

Avaliação da severidade do cancro cítrico em genótipos de laranja doce por meio de modelos mistos

Haward A. S. Américo^{1†}, Lucimary A. Santos², Vanderly Janeiro³

¹Acadêmico da Universidade Estadual do Paraná (UNESPAR).

²Docente da Universidade Estadual do Paraná (UNESPAR).

³Docente da Universidade Estadual de Maringá (UEM).

Resumo: A doença do cancro cítrico que afeta folhas e frutos das laranjeiras, ocasiona grande prejuízo econômico, assim, a utilização de metodologias estatísticas na análise dos dados de citricultura é fundamental. Em muitos estudos sobre o comportamento da doença podem ocorrer várias medidas realizadas em uma mesma unidade experimental, caracterizando-as como medidas repetidas e fazendo com que seja necessário o uso de métodos estatísticos que considerem este fato. Neste sentido, os modelos lineares de efeitos mistos têm se tornado uma ferramenta importante de análise. O objetivo, neste trabalho, foi estudar e aplicar a teoria de modelos lineares mistos a um conjunto de dados, proveniente de um experimento implementado na região Noroeste do Estado do Paraná cujo interesse foi o de avaliar a severidade do cancro cítrico em folhas de laranja doce, para catorze genótipos diferentes. Com auxílio de pacotes disponíveis no ambiente estatístico R, observou-se que o modelo linear misto proporcionou um bom ajuste do modelo aos dados. Ainda, concluiu-se que algumas das variedades estudadas são mais suscetíveis ao desenvolvimento do cancro cítrico que outras.

Palavras-chave: Modelos lineares mistos; Cancro cítrico; severidade; ambiente estatístico R.

Assessment of citrus canker severity in sweet orange genotypes by mixed models

Abstract: The citrus canker disease affects leaves and fruits of orange trees and causes great economic losses, therefore, the use of statistical methodologies in the analysis of citrus datasets is essential. In a lot of studies about the disease behavior, several measurements may be carried out in the same experimental unit, characterizing them as repeated measures, making it necessary to use statistical methods that take this fact into account. In this sense, linear mixed effects models have become an important analysis tool. The objective, in this work, was to study and apply the theory of linear mixed models to a dataset, coming from an experiment implemented in the Northwest region of the Paraná State, whose interest was to evaluate the severity of citrus canker in sweet orange leaves for fourteen different genotypes. Using available packages in the R statistical environment, it was observed that the mixed linear model provided a good fit of the model to the data. Even, it was concluded that some of the experimental varieties are more susceptible to the development of citrus canker than others.

Keywords: Linear mixed models; Citrus canker; severity; statistical environment R.

Introdução

O Brasil é reconhecido como o maior produtor mundial de laranjas, com destaque para a variedade pera, que compõe a maior parte de sua produção citrícola.

No estado do Paraná, por um longo período, a produção de cítricos, nas regiões Norte e Noroeste, ficou prejudicada por ter ocorrido interdição das áreas de plantio devido a doença do cancro cítrico. Atualmente, a produção se concentra nestas regiões, que abrigam as maiores usinas de suco de laranja concentrado, produto de grande relevância no mercado internacional e com significativo impacto positivo na economia do país. (Gonçalves-Zuliani, 2014).

[†]Autor correspondente: antunnyamerico@gmail.com.

A bactéria *Xanthomonas citri* subsp. *citri* é a grande responsável pela origem do cancro cítrico que traz prejuízos à economia. Contudo, graças à implementação de práticas de manejo integrado do cancro cítrico, o uso de variedades mais resistentes, a utilização de quebra-ventos e pulverizações com bactericidas cúpricos, a produção está sendo retomada com sucesso (Gonçalves-Zuliani, 2014).

Nesse contexto, a utilização de técnicas estatísticas na análise de dados da citricultura brasileira se torna fundamental para compreender e aprimorar os processos de produção, levando em conta as variáveis de interesse e os fatores que possam exercer influência no setor citrícola.

Diferentes metodologias foram aplicadas para avaliação da suscetibilidade de genótipos de cítricos ao cancro buscando identificar genótipos mais resistentes à doença. Volpato (2021) comparou distribuições de probabilidade, fazendo o uso de modelos lineares generalizados mistos, na análise de resistência ao cancro cítrico. Cazetta (2021) fez o uso de análise de variância e teste de Scott-Knott no estudo do diâmetro das lesões. Longhini *et al.* (2019) aplicaram análise de variância não paramétrica de Kruskal-Wallis pois pressupostos para utilização de metodologia paramétrica não foram atendidos.

Em várias ocasiões a variável dependente é medida mais de uma vez em uma mesma unidade experimental gerando correlação entre estas observações. Definimos este tipo de dados como de medidas repetidas que requerem ajuste de modelos que envolvem a estimação de parâmetros de covariância para capturar essa correlação. Modelos lineares de efeitos mistos (MLM) tem se tornado uma ferramenta bastante útil quando se tratam deste tipo de dados. Essa metodologia inclui fatores de efeitos fixos e de efeitos aleatórios ao modelo linear tradicional, de forma a permitir que a correlação linear padrão e a variabilidade existentes em uma unidade experimental possam ser modeladas adequadamente.

O estudo das características deste modelo e sua aplicabilidade, neste caso, aos dados de citricultura proporciona *insights* valiosos, que contribuem para a otimização da produção, o manejo adequado de doenças e pragas, e o fomento do desenvolvimento sustentável desse importante setor agrícola.

Dessa forma, o principal objetivo deste trabalho é o estudo das propriedades dos modelos mistos, e sua aplicabilidade, além da exploração computacional de rotinas existentes no ambiente estatístico R.

Materiais e Métodos

Materiais

A doença cancro cítrico está presente na agricultura, mundialmente, devido sua alta disseminação que acarreta debilitação das árvores, perda na qualidade e produtividade de frutos, e prejuízos econômicos consideráveis (Volpato, 2019, Gonçalves-Zuliani, 2014). Causada por meio da bactéria *Xanthomonas citri* subsp. *citri*, o patógeno demonstra os sintomas nas folhas e frutos da planta por meio de lesões salientes, necróticas de coloração marrom circundadas por um halo amarelo (Gottwald *et al.*, 2002 citado por Longhini *et al.*, 2019).

Na Figura 1 são apresentados os danos que a doença pode causar nos frutos e folhas das plantas. Os dados utilizados nesta aplicação, referem-se a um recorte dos dados apresentados em Cazetta (2021) e são provenientes da análise de 14 genótipos de citros, denominados G1, G2, G3, . . . , G14, para os quais foram coletados 5 ramos sadios, no mesmo estágio de maturação para cada genótipo. Após terem sido devidamente lavados e desinfetados, foi destacada 1 folha de cada ramo, e realizada a inoculação da bactéria por meio de uma agulha esterilizada. Cada folha foi perfurada 8 vezes e, após um período de 14 dias da inoculação do patógeno, por meio de um paquímetro eletrônico, avaliou-se o diâmetro das lesões.

O tamanho amostral considerado foi 560, ou seja, 14 genótipos; 5 folhas destacadas por genótipo; e 8 perfurações em cada folha.

Figura 1: Sintomas do cancro cítrico.



Fonte: Oliveira (2008).

Métodos

Os modelos lineares mistos (MLM), na literatura, são utilizados como alternativa na modelagem de diversas situações em que as suposições para utilização dos modelos tradicionais não são satisfeitas.

Aqui, em particular, consideramos o uso dos modelos lineares mistos aplicados a delineamentos com medidas repetidas, para os quais são observados os efeitos de um ou mais tratamentos, pelo menos duas vezes, numa mesma unidade experimental.

A utilização de modelos lineares mistos permite acomodar a presença de variáveis preditoras e covariáveis que apresentam efeitos fixos e aleatórios. Considerando o modelo apresentado por Pinheiro e Bates (2000),

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\varepsilon}_i, \quad i = 1, \dots, M \quad (1)$$

$$\mathbf{b}_i \sim N(0, \mathbf{G}), \quad \boldsymbol{\varepsilon}_i \sim N(0, \mathbf{R}),$$

em que \mathbf{y}_i é o vetor de resposta para o i -ésimo grupo; $\boldsymbol{\beta}$ é o vetor de efeitos fixos com dimensão p ; \mathbf{b}_i é o vetor de efeitos aleatórios de dimensão q ; \mathbf{X}_i é a matriz de efeitos fixos ($n_i \times p$); \mathbf{Z}_i é a matriz de efeitos aleatórios ($n_i \times q$) e $\boldsymbol{\varepsilon}_i$ é o vetor de erro aleatório. As matrizes de covariância dos efeitos aleatórios e dos erros aleatórios são definidas por $\mathbf{G} = \sigma_b^2 \mathbf{I}_q$ e $\mathbf{R} = \sigma^2 \mathbf{I}_n$.

Considerando que para a aplicação aqui apresentada, cada folha foi obtida aleatoriamente e, não havendo interesse na comparação entre as folhas e que foram feitas 8 avaliações em cada folha trata-se de medidas repetidas dentro de folhas, assim, folha será considerada como efeitos aleatórios no modelo. O modelo de efeitos aleatórios para o diâmetro da lesão no experimento é escrito como:

$$y_{ijl} = \mu + \tau_i + b_j + \varepsilon_{ijl}, \quad i = 1, \dots, 14, \quad j = 1, \dots, 70, \quad l = 1, \dots, 560. \quad (2)$$

$$b_j \sim N(0, \sigma_b^2), \quad \varepsilon_{ijl} \sim N(0, \sigma_\varepsilon^2),$$

em que y_{ijl} é o diâmetro da lesão no l -ésimo furo da j -ésima folha e i -ésimo genótipo; μ é o diâmetro médio da lesão na população; τ_i o efeito fixo do i -ésimo genótipo; b_j o efeito aleatório da j -ésima folha e ε_{ijl} o termo do erro aleatório.

Método para estimação dos parâmetros

Para a estimação dos parâmetros de efeitos fixos, β , e os parâmetros de covariância G e R utilizaremos o método de máxima verossimilhança restrita (MVR), pois tende a apresentar estimativas dos componentes de variância menos viesadas, uma vez que é possível fazer a correção do viés na estimação de parâmetros (West, Welch E Galecki, 2014). Contudo, a estimação das componentes de variância implica na maximização da verossimilhança de uma transformação linear ortogonal $y^* = U'y$, em que U é uma matriz de posto completo, de dimensão $(N \times N - p)$, com colunas ortogonais às colunas da matriz X e é da forma $\mathbb{E}(y^*) = 0$.

O resultado dessa transformação não depende do vetor de parâmetros β e nem da matriz U . Comumente, utiliza-se $U = I - X(X'X)^{-1}X'$, sendo a matriz de projeção que gera os resíduos do ajuste atingido por mínimos quadrados ordinários.

Logo, $y^* \sim N(0, U'VU)$ e a função log-verossimilhança restrita, $l_R(\theta)$, é

$$l_R(\theta) = -\frac{1}{2} \ln(2\pi) \sum_{i=1}^m n_i - \frac{1}{2} \sum_{i=1}^m \ln|V_i| - \frac{1}{2} \sum_{i=1}^m (y_i - X_i\beta)' V_i^{-1} (y_i - X_i\beta) \quad (3)$$

$$- \frac{1}{2} \ln \left| \sum_{i=1}^m X_i' V_i^{-1} X_i \right| \quad (4)$$

Crítérios para seleção do modelo

A utilização de um modelo adequado aos dados a serem analisados definem a eficácia das estimativas. Uma ferramenta útil na seleção de modelos são os critérios de informação, cuja regra de decisão é a de que quanto menor o valor, melhor será o modelo.

O critério de informação de Akaike (AIC) (Akaike, 1973) e o critério de informação Bayesiano (BIC), apresentados em Pinheiro e Bates (2000) nos auxiliam nesse processo, e podem ser definidos, respectivamente, como

$$AIC = -2l(\hat{\beta}, \hat{\theta}) + 2p \quad (5)$$

sendo $l(\beta, \theta)$ o logaritmo da função de verossimilhança e p o número total de parâmetros a serem estimados, e

$$BIC = -2l(\hat{\beta}, \hat{\theta}) + p \times \ln(n) \quad (6)$$

em que n é o número de observações.

Resultados e discussões

A realização de uma inspeção dos dados antes da análise é uma etapa fundamental para obtenção de resultados confiáveis. Essa etapa inicial permite identificar as particularidades dos dados, como distribuição, tendências, padrões e *outliers*, entre outros aspectos relevantes. Ao compreender melhor a natureza dos dados, é possível tomar decisões adequadas durante a análise e na escolha do modelo que melhor se adequa aos dados em estudo.

Apresentamos, inicialmente, uma análise exploratória dos dados. Na Tabela 1 são apresentadas medidas descritivas básicas para cada um dos genótipos, sendo: valores mínimo e máximo para o diâmetro da lesão, média dos valores observados, o desvio padrão (DP) e o coeficiente de variação (CV).

Tabela 1: Estatísticas Resumo por Genótipo

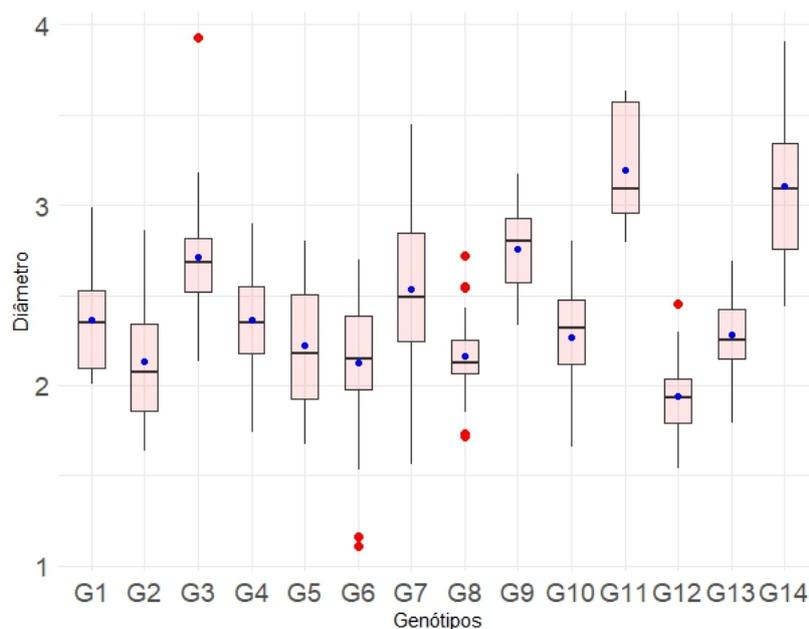
Genótipos	Mín.	Máx.	Média (mm)	DP	CV%
G1	2,01	2,99	2,3667	0,2833	11,9688
G2	1,64	2,86	2,1322	0,3374	15,8258
G3	2,13	3,93	2,7107	0,3010	11,1056
G4	1,74	2,90	2,3645	0,2975	12,5813
G5	1,67	2,80	2,2190	0,3323	14,9739
G6	1,11	2,70	2,1277	0,3693	17,3586
G7	1,56	3,45	2,5352	0,4572	18,0354
G8	1,72	2,72	2,1650	0,2081	9,6117
G9	2,33	3,17	2,7550	0,2271	8,2424
G10	1,66	2,80	2,2645	0,2972	13,1232
G11	2,79	3,63	3,1920	0,2920	9,1469
G12	1,54	2,45	1,9382	0,2038	10,5138
G13	1,79	2,69	2,2795	0,2085	9,1461
G14	2,44	3,91	3,1067	0,3787	12,1906

Fonte: Autores

Percebeu-se que, após 14 dias de inoculação da bactéria nas folhas, o genótipo G12 apresentou menor diâmetro médio de lesão (1,938mm), já os genótipos G11 e G14 foram os que apresentaram maior diâmetro médio de lesão, sendo 3,192mm e 3,106mm respectivamente. Ao avaliarmos o coeficiente de variação observou-se que o genótipo G9 apresentou a menor variabilidade em relação à média (8,2424%) e o G7, a maior variabilidade (18,0354%).

No gráfico de boxplot (Figura 2) percebem-se valores discrepantes (indicados por pontos vermelhos) que podem influenciar a média dos diâmetros das lesões, entretanto, como a variabilidade não é alta (coeficientes de variação menores que 20%), para este caso, em particular, provavelmente não ocasiona problemas que interfiram nas análises.

Figura 2: Distribuição dos genótipos de citros por diâmetro da lesão.



Fonte: Autores.

O genótipo G12 apresentou menor crescimento no diâmetro das lesões, indicando menor suscetibilidade à doença, enquanto os genótipos G11 e G14 mostraram maiores valores e maiores dispersões.

Com a finalidade de testar a hipótese $H_0 : \sigma_b = 0$, foram ajustados dois modelos, o primeiro considerando apenas o efeito fixo de genótipo (M1) e o segundo levando em consideração o efeito aleatório de folhas (M2) (como apresentado em (2)). Ambos ajustes foram feitos por usando MVR, por meio das funções `lme` e `gls` do pacote `nlme` com `method = "REML"`. Optou-se por fazer o ajuste do modelo com a função `lmer` do pacote `lme4` e, em seguida, usar a função `ranova` do mesmo pacote, na obtenção da estimativa da estatística da razão de verossimilhança entre os modelos cujo valor obtido foi 29,2873. Essa função realiza o teste considerando uma distribuição $\chi^2_{(1)}$, entretanto, como preconizado por Pinheiro e Bates (2000), novos cálculos foram realizados considerando uma mistura de distribuições χ^2 . Para este caso considerou-se $0,5\chi_0^2 + 0,5\chi_1^2$ que forneceu a probabilidade de significância $3,12 \times 10^{-8}$ levando a rejeição da hipótese H_0 , assim, o efeito aleatório de folha deve ser mantido no modelo.

Ainda, os valores obtidos por meio dos critérios AIC e BIC indicaram que modelo que considerou folhas como de efeitos aleatórios, M2, é mais adequado aos dados, conforme apresentado na Tabela 2.

Tabela 2: Valores para os critérios de informação

Modelo	AIC	BIC
M1	344,2585	408,7978
M2	316,9712	385,8131

Fonte: Autores

As estimativas para os parâmetros do modelo misto (2) foram obtidas por meio do método da máxima verossimilhança restrita e são apresentadas na Tabela 3.

Tabela 3: Estimativas dos parâmetros de efeitos fixos do modelo misto.

Genótipos	Estimativa	Erro Padrão	G.L.	Estatística-t	valor-p
Intercepto	2,3667	0,0727	490	32,5759	0,0000
G2	-0,2345	0,1027	56	-2,2823	0,0263
G3	0,3440	0,1027	56	3,3480	0,0015
G4	-0,0022	0,1027	56	-0,0219	0,9826
G5	-0,1477	0,1027	56	-1,4380	0,1560
G6	-0,2390	0,1027	56	-2,3261	0,0237
G7	0,1685	0,1027	56	1,6399	0,1066
G8	-0,2017	0,1027	56	-1,9636	0,0546
G9	0,3883	0,1027	56	3,7787	0,0004
G10	-0,1023	0,1027	56	-0,9952	0,3239
G11	0,8252	0,1027	56	8,0318	0,0000
G12	-0,4285	0,1027	56	-4,1704	0,0001
G13	-0,0872	0,1027	56	-0,8492	0,3994
G14	0,7400	0,1027	56	7,2021	0,0000

Fonte: Autores

Considerando o genótipo G1 como *baseline* e comparando-o aos demais, percebe-se que há diferença significativa de diâmetro médio das lesões com os genótipos G2, G3, G6, G9, G11, G12 e G14 ao nível de 5% de significância. As estimativas obtidas indicam que a maior e menor diferença com G1 é em relação aos genótipos G11 e G12, respectivamente. Para os fatores de

efeito aleatório do modelo as estimativas obtidas foram $\mathbf{G} = 0,1273449^2 \mathbf{I}_{(1)}$ e $\mathbf{R} = 0,28532^2 \mathbf{I}_{(560)}$ respectivamente, a variabilidade dos dados explicada pelo fator de efeitos aleatórios folhas, e o erro aleatório.

Como foram identificadas diferenças significativas entre os genótipos, efetuou-se o teste de comparações múltiplas de Scott-Knott, por meio da função SK do pacote `ScottKnott`, buscando identificar quais genótipos diferem entre si. Na Tabela 4, são apresentadas as comparações múltiplas, sendo que letras iguais indicam não haver diferença significativa, ao nível de 5% de significância.

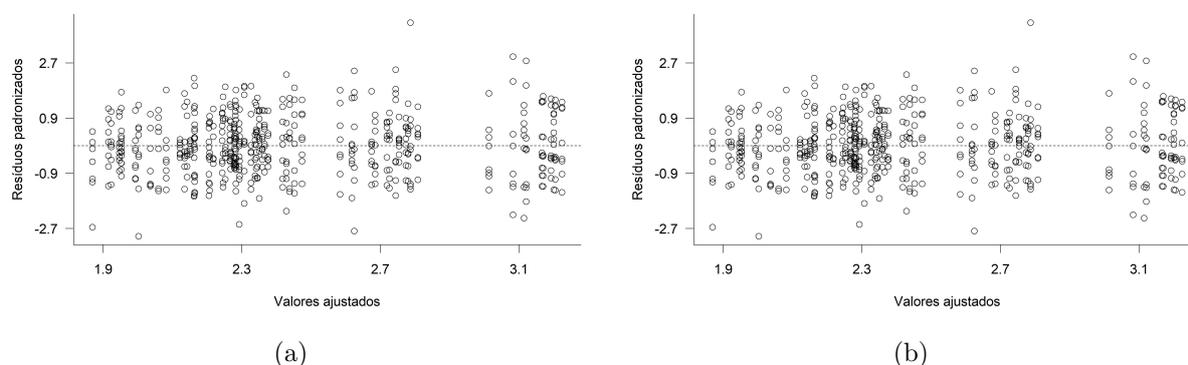
Tabela 4: Teste de comparações de Scott-Knott (5%).

Genótipos	Médias			
G11	3,19	a		
G14	3,11	a		
G9	2,76	b		
G3	2,71	b		
G7	2,54	c		
G1	2,37		d	
G4	2,36		d	
G13	2,28		d	
G10	2,26		d	
G5	2,22			e
G8	2,17			e
G2	2,13			e
G6	2,13			e
G12	1,94			f

Fonte: Autores

A verificação da adequação do modelo pode ser feita mediante análise gráfica dos resíduos. Nas Figuras 3a e 3b são apresentados gráficos de resíduos padronizados para o modelo linear misto.

Figura 3: Gráficos dos resíduos padronizados por genótipo para o modelo de efeitos mistos condicional.

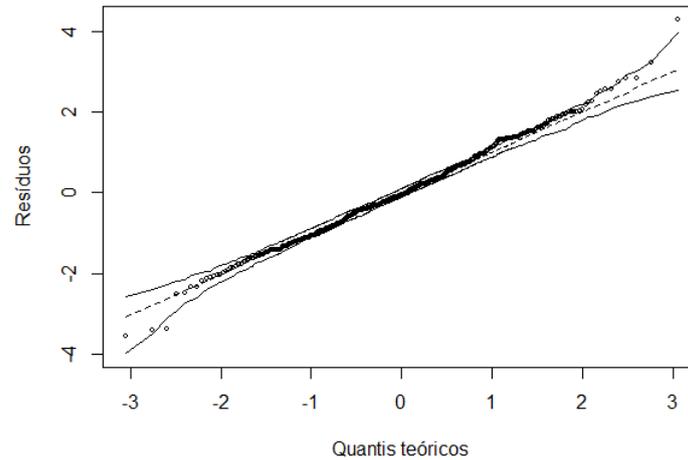


Fonte: Autores

Algumas observações que geraram valores como os apresentados na Figura 3b para os genótipos G3 e G6, foram removidas dos dados, porém, essa remoção não trouxe importantes alterações para as estimativas e assim, preferiu-se mantê-las, garantindo dados balanceados.

Na Figura 4, é apresentado o gráfico de quantis normais com envelope simulado e, embora a distribuição dos resíduos apresente alguns pontos fora das bordas de confiança (5,71%), a maioria deles encontra-se dentro das mesmas, sugerindo a qualidade de ajuste do modelo misto assumido para análise dos dados de genótipos para avaliação da suscetibilidade à doença do cancro cítrico.

Figura 4: Gráficos de quantis normais com envelope simulado para o modelo que considera efeitos aleatórios



Fonte: Autores

Considerações finais

O cenário pelo qual a citricultura brasileira passou e as condições pelas quais a doença do cancro cítrico condiciona a agricultura mundial evidencia a importância no combate a tal patógeno, prejudicial a economia do país e do mundo. A utilização da estatística no ambiente agrícola auxilia o desenvolvimento de métodos mais eficazes para o combate de doenças pertinentes à área.

Constatou-se que os genótipos G11 e G14 são mais suscetíveis a doença do cancro cítrico, enquanto o genótipo G12 se apresenta menos suscetível ao desenvolvimento da doença. Ainda, o modelo linear misto mostrou-se mais adequado ao conjunto de dados de severidade do cancro cítrico em genótipos de cítricos.

A eficácia da metodologia proposta se deve ainda, à utilização dos pacotes e rotinas implementadas no ambiente estatístico *R* (R DEVELOPMENT CORE TEAM, 2022).

Referências

BATES, Douglas; MAECHLER, Martin; BOLKER, Ben; WALKER, Steve. *Fitting Linear Mixed-Effects Models Using lme4*. *Journal of Statistical Software*, v. 67, n. 1, p. 1-48, 2015. DOI: 10.18637/jss.v067.i01.

CAZETTA, Andressa. *Genótipos De Citros Inoculados Com Xanthomonas citri subsp. citri: Um Estudo Sobre Severidade Da Doença E Anatomia Estomática Da Planta*. Dissertação de Mestrado, Universidade Estadual de Maringá, Maringá, 2021.

GONÇALVES-ZULIANI, ALINE MARIA ORBOLATO. *Resistência de genótipos de laranja doce (Citrus sinensis) ao cancro cítrico e diversidade genética de Xanthomonas citri subsp.*

Sigmae, Alfenas, v.12, n.3, p.99-107. 2023.

67ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria (RBras) e 20º Simpósio de Estatística Aplicada à Experimentação Agronômica (SEAGRO)

citri. Dissertação de Mestrado, Universidade Estadual de Maringá, Maringá, 2014.

JELIHOVSCHI, Enio G.; FARIA, Jose C.; ALLAMAN, Ivan Bezerra. *ScottKnott: A Package for Performing the Scott-Knott Clustering Algorithm in R*. *Trends in Applied and Computational Mathematics*, v. 15, n. 1, p. 3-17, 2014. URL: <https://tema.sbmac.org.br/tema/article/view/646/643>.

LONGHINI, Kleber Lopes et al. *Avaliação da incidência e severidade de cancro cítrico com uso de composto químico em plantas de laranja-doce*. In XI Encontro Internacional de Produção Científica, Maringá, PR, 2019.

OLIVEIRA, Roberto Pedroso et al. *Cancro cítrico: epidemiologia e controle*. Embrapa, Pelotas-RS, 2008. Disponível em: <https://shorturl.at/jltKR>. Acesso em: 04/03/2023.

PINHEIRO, José C.; BATES, Douglas M.. *Mixed-effects models in S and S-PLUS*. Springer Science e Business Media, 2000.

PINHEIRO, J.; BATES, D.; R Core Team. *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-162. URL: <https://CRAN.R-project.org/package=nlme>.

R Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Viena, Áustria. URL: <http://www.R-project.org/>.

VOLPATO, Sthefany Caroline. *Comparações de distribuições de probabilidade na análise à resistência ao cancro cítrico*. Dissertação de Mestrado em Ciências. Universidade Estadual de Maringá, Maringá, 2021.

WEST, Brady T.; WELCH, Kathleen B.; GALECKI, Andrzej T.. *Random coefficient models for longitudinal data: the autism example linear mixed models: a practical guide using statistical software*. 2^a ed. Nova York: CRC Press, Taylor e Francis, 2014.