

# Seleção De Covariáveis De Vizinhança Na Descrição Do Padrão Espacial Da Leprose Dos Citros

Elias Teixeira Krainski - mestrando UFMG  
Luziane Franciscan - EMBRAPA suínos e mestranda ESALQ  
Paulo Justiniano Ribeiro Junior - Prof. UFPR e Coord. LEG

24 de julho de 2007

Laboratório de Estatística e Geoinformação - UFPR.

<http://www.leg.ufpr.br> - {elias, luziane, paulojus}@leg.ufpr.br

# Resumo

# Resumo

- ▶ Várias avaliações no tempo de Leprose em um talhão dos citros.

# Resumo

- ▶ Várias avaliações no tempo de Leprose em um talhão dos citros.
- ▶ Aplicação do modelo autológico a dados de Leprose.

# Resumo

- ▶ Várias avaliações no tempo de Leprose em um talhão dos citros.
- ▶ Aplicação do modelo autologístico a dados de Leprose.
- ▶ Duas formas de considerar informação de vizinhança.

# Resumo

- ▶ Várias avaliações no tempo de Leprose em um talhão dos citros.
- ▶ Aplicação do modelo autológico a dados de Leprose.
- ▶ Duas formas de considerar informação de vizinhança.
  - ▶ Existência ou não de vizinhas doentes,
  - ▶ Número de vizinhas doentes.

# Resumo

- ▶ Várias avaliações no tempo de Leprose em um talhão dos citros.
- ▶ Aplicação do modelo autologístico a dados de Leprose.
- ▶ Duas formas de considerar informação de vizinhança.
  - ▶ Existência ou não de vizinhas doentes,
  - ▶ Número de vizinhas doentes.
- ▶ Reamostragem via *Gibbs Sampler* para testar significância da covariável.

# Resumo

- ▶ Várias avaliações no tempo de Leprose em um talhão dos citros.
- ▶ Aplicação do modelo autologístico a dados de Leprose.
- ▶ Duas formas de considerar informação de vizinhança.
  - ▶ Existência ou não de vizinhas doentes,
  - ▶ Número de vizinhas doentes.
- ▶ Reamostragem via *Gibbs Sampler* para testar significância da covariável.
- ▶ Teste Qui-quadrado para comparar ajustes.

# Leprose dos citros

## Leprose dos citros

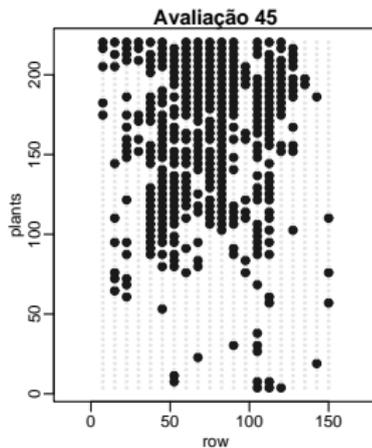
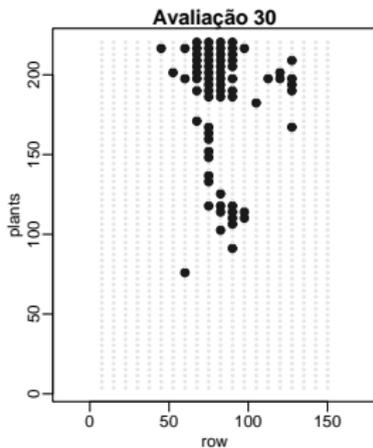
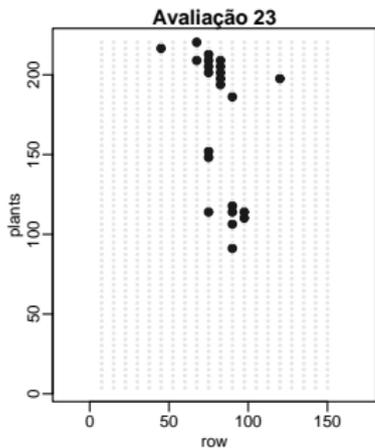
- ▶ Incidência de leprose dos citros num talhão em Santa Cruz do Rio Pardo - SP.

## Leprose dos citros

- ▶ Incidência de leprose dos citros num talhão em Santa Cruz do Rio Pardo - SP.
- ▶ 20 linhas com 58 plantas e espaçamento de  $7.5 \times 3.8$  metros.

# Leprose dos citros

- ▶ Incidência de leprose dos citros num talhão em Santa Cruz do Rio Pardo - SP.
- ▶ 20 linhas com 58 plantas e espaçamento de  $7.5 \times 3.8$  metros.
- ▶ 45 avaliações entre janeiro de 2002 e novembro de 2004.



# Objetivo

Verificar o efeito que uma planta doente exerce sobre as outras.

# Objetivo

Verificar o efeito que uma planta doente exerce sobre as outras.

- ▶ Esse efeito existe?

# Objetivo

Verificar o efeito que uma planta doente exerce sobre as outras.

- ▶ Esse efeito existe?
- ▶ Como é esse efeito?

## Modelo autológico (BESAG 1972)

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

## Modelo autológico (BESAG 1972)

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

- ▶  $p_{ij}$  é a probabilidade da planta na linha  $i$  e na coluna  $j$  estar doente;

## Modelo autológico (BESAG 1972)

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

- ▶  $p_{ij}$  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;

## Modelo autológico (BESAG 1972)

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

- ▶  $p_{ij}$  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;

## Modelo autológico (BESAG 1972)

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

- ▶  $p_{ij}$  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.

## Estimação (BESAG 1975)

Utilizar uma pseudo-verossimilhança:

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

## Estimação (BESAG 1975)

Utilizar uma pseudo-verossimilhança:

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

$f(\cdot)$  é a densidade de uma distribuição de Bernoulli.

## Estimação (BESAG 1975)

Utilizar uma pseudo-verossimilhança:

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

$f(\cdot)$  é a densidade de uma distribuição de Bernoulli.

- ▶ **Problema:** Estimativas de variância dos efeitos são subestimadas.

## Estimação (BESAG 1975)

Utilizar uma pseudo-verossimilhança:

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

$f(\cdot)$  é a densidade de uma distribuição de Bernoulli.

- ▶ **Problema:** Estimativas de variância dos efeitos são subestimadas.
- ▶ **Uma solução:** Reamostragem. Mas como fazer, pois os dados são espacialmente estruturados?

# Estimação (BESAG 1975)

Utilizar uma pseudo-verossimilhança:

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

$f(\cdot)$  é a densidade de uma distribuição de Bernoulli.

- ▶ **Problema:** Estimativas de variância dos efeitos são subestimadas.
- ▶ **Uma solução:** Reamostragem. Mas como fazer, pois os dados são espacialmente estruturados?
- ▶ **Reamostragem via amostrador de Gibbs**(GUMPERTZ & RISTAINO 1997) : Simular  $y_{ij}$  condicionando ao *status* das vizinhas, usando (1) com  $\gamma_1$  e  $\gamma_2$  estimados para os dados observados.

# Dois modelos

# Dois modelos

- ▶ **Modelo 1** Número de vizinhas doentes (**contagem**),

## Dois modelos

- ▶ **Modelo 1** Número de vizinhas doentes (**contagem**),
- ▶ **Modelo 2** Existência ou não de vizinhas doentes (**binária**).

# Dois modelos

- ▶ **Modelo 1** Número de vizinhas doentes (**contagem**),
- ▶ **Modelo 2** Existência ou não de vizinhas doentes (**binária**).

## Logito considerando vizinhas doentes na linha

Binária		Contagem Vizinhas Doentes		
Zero	Uma	Duas	Ausência	Presença
$\beta_0$	$\beta_0 + \gamma_1$	$\beta_0$	$\beta_0 + \gamma_1$	$\beta_0 + 2\gamma_1$

## Dois modelos

- ▶ **Modelo 1** Número de vizinhas doentes (**contagem**),
- ▶ **Modelo 2** Existência ou não de vizinhas doentes (**binária**).

### Logito considerando vizinhas doentes na linha

Binária		Contagem Vizinhas Doentes		
Zero	Uma	Duas	Ausência	Presença
$\beta_0$	$\beta_0 + \gamma_1$	$\beta_0$	$\beta_0 + \gamma_1$	$\beta_0 + 2\gamma_1$

- ▶ *Software*: Pacote Rcitrus (Krainiski & Ribeiro Junior 2005)
- ▶ Implementado em **R** (R Development Core Team 2007).

## Detalhes - Avaliação 31

### Freqüências observadas por número de viz. doentes

	Covariáveis de Vizinhança					
	Linha			Coluna		
	Zero	Uma	Duas	Zero	Uma	Duas
Plantas sadias	1033	51	5	1014	74	1
Plantas doentes	20	26	25	25	26	20

## Detalhes - Avaliação 31

### Freqüências observadas por número de viz. doentes

	Covariáveis de Vizinhaça					
	Linha			Coluna		
	Zero	Uma	Duas	Zero	Uma	Duas
Plantas sadias	1033	51	5	1014	74	1
Plantas doentes	20	26	25	25	26	20

- ▶ 1160 plantas na avaliação 31, sendo 71 plantas doentes. Destas, 20 não tinham vizinhas doentes na mesma linha e 25 não tinham vizinhas na linha adjacente (coluna).

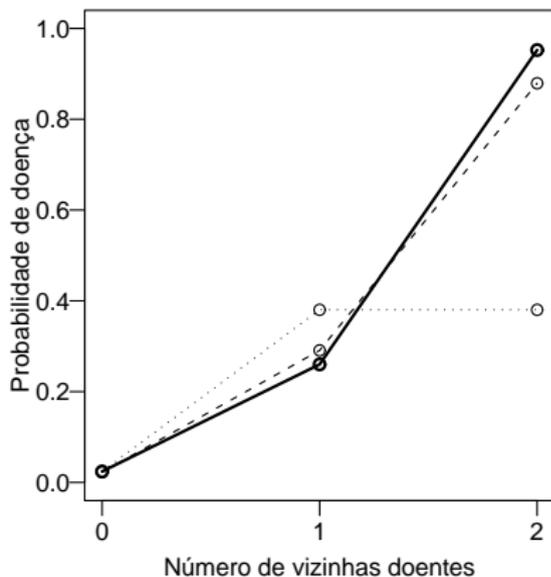
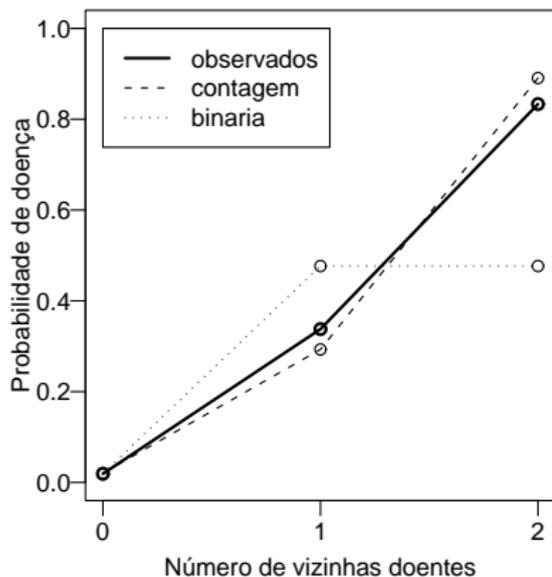
## Probabilidades estimadas

### Probabilidades observadas e estimadas por número de vizinhas doentes observadas

	Covariáveis de Vizinhança					
	Linha			Coluna		
	Zero	Uma	Duas	Zero	Uma	Duas
Prob. observadas	0.02	0.34	0.83	0.02	0.26	0.95
Cov. binária	0.02	0.48	0.48	0.02	0.38	0.38
Cov. contagem	0.02	0.29	0.89	0.02	0.29	0.88

- ▶ A proporção de plantas doentes com nenhuma vizinha doente na linha é 0.02, 0.34 se tiver uma vizinha e 0.83 se tiver duas vizinhas doentes.

## Proporções observadas e probabilidades estimadas pelos modelos, em função do número de vizinhas doentes



As probabilidades estimadas pelo modelo dado o número de vizinhas doentes observado tem comportamento mais parecido com as probabilidades observadas.

## Valores p do teste qui-quadrado

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			

- ▶ Diferença significativa: modelo mal ajustado.
- ▶ Ambos se ajustam bem nas primeiras avaliações.
- ▶ Considerar contagem é mais adequado.

## Valores p para covariáveis

99 reamostras para estimar erro padrão

### Valores p do teste de significância das covariáveis de contagem de vizinhos doentes

Avaliação	$\gamma_1$	$\gamma_2$	Avaliação	$\gamma_1$	$\gamma_2$
23	0.0000	0.4280	35	0.0000	0.0007
24	0.0000	0.9246	36	0.0000	0.0053
25	0.0000	0.6243	37	0.0000	0.0000
26	0.0000	0.6560	38	0.0000	0.0020
27	0.0000	0.8548	39	0.0000	0.0005
28	0.0000	0.2705	40	0.0000	0.0011
29	0.0000	0.0024	41	0.0000	0.0026
30	0.0020	0.0139	42	0.0000	0.0001
31	0.0000	0.0067	43	0.0000	0.0001
32	0.0000	0.0205	44	0.0000	0.0000
33	0.0000	0.0074	45	0.0000	0.0001
34	0.0000	0.0095			

# Conclusões

- ▶ A informação do *status* das vizinhas foi relevante;

# Conclusões

- ▶ A informação do *status* das vizinhas foi relevante;
- ▶ A forma de considerá-la também;

# Conclusões

- ▶ A informação do *status* das vizinhas foi relevante;
- ▶ A forma de considerá-la também;
- ▶ Nas primeiras avaliações, ambas são satisfatórias,

# Conclusões

- ▶ A informação do *status* das vizinhas foi relevante;
- ▶ A forma de considerá-la também;
- ▶ Nas primeiras avaliações, ambas são satisfatórias,
- ▶ Até a 28<sup>a</sup> avaliação, a vizinhança entre linha não é significativa.

# Agradecimentos

- ▶ Ao FUNDECITRUS, na pessoa do Renato B. Bassanezi,
- ▶ à Ana Beatriz C. Czermainski - EMBRAPA e doutora pela ESALQ.