

Padrões Espaciais da Morte Súbita dos Citros

Elias Teixeira Krainski, Paulo Justiniano Ribeiro Júnior, Renato Beozzo Bassanezi

LEG - Laboratório de Estatística e Geoinformação <http://www.est.ufpr.br/leg> - Departamento de Estatística - UFPR

Resumo

O padrão espacial da incidência da doença dentro de talhões de plantas é avaliado através de técnicas de estatística espacial. Foram empregados métodos baseados em *quadrat counts* e distância euclidiana entre plantas doentes. Para testar a hipótese de aleatoriedade espacial, fez-se um teste Qui-quadrado na avaliação por *quadrat counts* e um teste de Monte Carlo na avaliação por distância. Foram desenvolvidas funções em R (6) para a aplicação dos testes. Também foram desenvolvidas funções para a organização e validação dos dados em formato espaço-temporal. Os resultados indicaram padrão espacial na maioria das avaliações. O método da distância mínima média é mais poderoso para detectar agregação, seguido pelo índice de dispersão binomial.

Palavras-chave: *padrão espacial, índice de dispersão, método da distância mínima.*

1 Introdução

A morte súbita dos citros, MSC, é uma nova doença dos citros, que provoca rápido definhamento e morte de plantas enxertadas em limoeiro *Cravo*. Até o momento pouco se sabe sobre a etiologia da doença, apenas que se conseguiu a transmissão da doença de uma planta para outra (1).

A análise de agregação é feita para avaliar o padrão da incidência de uma determinada doença no espaço: regular, aleatório ou agregado. Esta análise é preliminar à uma modelagem estatística espacial (3).

2 Metodologia

Os métodos empregados para avaliação do padrão espacial consideraram a avaliação por *quadrat counts* e pela distância mínima média entre plantas doentes.

2.1 Análise por *quadrat counts*

Define-se *quadrat* uma área retangular de plantas abrangendo determinado número de linhas l e de plantas na linha c sendo o número de plantas em um *quadrat* $n = l * c$.

Na análise por *quadrat counts*, o número de plantas doentes em cada *quadrat* é usado na avaliação da dispersão espacial da doença no talhão. Para isto utiliza-se o índice de dispersão D , que é a razão entre a variância observada em um conjunto de dados e a variância esperada segundo um modelo.

O índice de dispersão é calculado para medir a agregação e para testar superdispersão, podendo-se avaliar a distribuição espacial. Se os dados tem distribuição aleatória, D será igual a 1, se a distribuição dos dados é regular D é menor que 1 e se existe agregação dos dados D é maior que 1 (4).

Sendo N o número de *quadrats* avaliados, a quantidade $D(N - 1)$ tem distribuição χ^2_{N-1} , podendo ser utilizada para testar a hipótese de aleatoriedade espacial da incidência da doença.

2.1.1 Índice de Dispersão Binomial

O modelo Binomial pode ser utilizado para o número de plantas doentes por *quadrat*, subdividindo-se o talhão em *quadrats* de tamanho fixo.

Devido ao fato de se ter n fixo, o número de plantas consideradas nesta análise pode não corresponder ao número total de plantas no talhão. Isso ocorre quando a dimensão do talhão não é múltipla da dimensão dos *quadrats*, ou quando existem falhas em um *quadrat*.

2.1.2 Índice de Dispersão Poisson

Na tentativa de se considerar maior número de plantas na análise sem desprezar os *quadrats* incompletos, pode-

se avaliar a agregação utilizando-se um modelo Poisson, calculando-se o índice de dispersão Poisson.

Esta metodologia é equivalente a um teste qui-quadrado, onde o número esperado de plantas em um *quadrat* é calculado em função de seu tamanho e da incidência da doença no talhão.

2.2 Método da Distância Mínima Média

Define-se distância mínima a distância entre uma planta doente e a planta doente mais próxima a ela. Calculando-se esta distância para cada planta doente, pode-se obter a distância mínima média. Quanto maior a agregação da doença, menor será a distância mínima média. Portanto para um mesmo número de plantas doentes, a distância mínima será menor quanto mais agregação.

Desconhecendo-se a distribuição dessa estatística, pode-se realizar um teste Monte Carlo para testar a hipótese de aleatoriedade espacial. O teste Monte Carlo neste caso consiste em calcular a distância mínima média observada no conjunto de dados e simular $s - 1$ conjuntos de dados com padrão aleatório, calculando para cada um a distância mínima média. O (p-valor) é dado pela proporção de distâncias menores ou iguais à distância mínima média observada.

Seja dm_1 a distância mínima média observada e dm_i : $i = 2, \dots, s$ a i -ésima distância sob a hipótese de aleatoriedade. Seja $dm_{(j)}$ a j -ésima estatística de ordem, então

$$P[dm_1 = dm_{(j)}] = s^{-1} : j = 1, \dots, s. \quad (1)$$

e para o cálculo do p-valor, seja $dm_1 = dm_{(k)}$, então o p-valor será k/s .

Para um teste ao nível de 5% de significância, $s = 100$ é adequado (3).

3 Implementações

Foram programados funções e procedimentos no *software* R, incluindo rotinas para leitura e formatação automática dos dados, validação dos dados, análise por *quadrat counts*, análise por método da distância mínima.

3.1 Leitura e Formato e Validação Temporal dos Dados

Os dados foram coletados e armazenados em planilhas e são convertidos para o formato espaço-temporal.

Também implementou-se uma função para correção de erros de digitação e uma para separação de dados com determinado código utilizada para selecionar os dados a serem analisados, separando plantas ou posições com códigos R e F, que indicam replante e falha, respectivamente.

Considerando que a Morte Súbita dos Citrus não regrida no estado da doença, foi implementada uma função para detectar e corrigir a inconsistência dos dados (1).

A importância de se ter funções para a validação, visa automatizar a validação dos dados de milhares de talhões que serão analisados posteriormente.

3.2 Análise por *quadrat counts*

Para análise por *quadrats*, pode ou não ser necessário que todos os *quadrats* tenham o mesmo número de plantas, dependendo do modelo utilizado.

Na função implementada para esse método, os dados em formato de planilha ou de coordenada-atributo, a dimensão do *quadrat* dada pelo número de plantas nos sentidos de entre linhas e plantas da rua e o nível de significância.

3.2.1 Índice de Dispersão Binomial

Para esse modelo, podem ser avaliados todos os possíveis *quadrats* na dimensão especificada. Quando as dimensões da matriz de dados não é múltipla das dimensões

do *quadrat*, os dados “excedentes” são desconsiderados da análise. Se faltar alguma planta para completar um *quadrat* dentro do talhão, este também é desconsiderado da análise.

3.2.2 Índice de Dispersão Poisson

Para esse modelo considera-se os dados de todos os *quadrats* onde o número de plantas considerado é maior que a metade do número de plantas definido pelo tamanho nominal do *quadrat* e não considera os dados de *quadrats* com menor número de plantas. Em geral, o número de plantas consideradas nesta análise é maior que na análise por dispersão binomial.

3.3 Método da Distância Mínima

A função em R para esse método, recebe os dados no formato *geodata* contendo coordenadas e atributo da planta no talhão, o código atribuído às plantas doentes e o número de simulações, e retorna o valor observado, os valores simulados e o p-valor obtido no teste de Monte Carlo.

4 Conclusões, Discussões e Trabalhos Futuros

Nos resultados das avaliações dos nove talhões analisados, obteve-se resultados indicando padrões agregado, aleatório e regular, exceto na avaliação pelo índice de dispersão Binomial, onde não houve nenhum resultado apontando padrão regular, ao nível de 10%. Nos dados analisados o padrão de agregação é aleatório para baixas incidências da doença e, em geral, variou do agregado para o regular a medida que aumentava a incidência da doença.

O resultado para avaliação por *quadrat counts* varia quando quando se varia o tamanho ou forma dos *quadrats*. O resultado pelo método da distância mínima média é único, podendo variar apenas devido ao número de simulações de Monte Carlo. Além disso ressalta-se que este método é o único que considera o espaçamento de plantio das plantas e considera todas as plantas do talhão, independentemente da forma do mesmo.

A avaliação por *quadrat counts* não detecta o que podemos chamar de agregação local, isto é, quanto tem-se muitos focos de doença no talhão e em cada foco existem várias plantas doentes próximas umas das outras.

Nos dados analisados, o padrão de agregação é aleatório para baixas incidências da doença e, em geral, variou do agregado para o regular a medida que aumenta a incidência da doença.

Na seqüência serão empregados métodos de análise de processos pontuais e análise de sobrevivência, utilizando interfaces de integração com Sistemas de Informações Geográficas - SIG, para análises em toda a região produtora.

Referências

- [1] BASSANEZI, R. B. and FERNANDES, N. G. and YAMMAMOTO, P. T., Morte Súbita do Citros, FUNDECITRUS, 2003, Araraquara, SP, Brasil
- [2] CAMPBELL, C. L. and MADDEN, L. V. *Introduction to plant disease epidemiology* John Wiley, 1990.
- [3] DIGGLE, PETER J., *Statistical Analysis of Spatial Point Patterns*, Oxford University Press Inc, 2003.
- [4] MADDEN, L. V. and HUGHES, G. *Plant Disease Incidence: Distributions, Heterogeneity, and Temporal Analysis* Phytopathology, 1995.
- [5] ORTIZ, G. C. and RIBEIRO JR, P. J. Análises Preliminares dos Dados de MSC nos Talhões, FUNDECITRUS, 2004.
- [6] R Development Core Team. *R: A language and environment for statistical computing*. Foundation for Statistical Computing, Vienna, Austria, 2003. ISBN 3-900051-00-3.