

# Seleção de covariáveis de vizinhança na descrição do padrão espacial da leprose dos citros

Elias Teixeira Krainski <sup>1</sup> Luziane Franciscan <sup>2</sup> Paulo Justiniano Ribeiro Jr <sup>3</sup>

## 1 Introdução

A leprose dos citros é uma doença que vem sendo estudada com objetivo de entender seu mecanismo de atuação. Em um experimento realizado pelo FUNDECITRUS (Fundo de Defesa da Citrucultura), a ocorrência de leprose em plantas de um talhão foram observadas em vários instantes de tempo. Neste caso, há dados descrevendo a evolução temporal da doença e também a dinâmica espacial dentro do talhão. Um objetivo é verificar o efeito que uma planta doente exerce sobre as outras. Esse efeito existe? Ocorre apenas em plantas vizinhas no espaço? No espaço e no tempo? Como é esse efeito?

O modelo autológico é uma alternativa para descrever a dinâmica espacial da incidência de doenças em um talhão, sendo um modelo para a probabilidade de doença de uma planta em função do *status* das plantas vizinhas. Essa informação é considerada como uma covariável do modelo, covariável de vizinhança. A covariável de vizinhança pode ser considerada de duas formas: a informação de existência ou não de vizinhas doentes ou, o número de vizinhas doentes. No primeiro caso, plantas com uma vizinha doente serão consideradas de forma igual às plantas com mais de uma vizinha doente. Essas duas possibilidades de considerar a informação das vizinhas devem ser consideradas na análise. Isso porque, ambas contemplam diferentes formas de contágio. No primeiro caso, o efeito de uma planta doente é o mesmo de duas plantas doentes.

Neste trabalho, foi utilizado um conjunto de dados de leprose dos citros para ilustrar o uso das duas formas de considerar a informação da vizinhança.

## 2 Material e métodos

O modelo autológico descreve a probabilidade de uma planta estar doente fazendo

$$\text{logit}(p_{ij}) = \beta_0 + \gamma_1(y_{i-1,j} + y_{i+1,j}) + \gamma_2(y_{i,j-1} + y_{i,j+1}), \quad (1)$$

---

<sup>1</sup>LEG/UFPR e UFMG, ekrainski@ufmg.br

<sup>2</sup>LEG/UFPR e ESALQ/USP, lfrancis@esalq.usp.br

<sup>3</sup>LEG/UFPR, paulojus@ufpr.br

em que  $p_{ij}$  é a probabilidade da planta na linha  $i$  e na coluna  $j$  estar doente;  $y_{i-1,j}$  e  $y_{i+1,j}$  são as vizinhas das linhas adjacentes, formando a covariável de vizinhança entre linha;  $y_{i,j-1}$  e  $y_{i,j+1}$  são as vizinhas das colunas adjacentes, formando a covariável de vizinhança dentro da linha;  $\gamma_1$  e  $\gamma_2$  são os parâmetros que medem o efeito das covariáveis de vizinhança. Esta estrutura é flexível para estudar o caso com doenças de plantas, onde o espaçamento entre linhas é diferente do espaçamento dentro das linhas.

Besag (1972) sugere que as estimativas de  $\{\gamma_1, \gamma_2\} = \gamma$  sejam obtidas maximizando a função de pseudo-verossimilhança, dada por

$$\tilde{L}(\gamma, y) = \prod_i \prod_j f(p_{ij}, y), \quad (2)$$

em que  $f(\cdot)$  é a densidade de uma distribuição de Bernoulli. Não há grandes problemas ao estimar os efeitos, mas as estimativas de variância dos efeitos são subestimadas.

Uma metodologia sugerida para estimar a variância das estimativas, é utilizar reamostragem. Consiste em gerar re-amostras dos dados originais e utilizá-las para estimar a quantidade de interesse. No caso de dados com estrutura de dependência espacial, ou seja, a configuração espacial dos dados, a reamostragem não é tão simples. Gumpertz, Graham & Ristano (1997) propõe que os erros sejam estimados por reamostragem, utilizando o algoritmo amostrador de Gibbs. A idéia é de retirar amostras da distribuição de cada observação  $y_{ij}$  condicionando ao *status* das vizinhas, usando (1).

As expressões para as duas formas de considerar a covariável de vizinhança estão resumidas na tabela 2, para contagem e presença/ausência de vizinhos doentes.

Tabela 1: Logito considerando contagem e ausência/presença de vizinhas doentes na linha

Contagem			Binária	
0	1	2	Ausência	Presença
$\beta_0$	$\beta_0 + \gamma_1$	$\beta_0 + 2\gamma_1$	$\beta_0$	$\beta_0 + \gamma_1$

A leprose dos citros é uma doença de relevante importância na citrucultura brasileira, pois compromete a produção das plantas infectadas e a vida útil das mesmas, gerando perdas de produção. Os dados analisados nesse trabalho foram cedidos pelo FUNDECI-TRUS. Trata-se de dados de incidência de leprose dos citros em um talhão de citros, no município de Santa Cruz do Rio Pardo/SP. O talhão tem espaçamento de 7,5 m  $\times$  3,8 m com 20 linhas de 58 plantas. Foram feitas 45 avaliações entre os períodos de janeiro de 2002 e novembro de 2004.

### 3 Resultados e discussão

O modelo autológico foi ajustado usando o pacote **Rcitrus** (Krainiski & Ribeiro Jr. 2005) implementado em R (R Development Core Team 2007). Para avaliar a forma mais adequada de considerar a informação das vizinhas, comparou-se as frequências esperadas com as observadas utilizando um teste qui-quadrado.

A Tabela 2 apresenta a distribuição de frequências das 1160 plantas na avaliação 31, sendo 71 plantas doentes. Destas, 20 não tinha nenhuma vizinha doente na mesma linha e 25 não tinha nenhuma vizinha na linha adjacente (coluna). A proporção de plantas doentes com nenhuma vizinha doente na linha é 0.019, 0.3377 se tiver uma vizinha e 0.8333 se tiver duas vizinhas doentes. Nota-se uma grande diferença entre as proporções de doentes com uma vizinha na linha e com duas vizinhas na linha. As probabilidades ajustadas pelos modelos considerando covariáveis de forma binária e contagem, também estão descritas na mesma Tabela.

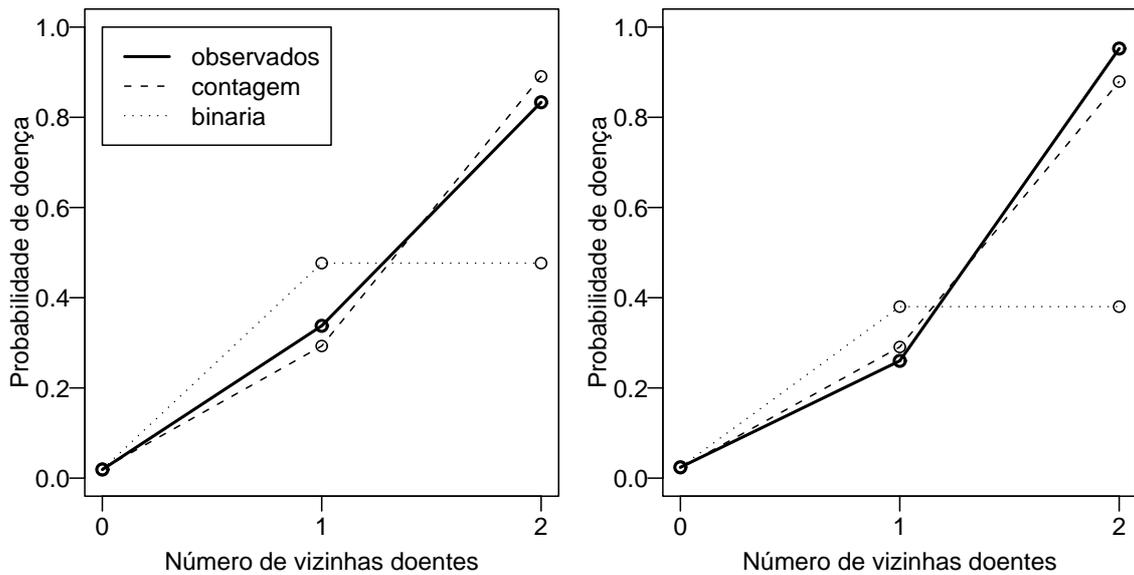
Tabela 2: Frequências observadas, probabilidades observadas e estimadas

	Covariáveis de Vizinha					
	Linha			Coluna		
	Nenhuma	Uma	Duas	Nenhuma	Uma	Duas
Plantas sadias	1033	51	5	1014	74	1
Plantas doentes	20	26	25	25	26	20
Prob. observadas	0.0190	0.3377	0.8333	0.0241	0.2600	0.9524
Cov. binária	0.0190	0.4766	0.4766	0.0241	0.3802	0.3802
Cov. contagem	0.0206	0.2930	0.8907	0.0226	0.2907	0.8791

O teste qui-quadrado para plantas vizinhas na linha, indicou como adequada a informação da covariável em forma de contagem com valor de p 0.8560, ou seja, o modelo se ajusta bem aos dados. Na forma binária o valor p do teste 0.0111 implica a não adequação desse modelo aos dados. Ao considerar vizinhas na coluna, o resultado é semelhante, com valor de p de 0.9068 no caso de contagem e 0.001 no caso binário.

A Figura 1 mostra as proporções observadas e as probabilidades estimadas pelos modelos, em função do número de vizinhas doentes. Comparando as retas, nota-se que das probabilidades estimadas pelo modelo considerando contagem de vizinhas doentes é mais parecida com a observada. Isso evidencia que o modelo mais adequado aos dados

Figura 1: Retas das proporções observadas e das probabilidades estimadas pelos modelos, em função do número de vizinhas doentes



de leprose de citros é quando considera-se a contagem de vizinhas doentes, tanto para covariáveis na linha como para covariáveis na coluna.

Os dois modelos foram ajustados considerando a covariável de vizinhança na linha para as avaliações 23 à 45. Os valores p do teste qui-quadrado aplicado para cada avaliação estão na Tabela 3. Nas primeiras avaliações, ambas as formas de modelagem são satisfatórias. Isso pode ser explicado pela baixa incidência de plantas doentes nessas avaliações, pois embora as proporções em função do número de doentes sejam muito diferentes, as frequências são baixas e fazem com que o teste qui-quadrado não seja significativo.

## 4 Conclusões

A forma de considerar a informação do *status* das vizinhas foi relevante na maioria das avaliações analisadas dos dados de leprose. Portanto, deve-se verificar a forma mais adequada de considerar a informação da vizinhança no modelo antes se fazer inferência.

## Agradecimentos

Ao FUNDECITRUS, na pessoa do Renato B. Bassanezi, e à Ana Beatriz C. Czermainski.

Tabela 3: Valores p do teste Qui-quadrado para avaliações 23 à 45 de leporse dos citros

Avaliação	Contagem	Binária	Avaliação	Contagem	Binária
23	0.9197	0.4288	35	0.9978	0.0000
24	0.5873	0.7797	36	0.5026	0.0008
25	0.2740	0.9563	37	0.9909	0.0000
26	0.3409	0.8513	38	0.9843	0.0000
27	0.1319	0.6801	39	0.9644	0.0000
28	0.3203	0.3825	40	0.8553	0.0000
29	0.3940	0.0825	41	0.9337	0.0000
30	0.5864	0.0433	42	0.8717	0.0000
31	0.8560	0.0111	43	0.8213	0.0000
32	0.9751	0.0023	44	0.7123	0.0000
33	0.9515	0.0028	45	0.1578	0.0010
34	0.9567	0.0009			

## Referências

- Besag, J. (1972). Nearest-neighbour systems and the auto-logistic model for binary data, *Journal of the Royal Statistics Society, Series B* **34**: 75–83.
- Gumpertz, M. L., Graham, J. M. & Ristano, J. B. (1997). Autologistic model of spatial pattern of phytophthora epidemic in bell pepper: Effects of soil variables on disease presence, *Journal of Agricultural, Biological and Environmental Statistics* **2**(2): 131–156.
- Krainski, E. & Ribeiro Jr., P. (2005). *Rcitrus: Funções em R para análise de dados de doenças de citros*. R package version 0.3-0.  
\*<http://www.leg.ufpr.br/Rcitrus>
- R Development Core Team (2007). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.  
\*<http://www.R-project.org>