Modelo autologístico:
Uma aplicação a dados de morte súbita dos citros

Elias Teixeira Krainski - mestrando UFMG
Luziane Franciscon - EMBRAPA e mestranda ESALQ
Paulo Justiniano Ribeiro Junior - Prof. UFPR e Coord. LEG
Renato Beozzo Bassanezi - FUNDECITRUS

25 de julho de 2007

Laboratório de Estatística e Geoirformação - UFPR.
`http://www.leg.ufpr.br`
{elias, luziane, paulojus}@leg.ufpr.br

Motivação

- *Morte Súbita dos Citrus*
  - Rápido progresso da doença e morte da planta
  - Rápido avanço sobre área citrícola

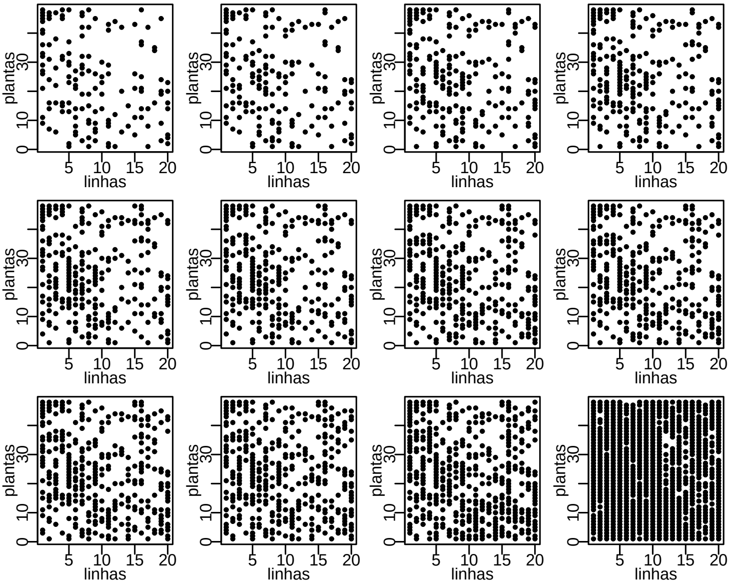
Motivação

- *Morte Súbita dos Citrus*
  - Rápido progresso da doença e morte da planta
  - Rápido avanço sobre área citrícola
- Análise
  - Disposição espacial dos dados modelo autoregressivo
  - Dados binários: modelo autologístico (BESAG 1972)

Motivação

- *Morte Súbita dos Citrus*
  - Rápido progresso da doença e morte da planta
  - Rápido avanço sobre área citrícola
- Análise
  - Disposição espacial dos dados modelo autoregressivo
  - Dados binários: modelo autologístico (BESAG 1972)
- Dados analisados
  - 20 linhas com 48 plantas por linha,
  - Espaçamento 4×7.5 metros,
  - 11 avaliações feitas de 05/11/2001 até 07/10/2002.

Dados de *Morte Súbita dos Citrus*



Objetivos

- Influência da vizinhança na probabilidade de doença

Objetivos

- Influência da vizinhança na probabilidade de doença
- Influência temporal

Objetivos

- Influência da vizinhança na probabilidade de doença
- Influência temporal
  - Vizinhança na mesma avaliação,
  - Vizinhança na avaliação anterior,
  - Vizinhança em ambos os casos.

Modelo autologístico (BESAG 1972)

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

Modelo autologístico (BESAG 1972)

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

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

Modelo autologístico (BESAG 1972)

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

- 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 autologístico (BESAG 1972)

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

- 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 autologístico (BESAG 1972)

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

- 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;
- γ_1 e γ_2 são os parâmetros que medem o efeito das covariáveis de vizinhança.

Estimação (BESAG 1975)

Utilizar uma pseudo-verossimilhança:

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

Estimação (BESAG 1975)

Utilizar uma pseudo-verossimilhança:

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

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

Estimação (BESAG 1975)

Utilizar uma pseudo-verossimilhança:

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

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

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

Estimação (BESAG 1975)

Utilizar uma pseudo-verossimilhança:

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

$f(.)$ 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:

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

$f(.)$ 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 γ_1 e γ_2 estimados para os dados observados.

Bootstrap via *Gibbs Sampler*

- O algoritmo amostrador de Gibbs é utilizado para preservar o padrão espacial

Bootstrap via *Gibbs Sampler*

- O algoritmo amostrador de Gibbs é utilizado para preservar o padrão espacial
- Gerar B amostras e estimar a quantidade de interesse,

Bootstrap via *Gibbs Sampler*

- O algoritmo amostrador de Gibbs é utilizado para preservar o padrão espacial
- Gerar B amostras e estimar a quantidade de interesse,
- Passos:
  1. Simular uma nova *lattice*, uma planta por vez:

Bootstrap via *Gibbs Sampler*

- O algoritmo amostrador de Gibbs é utilizado para preservar o padrão espacial
- Gerar B amostras e estimar a quantidade de interesse,
- Passos:
  1. Simular uma nova *lattice*, uma planta por vez:
     - Usar $\hat{\gamma}$ estimado nos dados observados,
     - Condicionar no *status* atual das vizinhas,
     - Calcular a probabilidade de doença,
     - Simular $\{0,1\}$ com a probabilidade estimada,
     - Ir para próxima planta, aleatoriamente.

Bootstrap via *Gibbs Sampler*

- O algoritmo amostrador de Gibbs é utilizado para preservar o padrão espacial
- Gerar B amostras e estimar a quantidade de interesse,
- Passos:
  1. Simular uma nova *lattice*, uma planta por vez:
     - Usar $\hat{\gamma}$ estimado nos dados observados,
     - Condicionar no *status* atual das vizinhas,
     - Calcular a probabilidade de doença,
     - Simular $\{0,1\}$ com a probabilidade estimada,
     - Ir para próxima planta, aleatoriamente.
  2. Obter a estimativa de γ em cada *lattice* simulado,
  3. Repetir B vezes.

Modelos propostos

- Modelo $m1$:

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

Modelos propostos

- Modelo $m1$:

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

- Modelo $m2$:

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

Modelos propostos

- Modelo $m1$:

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

- Modelo $m2$:

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

- Modelo $m3$:

$$\begin{aligned} logit(p_{ij}^t) = \; & \beta_0 + \gamma_1(y_{i-1,j}^{t-1} + y_{i+1,j}^{t-1}) + \gamma_2(y_{i,j-1}^{t-1} + y_{i,j+1}^{t-1}) + \\ & \gamma_3(y_{i-1,j}^t + y_{i+1,j}^t) + \gamma_4(y_{i,j-1}^t + y_{i,j+1}^t) \; . \end{aligned}$$

Resultados

	Modelo 1		Modelo 2		Modelo 3			
					Temp Atual		Temp Ant.	
Av-Incid	$\hat{\gamma}_1$	val p	$\hat{\gamma}_1$	val p	$\hat{\gamma}_1$	val p	$\hat{\gamma}_1$	val p
1-0.15	0.392	.045						
2-0.17	0.642	.002	0.358	.014	0.417	.046	-0.034	.435
3-0.22	0.707	.000	0.441	.001	1.027	.000	-0.506	.004
4-0.24	0.609	.000	0.639	.000	0.916	.000	-0.239	.060
5-0.26	0.654	.000	0.618	.000	0.390	.016	0.243	.024
6-0.28	0.626	.000	0.595	.000	0.887	.000	-0.245	.031
7-0.32	0.640	.000	0.582	.000	0.544	.001	0.097	.196
8-0.33	0.614	.000	0.611	.000	0.573	.000	0.070	.259
9-0.34	0.473	.000	0.608	.000	0.152	.167	0.472	.000
10-0.38	0.539	.000	0.503	.000	0.064	.334	0.444	.000
11-0.46	0.694	.000	0.436	.000	0.637	.000	-0.120	.118

AIC

- $m1$: Avaliações (2,4,5,6,7,8 e 11);
- $m2$: Avaliações 9 e 10,
- $m3$: Avaliação 3.

	Modelo 1	Modelo 2	Modelo 3
2	**725.55**	726.76	727.54
3	813.25	824.66	**812.33**
4	**851.58**	858.66	853.08
5	**908.32**	909.09	909.81
6	**932.52**	936.61	934.17
7	**992.94**	997.26	994.80
8	**1003.70**	1004.79	1005.68
9	1019.30	**1018.58**	1020.50
10	1067.11	**1064.87**	1066.82
11	**1109.49**	1121.87	1111.08

Conclusões da análise

- A dependência espacial é de curto alcance, pois o espaçamento dentro da linha é de 4 metros e entre linhas é de 7,5 metros,
- A covariável de vizinhança no tempo anterior foi significativa,
- Há colinearidade quando incluimos covariável no tempo anterio e atual.

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.

BESAG, J. (1975). Statistical analisys of non-lattice data, *The Statistican* **24**: 179–195.

GUMPERTZ, M. L. ; GRAHAM, J. M. & RISTAINO, 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.