

Modelo Autológico Aplicado a Dados de Citrus

Luziane FRANCISCON

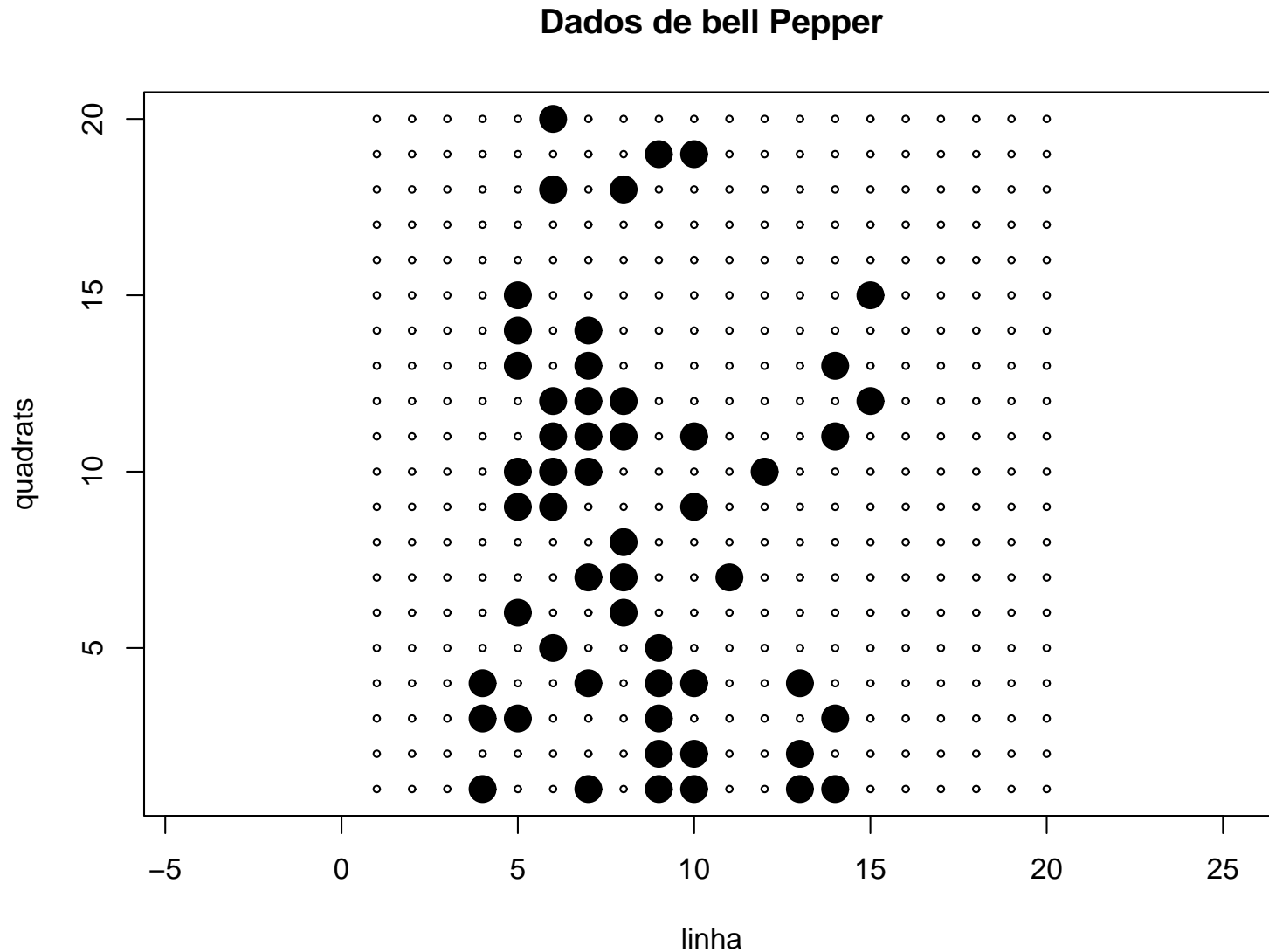
Orientador: Paulo Justiniano RIBEIRO Jr.

`luziane@ufpr.br`

Departamento de Estatística - UFPR

Dados de pimentas do sino

De GUMPertz M. L. ; GRAHAM e RISTAINO (1997)



Objetivos

- Aplicar o modelo autologístico a dados de citrus
- Estudar métodos de estimação e inferência
- Implementação do modelo no pacote Rcitrus

Idéias Iniciais

- O modelo espacial para dados binários - Modelo Autologístico BESAG (1972)
- Modelo logístico com covariadas de vizinhança
- A expressão é dada por:

$$P(Y = y_i | x_{ji}, y_{(i)}) = \text{logit}(y_i) = \sum_{j=1}^p \beta_j x_{ji} + \sum_{t=1}^q \gamma_t y_{(i)} .$$

- Os parâmetros γ captam a dependência espacial
- A probabilidade de doença depende de covariadas e do status das vizinhas

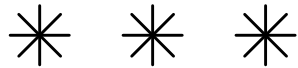
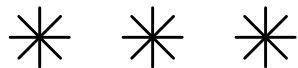
Estrutura de Vizinhaça

Deferentes possibilidades

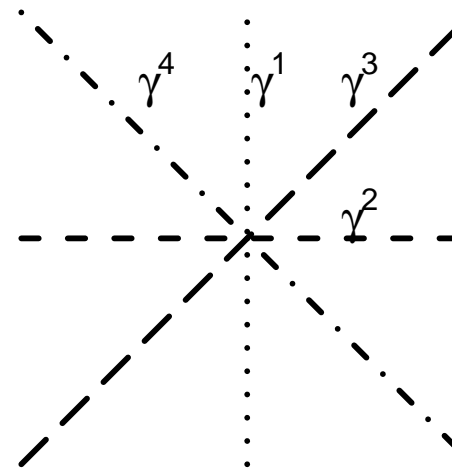
Primeira Ordem



Segunda Ordem



Segunda Ordem Separadamente



Inferência para o Modelo Autologístico

- As observações não são independentes
- O GLM usual não leva em conta a autocorrelação
- O GLM então é chamado de pseudo-verossimilhança
- É preciso um procedimento para estimar adequadamente os erros-padrão das estimativas
- Será utilizado o método bootstrap

Método de Bootstrap

- Consiste em gerar amostras dos dados para estimar a quantidade de interesse
- Não é possível o procedimento usual devido à configuração espacial
- Os passos:
 1. Simular N *lattices* usando a expressão do modelo autológico e as estimativas dos parâmetros para dados observados
 2. Obter a estimativa de γ em cada *lattice* simulado.
- Cada observação dentro do *lattice* deve estar condicionada aos seus vizinhos
- O amostrador de Gibbs foi utilizado para preservar o padrão espacial

Amostrador de Gibbs

O algoritmo utilizado para obter N amostras $\hat{\gamma}$ é:

1. Faz-se um ciclo pelas plantas atualizando o status de cada planta em ordem aleatória;
2. Para atualizar cada planta, simular o status (doente/sadia) de acordo com o modelo ajustado aos dados originais e o estado atual dos vizinhos;
3. Completa-se um ciclo ao atualizar o status de todas as plantas;
4. Repetir os passos anteriores N vezes.

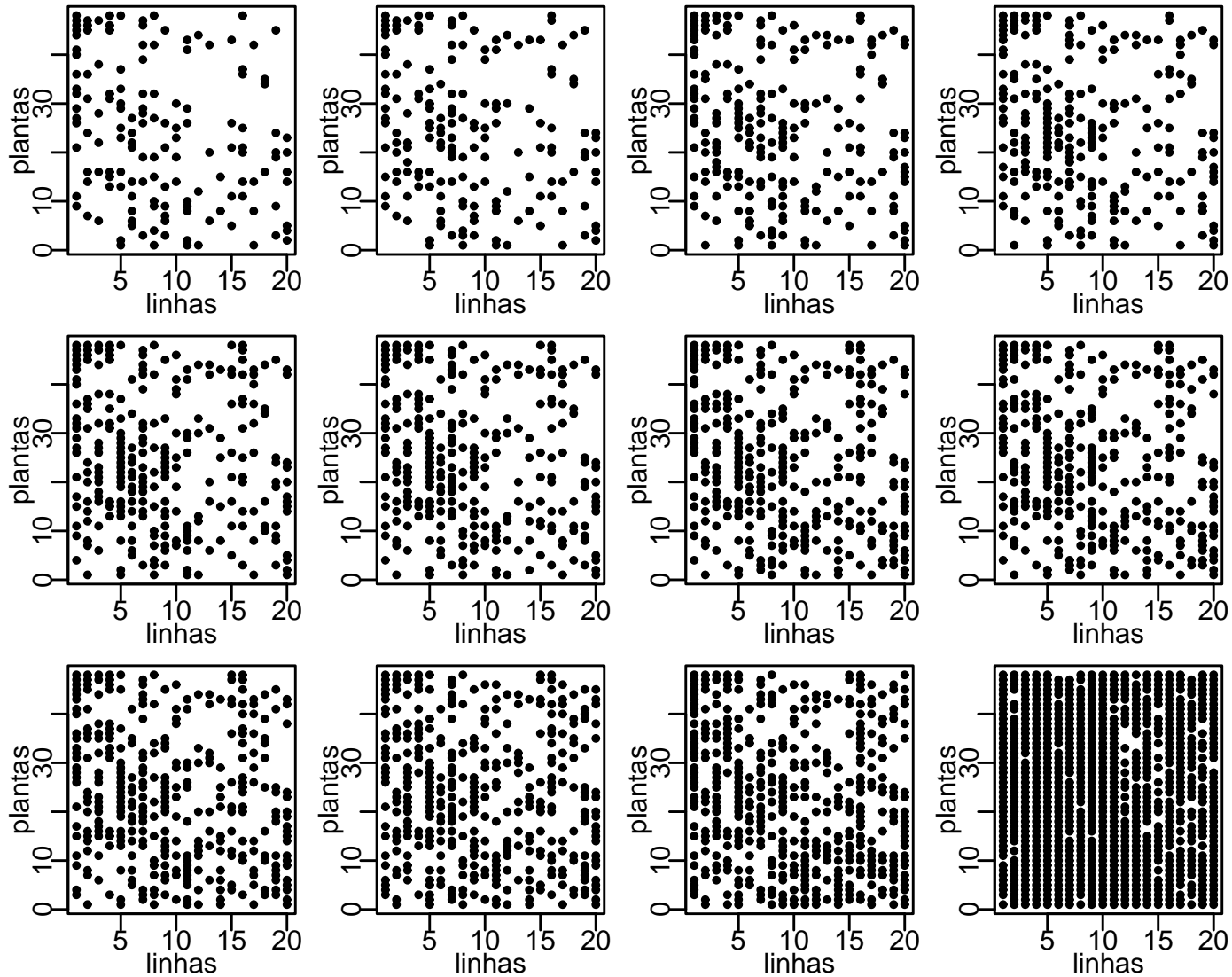
Alguns detalhes

- Descartou-se os m primeiros *lattices*;
- Ajustar o modelo para cada um dos $N - m$ *lattices*;
- Os $N - m$ vetores de $\hat{\gamma}$ são utilizados para obter o erro-padrão para $\hat{\gamma}$ estimado com os dados observados.

Dados de Morte Súbita em Citrus

- A gravidade da doença na citricultura brasileira.
- Dados de um talhão da fazenda Vale Verde, MG.
- O talhão com 20 linhas com 48 plantas por linha, com espaçamento $4 \times 7,5$ metros.
- Resposta: presença ou ausência da doença nas plantas.
- Analisadas 11 avaliações feitas no talhão.

Dados de Morte Súbita dos Citrus



Modelo proposto

- Ajustou-se o modelo autologístico para estes dados.
- O modelo proposto,

$$\text{Modelo : } \text{logit}(p_{kl}) : \beta_0 + \gamma_1 L_{kl} + \gamma_2 C_{kl} + \gamma_3 D_{kl1} + \gamma_4 D_{kl2} .$$

Estimativas dos parâmetros

Estimativas dos parâmetros por pseudo-verossimilhança

Avaliação	Incidência	$\hat{\beta}_0$	Dentro Linha	Entre Linhas	Diagonal (1,1)	Diagonal (-1,1)
01	0.14895	-2.02052	0.32079	-0.02221	0.00699	0.20609
02	0.17293	-1.97306	0.34912	0.22879	0.13854	0.16093
03	0.21875	-1.84436	0.62823	-0.02529	0.16258	0.23295
04	0.23840	-1.78096	0.70992	-0.09897	0.21175	0.20843
05	0.26354	-1.68169	0.58892	-0.02604	0.30987	0.16706
06	0.27812	-1.63307	0.63199	-0.00680	0.18708	0.23912
07	0.32292	-1.45117	0.60624	0.06947	0.09455	0.19067
08	0.33125	-1.39161	0.62401	0.13288	0.02720	0.13081
09	0.34167	-1.28953	0.60778	0.07711	-0.05904	0.18728
10	0.37500	-0.90676	0.47809	0.01132	-0.11679	0.07101
11	0.45729	-0.90008	0.52397	0.12469	-0.07815	0.16272

Estimativa dos erros-padrão

Avaliação	$\hat{EP}(\beta_0)$	$\hat{EP}(\gamma_1)$	$\hat{EP}(\gamma_2)$	$\hat{EP}(\gamma_3)$	$\hat{EP}(\gamma_4)$
01	0.19422	0.29304	0.29814	0.30349	0.28006
02	0.18722	0.25284	0.25847	0.25426	0.25613
03	0.18115	0.20525	0.22291	0.21430	0.20696
04	0.18349	0.19918	0.21155	0.19625	0.19798
05	0.18163	0.18264	0.19417	0.17766	0.18960
06	0.18481	0.17240	0.19332	0.18756	0.17759
07	0.18510	0.16213	0.17422	0.16269	0.16444
08	0.20016	0.16374	0.17241	0.17010	0.16239
09	0.19248	0.15058	0.17022	0.15719	0.16110
10	0.21185	0.15399	0.15652	0.15091	0.15291
11	0.22392	0.14430	0.14845	0.14823	0.14282

P-valores

P-valores obtidos utilizando os erros-padrão obtidos por bootstrap

Avaliação	$\hat{\beta}_0$	Dentro da Linha	Entre as Linhas	Diagonal (1,1)	Diagonal (-1,1)
01	0.00000	0.27365	0.94062	0.98163	0.46181
02	0.00000	0.16735	0.37606	0.58584	0.52980
03	0.00000	0.00221	0.90968	0.44806	0.26035
04	0.00000	0.00036	0.63991	0.28060	0.29245
05	0.00000	0.00126	0.89330	0.08113	0.37826
06	0.00000	0.00025	0.97196	0.31854	0.17815
07	0.00000	0.00018	0.69008	0.56112	0.24626
08	0.00000	0.00014	0.44090	0.87295	0.42052
09	0.00000	0.00005	0.65054	0.70722	0.24503
10	0.00002	0.00190	0.94235	0.43899	0.64234
11	0.00006	0.00028	0.40094	0.59804	0.25456

Conclusões da análise

- Existe dependência espacial somente dentro da linha a partir da terceira avaliação.
- 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.
- Quando há baixa incidência não existe dependência espacial.
- Através do modelo ajustado podemos calcular a chance de determinada planta estar doente condicionada ao status da doença das plantas vizinhas dentro da linha.
- As funções estão incluídas no pacote **Rcitrus**:
<http://www.est.ufpr.br/Rcitrus>

Bibliografia

Referências

BESAG, J. Nearest-neighbour systems and the auto-logistic model for binary data. *Journal of the Royal Statistics Society, Series B*, 1972.

GUMPERTZ M. L. ; GRAHAM, J. M.; RISTAINO, J. B. 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*, 1997.

Agradecimentos

- Ao Elias T. Krainski pela ajuda e incentivos
- A FUNDECITRUS por ter disponibilizado os dados de *Morte Súbita dos Citrus*.
- Ao Laboratório de Estatística e Geoinformação (LEG) pelo espaço físico e recursos computacionais.