

Análise de ensaios de sanidade pós-colheita considerando a estrutura longitudinal das avaliações

Aline de Holanda Nunes Maia¹

Daniel Terao¹

Resumo: O uso de técnicas de exploração, visualização e análise inferencial de dados provenientes de estudos de patologias pós-colheita constitui um importante componente no processo de geração e comunicação de evidências experimentais para orientar o desenvolvimento de inovações tecnológicas para controle dessas patologias. Tradicionalmente, para análise inferencial de ensaios de doenças pós-colheita são utilizados modelos lineares ordinários, isoladamente, para cada data de avaliação. Os dados originais são geralmente transformados, buscando evitar violações de pressupostos de normalidade e homogeneidade de variâncias. Esses métodos apresentam limitações, especialmente por não considerarem de forma adequada a natureza das variáveis e as prováveis correlações temporais entre as observações feitas num mesmo fruto ao longo do tempo. A escolha dos métodos adequados de análise inferencial de variáveis em qualquer ensaio depende tanto do delineamento experimental utilizado, quanto da natureza das variáveis de interesse. Quando variáveis são observadas ao longo do tempo numa mesma unidade experimental, caracterizando um estudo longitudinal ou de medidas repetidas é necessária especial atenção para a modelagem das prováveis correlações temporais entre as medidas, além da questão do tipo de distribuição de probabilidade da variável. Para variáveis Normais, essa modelagem pode ser feita via Modelos Lineares Gerais; para variáveis não Normais, deve-se utilizar uma generalização dos GLM, os Modelos Lineares Generalizados Mistos. Apresentamos, um exemplo de análises exploratória, análise inferencial e visualização de dados de um ensaio para avaliar a eficiência de tratamentos térmicos para controle do bolor verde da laranja.

Palavras-chave: análise de medida repetidas, visualização de dados, análise de agrupamentos, pós-colheita

¹ Embrapa Meio Ambiente. E-mail: aline.maia@embrapa.br

Analysis of post-harvest pathology experiments accounting for the longitudinal nature of disease monitoring data

Aline de Holanda Nunes Maia¹

Daniel Terao¹

Abstract: We present an exploratory analysis, data visualization and inferential analysis for post-harvest pathology assays. We used Generalized Linear Models with an AR1 covariance structures for each treatment, which accounts for the longitudinal natures of disease monitoring in fruit.

Key words: repeated measures, data visualization, cluster analysis

¹ Embrapa Meio Ambiente. E-mail: aline.maia@embrapa.br

Análise de ensaios de sanidade pós-colheita considerando a estrutura longitudinal das avaliações

Aline de Holanda Nunes Maia¹

Daniel Terao¹

1 Introdução

O uso de técnicas de exploração, visualização e análise inferencial de dados provenientes de estudos de patologias pós-colheita constitui um importante componente no processo de geração e comunicação de evidências experimentais para orientar o desenvolvimento de inovações tecnológicas para controle dessas patologias. A eficiência de métodos de controle de doenças pós-colheita em frutos é geralmente avaliada via ensaios realizados em ambientes controlados onde variáveis de interesse de natureza diversa (severidade, porcentagem de infecção, diâmetro de lesão) são monitoradas ao longo do tempo. Tradicionalmente, para análise inferencial de ensaios dessa natureza, são utilizados modelos lineares ordinários, isoladamente, para cada data de avaliação. Os dados originais são geralmente transformados, buscando evitar violações de pressupostos de normalidade e homogeneidade de variâncias. Esses métodos apresentam limitações, especialmente por não considerarem de forma adequada a natureza das variáveis e as prováveis correlações temporais entre as observações feitas num mesmo fruto ao longo do tempo. A escolha dos métodos adequados de análise (estatística) inferencial de variáveis em qualquer ensaio depende tanto do delineamento experimental utilizado, quanto da natureza das variáveis de interesse. Quando variáveis são observadas ao longo do tempo numa mesma unidade experimental, caracterizando um estudo longitudinal ou de medidas repetidas é necessária especial atenção para a modelagem das prováveis correlações temporais entre as medidas, além da questão do tipo de distribuição de probabilidade da variável. Para variáveis Normais, essa modelagem pode ser feita via Modelos Lineares Gerais (GLM, *General Linear Models*; McCULLAGH & NELDER, 1989) com medidas repetidas utilizando o Proc Mixed do SAS[®]/STAT (SAS[®], 2008); para variáveis não Normais (ex. Poisson, Binomial), deve-se utilizar uma generalização dos LMM, os Modelos Lineares Generalizados Mistos (GLMM, *generalized Linear Mixed Models*; BRESLOW & CLAYTON, 1983), disponíveis no Proc GLIMMIX (SAS[®], 2008). Apresentamos, um exemplo de análises exploratória, análise inferencial e visualização de dados de um ensaio para avaliar a eficiência de tratamentos térmicos para controle do bolor verde da laranja. Esses métodos são úteis para seleção de tratamentos sanitários preventivos e comunicação de resultados em estudos sobre alternativas de controle de patologias pós-colheita.

2 Material e Métodos

Para análise exploratória dos dados e visualização dos dados, foram construídos gráficos dos perfis longitudinais de infecção em cada tratamento, mensurados pelo diâmetro médio de lesão. Foram construídos dendogramas para avaliar a similaridade entre tratamentos com relação ao progresso da doença. A abordagem proposta foi demonstrada utilizando os dados de um ensaio para avaliação de eficiência de tratamentos térmicos para controle do bolor verde da laranja, uma patologia pós-colheita causada pelo fungo *Penicillium digitatum*. Foram testados os seguintes tratamentos: i) testemunha absoluta - laranjas não tratadas; ii) fungicida – laranjas tratadas com fungicida tradicional; iii) imersão em água a 52°C por 120 s – tratamento térmico convencional; iv) imersão em água a 52°C por 60 s; v) imersão em água a 55°C por 30 s; vi) imersão em água a 60°C por 30 s; vii) imersão em água a 65°C por 30 s e viii) imersão em água a 70°C por 30 s.

Foi ajustado um modelo linear para avaliar o efeito dos tratamentos sobre a severidade da infecção, mensurada pelo diâmetro médio de lesão, correspondente à média dos diâmetros transversal e longitudinal da lesão decorrente da inoculação proposital do fungo em cada fruto. Foram consideradas

¹ Embrapa Meio Ambiente. E-mail: aline.maia@embrapa.br

estruturas de correlação autoregressivas de ordem um (AR1) entre os diâmetros médios de lesão medidos num mesmo fruto, com parâmetros ρ específicos para cada tratamento. Para exemplificar o uso de contrastes para testar hipóteses sobre combinações lineares das médias dos tratamentos, comparamos os tratamentos térmicos alternativos com a testemunha absoluta.

3 Resultados e Discussão

Pela análise de contrastes, observa-se que todos os tratamentos apresentaram desempenho superior à testemunha absoluta e ao fungicida tradicional (Tabelas 1 e 2). Na análise gráfica, vemos os tratamentos térmicos alternativos com comportamento similar e um comportamento diferenciado para a testemunha absoluta, tratamento com fungicida e imersão tradicional. Nota-se a ineficiência relativa do fungicida, com altas taxas de infecção ao longo do período e o bom desempenho da imersão tradicional (Figuras 1 e 2).

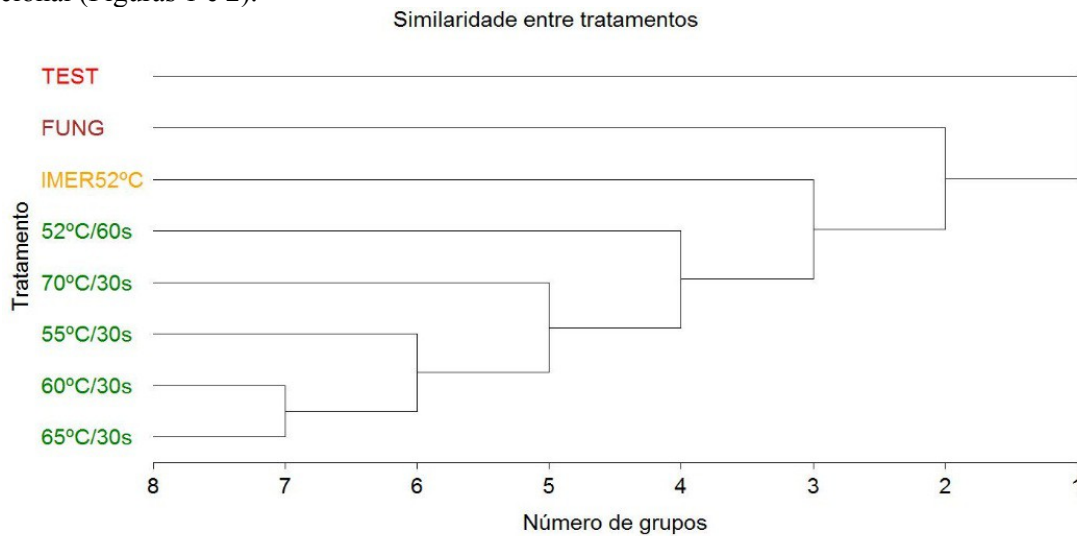


Figura 1. Árvore de classificação ou dendograma representando o agrupamento dos tratamentos com relação à similaridade entre perfis de evolução temporal de infecção de 7 a 18 dias após inoculação do agente causal do bolor verde da laranja (*Penicillium digitatum*).

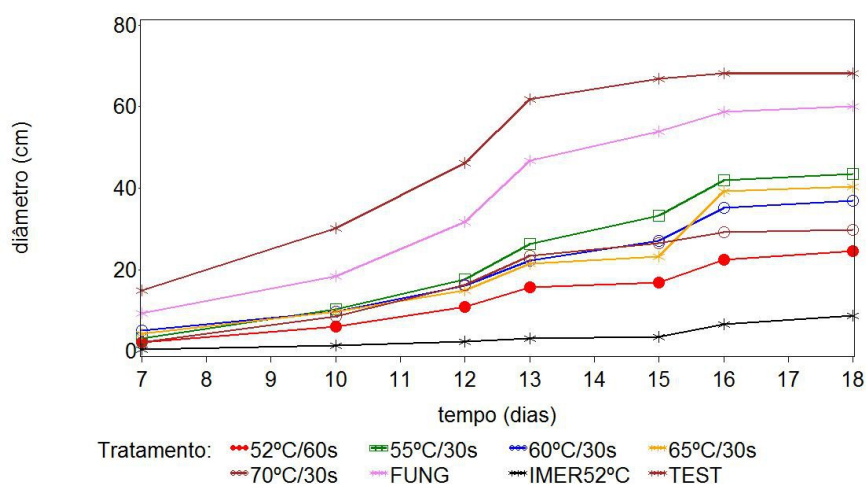


Figura 2. Evolução temporal da infecção no período de 7 a 14 dias após inoculação do agente causal do bolor verde da laranja (*Penicillium digitatum*).

Tabela 1. Contrastes entre a testemunha absoluta e os demais tratamentos testados, aos 7, 13 e 18 dias após a inoculação.

Contraste	Estimativa	Erro Padrão	t	Valor p
52°C/60s x TEST-T7	-12.81	4.25	-3.02	0.0026
52°C/60s x TEST - T13	-46.11	4.25	-10.86	<0.0001
52°C/60s x TEST - T18	-43.77	4.25	-10.30	<0.0001
55°C/30s x TEST-T7	-11.78	4.47	-2.64	0.0084
55°C/30s x TEST - T13	-35.52	4.47	-7.95	0.0001
55°C/30s x TEST - T18	-24.81	4.47	-5.55	>0.0001
60°C/30s x TEST-T7	-9.91	4.66	-2.13	0.0336
60°C/30s x TEST - T13	-39.55	4.66	-8.49	<0.0001
60°C/30s x TEST - T18	-31.30	4.66	-6.72	<0.0001
65°C/30s x TEST-T7	-10.55	4.52	-2.33	0.0198
65°C/30s x TEST - T13	-40.30	4.52	-8.91	<0.0001
65°C/30s x TEST - T18	-27.98	4.52	-6.19	<0.0001
70°C/30s x TEST-T7	-12.65	4.44	-2.85	0.0045
70°C/30s x TEST - T13	-38.44	4.44	-8.66	<0.0001
70°C/30s x TEST - T18	-38.43	4.44	-8.65	<0.0001
FUNG x TEST-T7	-5.65	3.64	-1.55	0.1207
FUNG x TEST - T13	-15.08	3.64	-4.14	<0.0001
FUNG x TEST - T18	-8.24	3.64	-2.26	0.0237
IMER52 x TEST-T7	-14.47	3.00	-4.82	<0.0001
IMER52 x TEST - T13	-58.56	3.00	-19.51	<0.0001
IMER52 x TEST - T18	-59.46	3.00	-19.81	<0.0001

Tabela 1. Contrastes entre o tratamento com fungicida e os tratamentos térmicos alternativos, aos 7, 13 e 18 dias após a inoculação.

52°C/60s x FUNG-T7	-7.16	5.08	-1.41	0.1586
52°C/60s x FUNG - T13	-31.04	5.08	-6.11	<.0001
52°C/60s x FUNG - T18	-35.53	5.08	-7.00	<.0001
55°C/30s x FUNG-T7	-6.13	5.26	-1.16	0.2443
55°C/30s x FUNG - T13	-20.45	5.26	-3.88	0.0001
55°C/30s x FUNG - T18	-16.57	5.26	-3.15	0.0017
60°C/30s x FUNG-T7	-4.26	5.43	-0.78	0.4327
60°C/30s x FUNG - T13	-24.47	5.43	-4.51	<.0001
60°C/30s x FUNG - T18	-23.06	5.43	-4.25	<.0001
65°C/30s x FUNG-T7	-4.90	5.31	-0.92	0.3565
65°C/30s x FUNG - T13	-25.22	5.31	-4.75	<.0001
65°C/30s x FUNG - T18	-19.74	5.31	-3.72	0.0002
70°C/30s x FUNG-T7	-6.99	5.24	-1.33	0.1823
70°C/30s x FUNG - T13	-23.36	5.24	-4.46	<.0001
70°C/30s x FUNG - T18	-30.19	5.24	-5.76	<.0001

Apesar da disponibilidade de métodos mais adequados para a análise de variáveis não Normais com medidas repetidas, seu uso é ainda restrito na análise de dados biológicos, para a qual prevalecem abordagens mais tradicionais que utilizam transformações de dados e comparação de médias para cada diferente momento de avaliação ou no momento final, ignorando a natureza longitudinal dos dados, de importância fundamental nesse tipo de problema, onde a evolução temporal da doença é informação crítica para o planejamento de processos de transporte e armazenamento dos produtos agrícolas. O uso de métodos apropriados de análises inferenciais permite quantificar adequadamente as incertezas envolvidas nos processos de decisão baseados em resultados experimentais; já a visualização de dados facilita a escolha dos métodos de análise inferencial e numa etapa posterior, a comunicação das informações obtidas nos ensaios aos clientes potenciais (TUFTE, 2001), fomentando o desenvolvimento de alternativas inovadoras para o controle de patologias pós-colheita.

4 Conclusões

Os métodos inferenciais e de visualização de dados aqui apresentados são complementares e possibilitam uma exploração mais rica de dados de ensaios de patologias pós-colheita e melhor divulgação dos resultados para clientes potenciais.

O uso de um modelo com estrutura de correlações entre as medidas é mais compatível com estudos dessa natureza onde medições são feitas ao longo do tempo no mesmo fruto que o uso de uma análise de variância com pressupostos de independência e homocedasticidade.

5 Bibliografia

[1] BRESLOW, N.E.; CLAYTON, D.G. Approximate inference in generalized linear mixed models. **JASA**, n. 88, v.421, p. 9–25, 1983.

[2] McCULLAGH, P.; NELDER, J. A. **Generalized linear models**, 2nd. ed. London: Chapman & Hall, 511 p., 1989.

[3] SAS INSTITUTE INC. **SAS/STAT user's guide/** Version 9.2. Cary, NC: SAS Institute Inc., 2008.

[4] TUFTE, E. F. **The Visual Display of Quantitative Information**, 2nd. ed. Cheshire, Graphics Press, Cheshire, 2001.