



Modelos de regressão espacial aplicados a dados epidemiológicos

Stéfane Lele Kanashiro (PIBIC/CNPq/FA/Uem), Diogo Francisco Rossoni (Orientador), e-mail: diogo.rossoni@gmail.com.

Universidade Estadual de Maringá /Centro de Ciências Exatas/Maringá, PR.

Ciências Exatas e da Terra / Probabilidade e Estatística.

Palavras-chave: SAR, CAR, autocorrelação.

Resumo:

Os dados epidemiológicos escolhidos foram sobre a incidência de casos de leishmaniose cutânea nos 399 municípios do estado do Paraná no ano de 2013 e serão ajustados em função da degradação da vegetação, considerado um fator importante para a proliferação da doença. Se tratando de uma doença endêmica, é possível verificar autocorrelação entre as áreas vizinhas, com um índice $I = 0.183$ (índice de Moran). Na presença de autocorrelação é necessário modelarmos a dependência espacial, que resulta em uma melhora na precisão geral do modelo. Os métodos escolhidos para a análise em questão foram o SAR (Simultaneous Autoregressive Model) com suas ramificações, e CAR (Conditional Autoregressive Model). O modelo escolhido através do critério de Akaike foi o CAR, sendo que não houve influência da degradação da vegetação nos casos de leishmaniose cutânea.

Introdução

A leishmaniose cutânea é uma doença infecciosa em que o parasita se multiplica dentro das células de defesa do indivíduo, causando feridas na pele, nariz, boca e garganta, transmitida por insetos que se alimentam de sangue, difíceis de detectar devido ao seu tamanho, além de serem capazes de atravessar telas de proteção das janelas. A devastação das matas faz com que os mamíferos que antes a habitavam, migrem para outra região, circunstância que deixa o mosquito sem suas fontes naturais de alimentação.





O fato de que os mosquitos saem para áreas próximas para se alimentarem, disseminando a doença, faz com que os dados das incidências sejam dependentes.

Um modelo de regressão mais adequado nos permitirá verificar com maior precisão a relação entre o desmatamento e as incidências, permitindo que atitudes adequadas sejam tomadas em relação a esse assunto.

Materiais e métodos

Os dados contém informações relacionadas à degradação da vegetação e incidência de leishmaniose no estado do Paraná durante o ano de 2013 e foi analisado no ambiente R (R DEVELOPMENT CORE TEAM, 2016).

Os modelos SAR em si analisam a autocorrelação de uma forma global pois aceitam vizinhos de ordem superiores a 1.

Considerando ρ como o índice de autocorrelação dos dados e λ como o índice de dependência espacial do erro, obtemos um modelo de regressão linear simples quando ambos são zero (ASSUNÇÃO; KRAINSKI, 2009).

Segundo Smith (2015), podemos definir os modelos como:

SAR lag-model: É utilizado quando a variável preditora é incluída no modelo de regressão linear simples, juntamente com o componente autoregressivo (ρ) e não existe dependência nos erros ($\lambda = 0$).

SAR error model: Nos casos em que há indícios de que os dados são correlacionados porém o SAR lag-model não obtém melhorias relevantes, podemos analisar se $\rho = 0$ porém os erros não são i.i.d ($\lambda \neq 0$).

CAR model: O modelo CAR verifica a dependência de primeira ordem ou autocorrelação espacial local e se utiliza das cadeias de Markov para a modelagem de dados desse tipo.

Para este caso, consideramos que a probabilidade dos valores que foram estimados em determinado local é condicionada sobre os valores dos vizinhos (BIVAND et al, 2008).

Resultados e Discussão

Na análise preliminar o teste de I de Moran obteve um índice de autocorrelação nos dados de $I = 0.183$, fazendo com que os modelos de regressão espacial fossem necessários. Além disso, foi detectada autocorrelação do tipo “alta-alta” em 14 municípios (Figura 1).





O modelo SAR lag-model (Tabela 1) não considerou a degradação da vegetação significativa para o modelo, além de detectar um $\rho = 0.383$, descartando a hipótese de que não existe dependência nos dados. O AIC obtido aqui foi de -3283.08.

O modelo SAR error model também produziu um valor-p alto em respeito à significância da degradação da vegetação. O valor apresentado para a dependência dos erros foi de $\lambda = 0.384$ também indicando a rejeição da hipótese de nulidade de autocorrelação. O AIC para este caso foi de -3283.04.

Já o modelo CAR chegou aos mesmos resultados em se tratando da significância da degradação, porém obteve $\lambda = 0.1191$ e $AIC = -3284.2$, sendo o modelo que melhor se adequou aos dados autocorrelacionados, resultando em um melhor ajuste.

Tabela 1 Parâmetros estimados dos modelos

SAR lag	estimated	p-value	SAR error	estimated	p-value
Intercept	5.9e-04	0.02	Intercept	9.2e-04	0.01
Degradac	-7.7e-06	0.64	Degradac	-8.9e-06	0.67
Rho	0.383	5.5e-08	Lambda	0.384	5.6e-08

CAR	estimated	p-value
Intercept	8.14e-04	0.03
Degradac	-6.0e-06	0.76
Lambda	0.119	3.0e-08

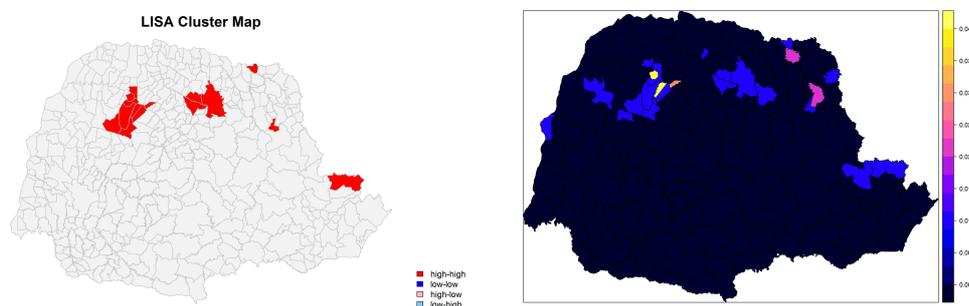


Figura 1 Mapa de clusters de autocorrelação e incidências, respectivamente.





Conclusões

Através dos mapas gerados, é possível verificar que alguns municípios da região Norte e da região Leste apresentaram correlação do tipo “alta-alta”. Nas demais regiões a ocorrência da variável dependente é considerada aleatória.

Além disso, verificou-se que a degradação da vegetação não influenciou na incidência de casos de leishmaniose cutânea, sendo que o melhor modelo ajustado foi o CAR.

Agradecimentos

CNPq - FA - UEM, Governo do Estado do Paraná.

Referências

ASSUNÇÃO, R; KRAINSKI, E. **Neighborhood Dependence in Bayesian Spatial Models**. Biometrical Journal, v. 51, n. 5, p. 851-869, 2009.

BIVAND, R. S.; PEBESMA, E. J.; G OMEZ-RUBIO, V. **Applied Spatial Data Analysis with R**. [s.l.] Springer, 2008. v. 747248717

R DEVELOPMENT CORE TEAM. **R: A language and environment for statistical computing**. Versão 3.3.0. Vienna: R Foundation for Statistical Computing, 2016. Disponível em: <<http://www.r-project.org>>. Acesso em: 22 jun. 2016.

SMITH, M. J. **Statistical Analysis Handbook – a web-based statistics resource**, The Winchelsea Press, 2015

