

Uma Abordagem de Modelos Estatísticos Aplicados a Geomedicina - Casos de Câncer de Fígado no Estado do Paraná

Gledson Luiz Picharski; Edicleia Furlanetto;
Paulo Justiniano Ribeiro Jr.; Silvia Emiko Shimakura
(LEG - Laboratório de Estatística e Geoinformação)

Resumo

Um dos principais objetivos da Geomedicina é estudar a relação entre estruturas de geologia, como elementos químicos coletados em águas superficiais, com a ocorrência de doenças. Neste contexto, este trabalho busca investigar a associação entre dados geoquímicos de águas superficiais em microbacias do Estado do Paraná coletados pela MINEROPAR, e o registro de casos de câncer de fígado nas microregiões do Paraná obtido do DATASUS no ano de 2004. Modelos estatísticos são apropriados para explicar estas relações considerando diversos fatores, dentre eles, a espacialidade das informações. Modelos lineares generalizados (GLM) e modelos aditivos generalizados (GAM) são duas estratégias de modelagem úteis para explicação destas relações. Para utilização destas abordagens faz-se necessário primeiramente um bom entendimento das variáveis. Isto pode ser feito através de uma análise exploratória, que pode sugerir possíveis transformações nas covariáveis e obter por exemplo mapas suavizados, indicando as regiões de maior ou menor incidência de casos. Através do sistema R pode-se obter o variograma dos elementos geoquímicos, fazer a krigagem e obter todos os resumos de informação e utilizar os modelos deste trabalho. O GAM mostrou a não significância do componente espacial para a análise por microregião, isso também foi percebido na razão de verossimilhança dos resíduos do GLM. O Manganês (MN) e o Nitrato (NO_3) foram as covariáveis mais significativas encontradas pelo GLM.

Introdução

Apesar de muitos esforços para cuidar da qualidade da água doce, são preocupantes os casos de doenças que surgem devido a contaminação das águas. Uma das fontes fundamentais para sobrevivência humana é a água, e ela está se tornando um bem escasso. Segundo Romão, a evolução humana tem criado um modelo de desenvolvimento industrial e urbano desordenado, com exploração inadequada dos recursos naturais e desequilíbrio ambiental. A deterioração da qualidade da água nos diferentes sistemas hídricos é preocupação crescente e com razão, os recursos de água doce constituem um componente essencial e indispensável à sobrevivência humana. A água potável, própria para consumo humano, segundo o artigo 40 da Portaria nº 518 de 25 de março de 2004 do Ministério da Saúde, precisa estar em conformidade com parâmetros

microbiológicos, físicos, químicos e radioativos que atendam ao padrão de portabilidade e não ofereça riscos à saúde.

Em se tratando de casos de doenças, alguns métodos de análise espacial para dados geoestatísticos são necessários para o objetivo do estudo desse trabalho, métodos estes que visam explicar pelo menos em parte os casos de câncer de fígado. Esses métodos, tais como modelos lineares generalizados, variogama, krigagem e modelos aditivos generalizados são ferramentas cabíveis para explicar o fenômeno em estudo.

Objetivo

Investigar associação entre os teores de elementos geoquímicos presentes nas águas superficiais à ocorrência de casos de câncer de fígado no Estado do Paraná. As variáveis medidas nas águas são dados geoestatísticos e a doença (casos de câncer de fígado) é mensurada por município. Deseja-se verificar a relação entre as características das águas com esta doença controlando por um possível efeito espacial residual.

Descrição do estudo

Os dados utilizados neste estudo foram disponibilizados pela MINEROPAR, os quais contém informações da qualidade das águas superficiais por microbacias do estado do Paraná. Foram medidos nas águas a quantidade de diversos elementos e características, entre eles Nitrato, Nitrito, Teor de Acidez, Zinco, Bromo, Sódio, Ferro, Alumínio, Fosfato, Magnésio, Manganês, Cálcio, Condutibilidade, etc. Os dados sobre a doença foram obtidos através do DATASUS.

Em uma análise exploratória percebe-se que os dados distinguem-se em três grupos, sendo eles: variáveis com clara espacialização; variáveis que talvez possam ser consideradas na análise mas talvez agrupadas ("presença", "ausência", ou mais grupos) pois apresentam grande número de valores repetidos; e variáveis aparentemente sem potencial para serem analisadas por apresentarem quase a totalidade dos valores iguais.

No primeiro grupo foi feita uma análise geoestatística, onde foram feitos resumos numéricos de cada covariável, percebendo as diversas medidas resumo dos dados. Histogramas, assim como gráficos de boxcox, auxiliaram a identificação da transformação necessária, a seguir verifica-se a variação espacial através dos variogramas e posteriormente é feita a krigagem, gerando os mapas das covariáveis distribuídas suavemente pela superfície do Paraná.

Para os demais grupos foram usadas variáveis "dummy" devido ao fato de algumas variáveis apresentarem alta incidência de valores iguais. Em sua maioria são censuras à esquerda que ocorrem por limites de detecção dos aparelhos. As técnicas para estimar dados com censura são muito mais apuradas e serão tratadas na seqüência do trabalho, contudo neste momento opta-se pela simplicidade das variáveis "dummy".

Todas as covariáveis foram medidas em 698 pontos do Estado do Paraná, e deseja-se obter a relação com a ocorrência do câncer de fígado que foi observada por município no ano de 2004. Como muitos dos municípios não possuem incidência da doença no ano, neste ponto do trabalho optou-se por tratar todas as informações por microrregião. Foi calculada a média dos valores de cada covariável para cada microrregião, assim busca-se agora verificar a significância do componente espacial e de cada uma das covariáveis no modelo.

Alguns modelos poderiam ser usados para esta situação, inicia-se o estudo com os modelos lineares generalizados (GLM), com ajuda do método stepwise e através do AIC foi possível

identificar as covariáveis mais significativas neste modelo. Por tratar-se de dados de contagens é adequado o uso de distribuições da família exponencial, em particular a distribuição de Poisson. Usando o resíduo padronizado de Pearson e os centróides das microrregiões, com o uso do pacote geoR, foi possível fazer uma análise da razão de verossimilhança que indicou que o modelo não espacial era mais significativo, de forma que pode-se considerar que as explicações espaciais por microrregião estariam contidas nas covariáveis. O modelo usado foi:

$$Y \sim \text{Poisson}(\mu)$$

$$g(\mu) = \log(MG) + MN + \log(NO3) + \log(CONDU) + \log(e)$$

Onde Y é o número de casos de câncer de fígado.

Buscando verificar possíveis suavizações nas covariáveis, ou ainda a consideração do componente espacial, construiu-se o modelo aditivo generalizado (GAM). A diferença em relação ao GLM é que no GAM foi usada uma função de suavização no espaço. Aqui o componente espacial é tratado como covariável, onde se conseguiu estimar um coeficiente e verificar o p-valor. Observou-se a não significância no componente espacial, assim como observado no teste da razão de verossimilhança do resíduo espacial analisado na geoR. O GAM utilizado foi:

$$Y \sim \text{Poisson}(\mu)$$

$$g(\mu) = \log(MG) + MN + \log(NO3) + \log(CONDU) + s(x, y) + \log(e)$$

Onde Y é a taxa de casos de câncer de fígado, $s(x,y)$ é o componente espacial do modelo e $\log(e)$ é o log do número esperado de casos sob a hipótese nula de risco constante nas microrregiões.

Resultados

Foi feita uma análise geoestatística nas covariáveis do grupo que possuía uma clara espacialização. Um dos elementos que passou por esta análise foi o Cálcio(Ca). Inicialmente utilizou-se o gráfico box-cox para verificar a necessidade de transformação da variável, percebe-se que mesmo estando um pouco fora do intervalo é possível usar transformação logarítmica. É feito um variograma com base na variável transformada e depois ajustado, neste caso usando distribuição exponencial. Observa-se espacialmente no estado do Paraná, pontos de localização com níveis diferentes de teor de Cálcio. Com o uso do variograma ajustado, foi feita a krigagem, podendo assim ver o mapa do cálcio no Estado do Paraná. Os 4 gráficos são representados na Figura 1. Após a utilização dos métodos, por duas técnicas relativamente distintas, GAM e análise espacial do resíduo do modelo Poisson, o componente espacial mostrou-se não significativo para microrregiões, indicando que para essa divisão de áreas tratadas não parece ser necessário usar modelos espaciais.

As suavizações de splines (componente suavizador) feitas para as covariáveis no GAM, não mostraram ser significativas para serem usadas, pois acrescentam em complexidade e não em informação para o modelo.

O modelo que melhor explica o câncer de fígado considerando as variáveis usadas é:

$$Y \sim \text{Poisson}(\mu)$$

$$g(\mu) = \log(MG) + MN + \log(NO3) + \log(CONDU) + \log(e)$$

A tabela 1 representa os coeficientes estimados para o modelo.

Dos elementos químicos encontrados, o Manganês, Magnésio e Nitrato foram os que melhor explicaram os casos de câncer de fígado observados nas microrregiões do Paraná.

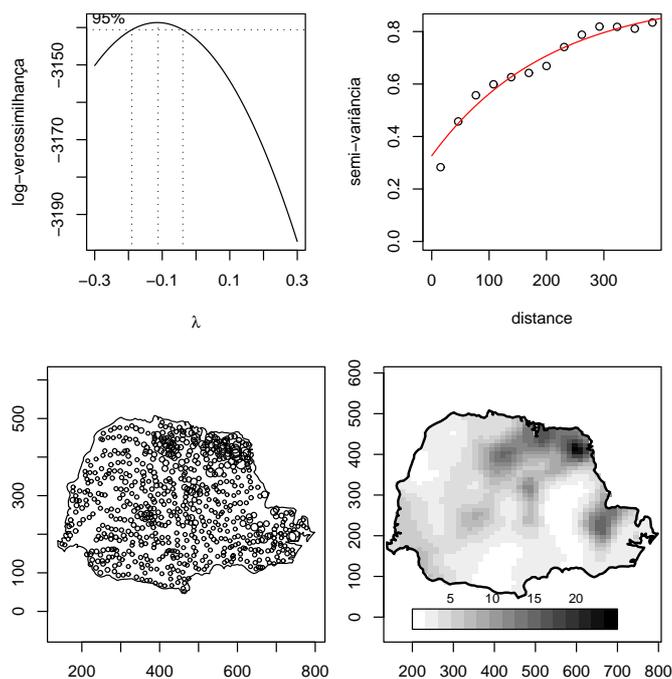


Figura 1: Análise geoestatística de Cálcio no Paraná

	Coefficiente	Erro padrão	p-valor
Intercepto	3.43971	1.89132	0.068960
log(MG)	0.81397	0.57861	0.159496
MN	-0.61753	0.16391	0.000165
log(NO3)	0.24080	0.08468	0.004461
log(CONDU)	-0.85928	0.58772	0.143724

Tabela 1: Coeficientes do GLM

Discussão

Para a análise de dados de geomedicina, assim como muitos outros, é necessário excluir alguns efeitos de confundimento. Para isto pode-se usar análise fatorial em conjunto com modelos estruturados, onde uma matriz de correlação pode ser o primeiro passo nesta direção. Este é um procedimento que será adotado posteriormente para melhor aproveitamento das informações das variáveis mais correlacionadas.

Outro procedimento é efetuar análise semelhantes por municípios, onde será investigada a significância no componente espacial e possivelmente algumas covariáveis se tornem significativas no modelo.

Agradecimentos

Agradecemos a equipe do Geomedicina do Instituto de Pesquisa Pelé Pequeno Príncipe pelas discussões e fornecimento dos dados, em especial Professor Bonald Cavalcante de Figueiredo, Humberto Ibañez e Otavio Augusto Boni Licht.