxa de homicídio e $z(u_i) = l(u_i)/n(u_i)$ como possíveis realizações de $Z(u_i)$. Supondo a existência da variável aleatória risco de homicídio $R(u)$, deseja-se estimar sua superfície.

Denotando $R(u_i)$ como o risco médio do centróide da área i e que o risco é a única fonte de correlação entre homicídios, então, a distribuição das taxas de ocorrência de homicídio é $Z(u_i)|R \sim \frac{1}{n(u_i)} Bi[r(u_i), n(u_i)]$.

O estimador proposto do semivariograma do risco é dado por:

$$\hat{\gamma}_{(u_i, u_j)}^{*R} = \hat{\gamma}_{(u_i, u_j)}^{*Z} - \frac{1}{2} [\hat{\mu}_w(1 - \hat{\mu}_w) - \hat{\sigma}_{R_w}^2] \left[\frac{n(u_i) + n(u_j)}{n(u_i)n(u_j)} \right], \text{ em que,}$$

$$\hat{\gamma}_{(u_i, u_j)}^{*Z} = \frac{1}{2 \sum_{i,j=1}^{M(u_i, u_j)} n(u_i)n(u_j)} \sum_{i,j=1}^{M(u_i, u_j)} \left\{ n(u_i)n(u_j) [z(u_i) - z(u_j)]^2 \right\}, \quad \text{é}$$

estimador de semivariograma empírico ponderado pela população. $M(u_i, u_j)$ é o número total de pares de pontos disponíveis para uma distância de análise (lag).

$$\hat{\mu}_w^* = \frac{1}{n_w} \sum_{i=1}^{n_w} z(u_i) \text{ e } \hat{\sigma}_{R_w}^2 = \sum_{i=1}^{n_w} \frac{[z(u_i) - \hat{\mu}_w^*]^2}{n_w - 1},$$

são escolhidos adequadamente dentre as W médias e variâncias zonais de risco, conforme as posições u_i e u_j , em que, n_w é o número de observações $z(u_i)$ em w .

A média dos pares de localização envolvidos no cálculo do vetor distância h é dada por $\frac{1}{M(h)} = \sum_{i=1}^{M(h)} \frac{n(u_i) + n(u_i+h)}{n(u_i)n(u_i+h)}$.

A co-krigeagem binomial é um interpolador linear usado para estimar o risco em uma localização qualquer na área, denotada por u_0 , por meio de uma combinação linear de k taxas vizinhas $z(u_i)$ dada por $\hat{R}(u_0) = \sum_{i=1}^k \lambda(u_i) z(u_i)$, para a centróide k .

O destaque no uso da co-krigeagem em relação aos demais métodos de interpolação, é a cálculo dos pesos considerando as covariâncias diretas entre $Z(u_i)$ e $Z(u_j)$, as covariâncias cruzadas entre $Z(u_i)$ e $R(u_0)$, a não tendenciosidade e variância mínima. Assim, a aplicação do método de co-krigeagem binomial gera uma superfície representando a média da distribuição espacial de risco.

A estrutura de correlação espacial do risco estimada influi diretamente na qualidade e na precisão dos mapas gerados pelo interpolador de co-krigeagem binomial, deste modo, antes da aplicação desta metodologia deve-se verificar se a estrutura estimada é adequada. Foi gerado um conjunto de mapas simulados de $Z(u_i)$, representando as possíveis realizações do que poderia ser a informação observada e, para cada mapa, o estimador foi aplicado, formando uma distribuição simulada para o semivariograma do risco e, a estrutura de correlação espacial do risco estimada é comparada à simulada.

Depois de selecionado o modelo de estrutura de correlação espacial, estimativas do risco de homicídios foram estabelecidas por co-krigeagem binomial. As superfícies geradas mostram a evolução da distribuição de homicídios em São Paulo de 2002 a 2004 e, verifica-se que algumas áreas concentram grande parte dos homicídios, enquanto em outras, as proporções são bem menores.

A região central da cidade é um foco de homicídios. Esta região contém grande aglomeração de pessoas durante o dia e atividade noturna intensa. Deslocando-se do centro em direção às periferias, o risco de homicídios vai diminuindo, até que envolvem distritos menos vulneráveis à criminalidade. Porém, surgem três áreas de alto risco, na periferia sul, leste e norte, que são áreas com grande densidade populacional, com ocupação desordenada e irregular, com falta de infra-estrutura básica e baixo grau de escolaridade.

PRINCIPAIS TESES DEFENDIDAS

Os procedimentos de simulação condicionada consideram a heterocedasticidade e permitem delimitar níveis de risco probabilísticos espacialmente, melhorando a representação e entendimento dos mapas. No caso abordado, esta metodologia é a mais adequada, possibilitando gerar cenários para o risco de homicídio com base na co-krigeagem binomial e de procedimentos de simulação estocástica

ANÁLISE CRÍTICA

O artigo expõe uma situação que exemplifica a importância da utilização de simulação condicionada sob o enfoque de um novo estimador de semivariograma. A citação de outras referências que utilizaram a mesma metodologia em outras áreas e obtiveram resultados similares deveria ocorrer. O mesmo deveria ser feito para casos em que a metodologia não se enquadra.

A abordagem e apresentação dos cenários otimistas e pessimistas em 2002, 2003 e 2004, viabilizam uma melhor interpretação dos resultados. Nos resultados, apenas foi citado que a explicação da distribuição é complexa envolvendo alguns fatores e que isto é tema de trabalhos futuros, porém, uma interpretação prévia poderia ser realizada com base em conhecimentos demográficos sob análise multivariada.