

Universidade Estadual de Maringá

Departamento de Estatística

Disciplina: 8076 – Modelos Mistos

Professora: Dra. Terezinha Aparecida Guedes

Acadêmico: André Felipe Berdusco Menezes

Modelos Mistos: aplicação em dados de frutos de laranja doce

Maringá
Agosto de 2017

Sumário

1	Introdução	2
2	Metodologia	3
3	Resultados e Discussões	6
4	Conclusão	15

1 Introdução

Em diversas áreas, é comum na prática o uso de experimentos longitudinais. Segundo Verbeke e Molenberghs (1997) um experimento onde cada sujeito é observado sob várias condições experimentais, é denominado de medidas repetidas. Por outro lado, os mesmos autores definem dados longitudinais quando o experimento mede repetidamente ao longo do tempo cada indivíduo.

Neste trabalho iremos analisar os dados provenientes de um estudo longitudinal, onde se avaliou em cinco tempos diferentes o diâmetro do pedúnculo da laranja doce sob o efeito de diferentes doses. As avaliações foram realizadas semanalmente aos 0 (avaliação inicial), 7, 14, 21 e 28 dias. Ressalta-se que nos materiais disponibilizados não há menção sobre o significado das doses, isto é, o que foi aplicado. Conforme o pesquisador, responsável pelos dados, nas doses utilizadas, não se esperava diferença significativa, porque as doses foram de regulador hormonal, somente para influenciar a cicatrização e interferir na doença.

Uma vez que a estrutura original dos dados não se adequavam com a metodologia estudada na disciplina Modelos Mistos (8076), tal estrutura foi modificada. Nesse sentido, além da variável resposta (diâmetro), as covariáveis avaliação (tempo em dias) e tratamento (doses) foram medidas em cada árvore (fruto) ao longo do tempo. Logo, o objetivo deste estudo se reduz em comparar o comportamento do diâmetro de cada fruto ao longo do tempo (avaliações) e identificar se existe influência das doses, tempo (avaliações) bem como da interação entre estes fatores.

Além desta introdução o presente trabalho estado organizado da seguinte forma. Na Seção 2 discuto a teoria dos modelos mistos aplicada sob a perspectiva da análise de dados longitudinais, além disso também descrevemos as possíveis estruturas dos modelos para responder os objetivos do pesquisador. Uma análise descritiva do conjunto de dados, resultados da modelagem e discussões são expostas na Seção 3. Por fim, algumas considerações gerais sobre o trabalho finalizam este relatório na Seção 4.

2 Metodologia

Esta seção é dedicada a um breve apanhado sobre a teoria dos modelos mistos. Maiores detalhes sobre os modelos mistos podem ser encontrados em: Pinheiro (1994) Verbeke e Molenberghs (1997), Pinheiro e Bates (2000) Verbeke e Molenberghs (2000), Diggle et al. (2002), Littell et al. (2006) Bates (2010), Demidenko (2013) e Singer, Nobre e Rocha (2017).

Os modelos lineares mistos são definidos, em uma forma compacta, da seguinte forma:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e} \quad (1)$$

em que:

- \mathbf{y} é um vetor $n \times 1$ das variáveis;
- \mathbf{X} é uma matriz $n \times p$ de especificação dos efeitos fixos;
- $\boldsymbol{\beta}$ é um vetor $p \times 1$ de parâmetros dos efeitos fixos;
- $\mathbf{Z} = [\mathbf{Z}_1, \dots, \mathbf{Z}_b]$, onde \mathbf{Z}_i é uma matriz $n \times q_i$ de especificação para o i -ésimo efeito aleatório;
- \mathbf{u} é um vetor $q \times 1$ de efeitos aleatórios, onde \mathbf{u}_i é outro vetor $q_i \times 1$ tal que $q = \sum_{i=1}^b q_i$, com $\mathbb{E}(\mathbf{u}) = \mathbf{0}$;
- \mathbf{e} é um vetor $n \times 1$ de erros aleatórios, com $\mathbb{E}(\mathbf{e}) = \mathbf{0}$.

Sob esta formulação assumimos que \mathbf{u} e \mathbf{e} são independentes e seguem distribuição Normal multivariada de modo que:

$$\begin{bmatrix} \mathbf{u} \\ \mathbf{e} \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \sigma^2 \begin{bmatrix} \mathbf{G}(\boldsymbol{\gamma}) & \mathbf{0} \\ \mathbf{0} & \mathbf{R}(\boldsymbol{\rho}) \end{bmatrix} \right) \quad (2)$$

em que $\boldsymbol{\gamma}$ e $\boldsymbol{\rho}$ são vetores $r \times 1$ e $s \times 1$ (com $s \leq n(n+1)/2$) desconhecidos dos parâmetros de variâncias correspondentes a \mathbf{u} e \mathbf{e} , respectivamente. Conforme Verbeke e Molenberghs (2000) escrevemos a matriz de variância-covariância dos dados, \mathbf{y} , como:

$$\text{Var}(\mathbf{y}) = \sigma^2 (\mathbf{Z}\mathbf{G}\mathbf{Z}^\top + \mathbf{R}) = \sigma^2 \mathbf{H} \quad (3)$$

A matriz \mathbf{H} consiste de duas componentes que são utilizadas para modelar a heterocedasticidade e correlação, sendo elas: um componente de efeito aleatório $\mathbf{Z}\mathbf{G}\mathbf{Z}$ e um componente dentro do grupo \mathbf{R} . Ressalta-se aqui, que a componente intra grupo, \mathbf{R} , em algumas aplicações, é utilizada diretamente para modelar a matriz de variância-covariância,

sem a necessidade de incorporar efeitos aleatórios no modelo para explicar a dependência entre as observações.

Como discutido em Singer, Nobre e Rocha (2017) grande parte do esforço empregado na modelagem de dados com medidas repetidas se concentra na estrutura de covariância. No contexto dos modelos mistos, a covariância entre as observações obtidas em uma mesma unidade amostral poderá ser modelada diretamente por meio da matriz \mathbf{R} . Além disso, Diggle et al. (2002) ressalta que os modelos para a matriz \mathbf{R} podem incluir três possíveis fontes de variação na estrutura de covariância: aquelas correspondente aos efeitos aleatórios, à correlação serial e aos erros de medida. Aqui deve ser ressaltado que para os dados em estudo iremos considerar as estruturas *Compound Symmetry*, Exponencial e Gaussiana. Para informações sobre outras estruturas o leitor pode consultar as referencias citadas anteriormente ou ainda o manual da PROC MIXED do SAS (SAS Institute Inc., 2011).

Dado o modelo formulado (1) é necessário estimar seus parâmetros. Neste caso, temos parâmetro fixos, β , e os parâmetros de variâncias dos efeitos aleatórios \mathbf{u} e do erro aleatório \mathbf{e} . Diferente métodos de estimação foram propostos, mas seguramente os mais utilizados são os métodos da máxima verossimilhança (ML) e máxima verossimilhança restrita (RML).

Para obter as estimativas das matrizes \mathbf{G} e \mathbf{R} maximizamos as funções associada a ML e RML. As correspondentes funções de log-verossimilhanças são definidas, respectivamente, por:

$$\begin{aligned} \text{ML:} \quad \ell(\mathbf{G}, \mathbf{R}) &= -\frac{1}{2} \log |\mathbf{V}| - \frac{1}{2} \mathbf{r}^\top \mathbf{V}^{-1} \mathbf{r} - \frac{n}{2} \log(2\pi) \\ \text{RML:} \quad \ell_R(\mathbf{G}, \mathbf{R}) &= -\frac{1}{2} \log |\mathbf{V}| - \frac{1}{2} \log |\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X}| - \frac{1}{2} \mathbf{r}^\top \mathbf{V}^{-1} \mathbf{r} - \frac{n-p}{2} \log(2\pi) \end{aligned}$$

em que $\mathbf{r} = \mathbf{y} - \mathbf{X} (\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{V}^{-1} \mathbf{y}$ e p é o posto de \mathbf{X} .

Para obter as estimativas dos efeitos fixos β e as predições dos efeitos aleatórios γ o método padrão é resolver as *equações de modelos mistos* definidas por Henderson (1984):

$$\begin{bmatrix} \mathbf{X}^\top \hat{\mathbf{R}}^{-1} \mathbf{X} & \mathbf{X}^\top \hat{\mathbf{R}}^{-1} \mathbf{Z} \\ \mathbf{Z}^\top \hat{\mathbf{R}}^{-1} \mathbf{X} & \mathbf{Z}^\top \hat{\mathbf{R}}^{-1} \mathbf{Z} + \hat{\mathbf{G}}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{\gamma} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^\top \hat{\mathbf{R}}^{-1} \mathbf{y} \\ \mathbf{Z}^\top \hat{\mathbf{R}}^{-1} \mathbf{y} \end{bmatrix} \quad (4)$$

As soluções podem ser escritas como:

$$\begin{aligned} \hat{\beta} &= \left(\mathbf{X}^\top \hat{\mathbf{V}}^{-1} \mathbf{X} \right)^{-1} \mathbf{X}^\top \hat{\mathbf{V}}^{-1} \mathbf{y} \\ \hat{\gamma} &= \hat{\mathbf{G}} \mathbf{Z}^\top \hat{\mathbf{V}}^{-1} (\mathbf{y} - \mathbf{X} \hat{\beta}) \end{aligned} \quad (5)$$

Inferências sob os parâmetros do modelo (1) são realizadas sob o paradigma da teoria da verossimilhança, em particular utilizando o fato de que os estimadores de máxima verossimilhança são assintoticamente normais, com matriz de variância-covariância dada pela inversa da informação de Fisher. Para comparação de modelos encaixados procedeu-se com o teste da razão de verossimilhança, a qual sua estatística é definida por:

$$S_{LR} = 2(\ell_c - \ell_r) \quad (6)$$

em que ℓ_r e ℓ_c são os valores das funções de log-verossimilhanças avaliadas sob hipótese nula e alternativa, respectivamente. A estatística S_{LR} tem distribuição assintótica qui-quadrado com p graus de liberdade, sendo p o número de restrições impostas.

Por outro lado, para comparar as estruturas de covariância os critérios de informação de Akaike (AIC) e Bayesiano (BIC) foram utilizados. Eles são definidos, respectivamente, por:

$$\text{AIC} = -2\log(L) + \frac{2np}{n - p - 1} \quad \text{e} \quad \text{BIC} = -2\log(L) + p\log(n) \quad (7)$$

em que L é o valor da função de verossimilhança avaliada nas estimativas de máxima verossimilhança, n é o número de observações e p é o número de parâmetros estimados.

Por fim, ressalta-se que toda a análise foi conduzida no software R (R Core Team, 2016) com auxílio das bibliotecas `lme4` (BATES et al., 2015) e `nlme` (PINHEIRO et al., 2016) para ajustar os modelos, comparação dos mesmos e análise de resíduos.

3 Resultados e Discussões

Como apontado por Bates (2010) uma representação gráfica interessante para avaliar em dados longitudinais é verificar se existe alguma tendência da variável resposta ao longo do tempo conforme o sujeito. No entanto para os dados considerados neste trabalho este gráfico não forneceu informações relevantes, uma vez que 80 sujeitos foram observados ao longo das cinco avaliações. Nesse sentido, a análise descritiva ficou restrita a verificar o comportamento da variável resposta de acordo com os níveis do tratamento e ao longo do tempo. Ou seja, sem considerar a trajetória do indivíduo ao longo do tempo.

Na Figura 1 é apresentada o comportamento da variável resposta conforme as doses observadas. Os boxplots indicam que a variabilidade entre os tratamentos, bem como a mediana são muito similares.

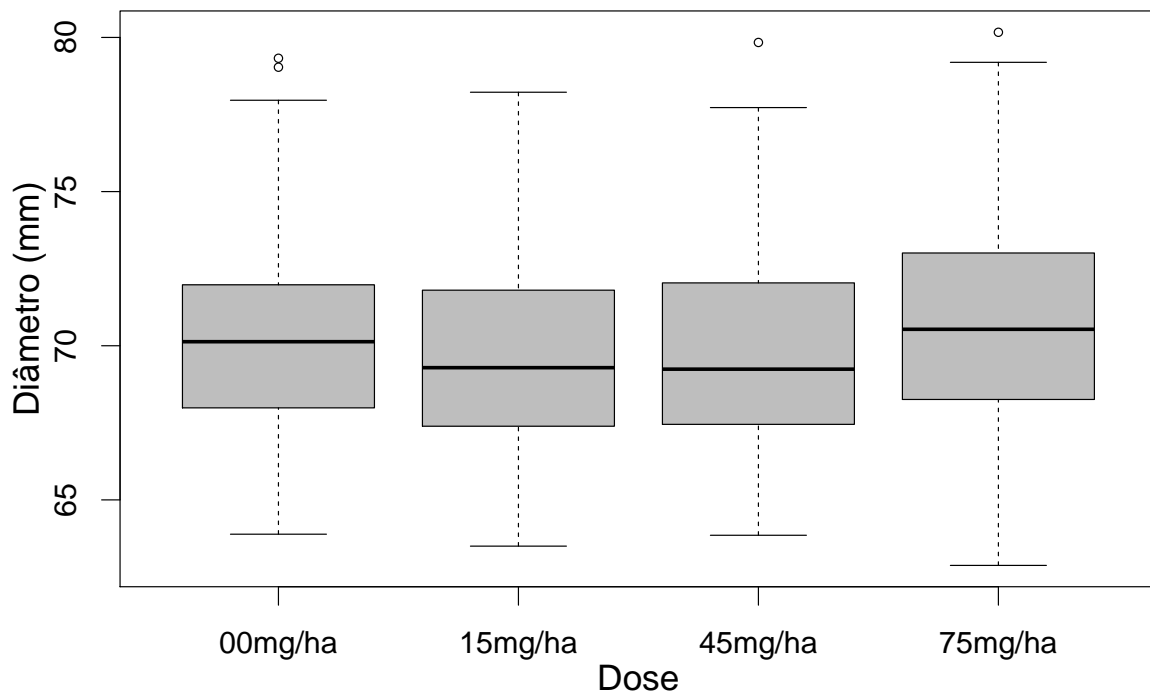


Figura 1: Boxplots do diâmetro conforme as doses aplicadas.

Por outro lado, o comportamento do diâmetro ao longo do tempo é representado na Figura 2. Inicialmente, nota-se diferenças entre os boxplots, tanto em posição (mediana) quanto em dispersão (amplitudes). Curiosamente, não existe uma tendência de crescimento ou decrescimento ao longo do tempo da variável resposta diâmetro do fruto. Os dados indicam que a maior média do diâmetro ocorre após 21 semana, por outro lado, as menores médias são na semana 7 e 28. Embora exista diferenças entre os tempos, este fato comprova a falta de tendência ao longo do tempo.

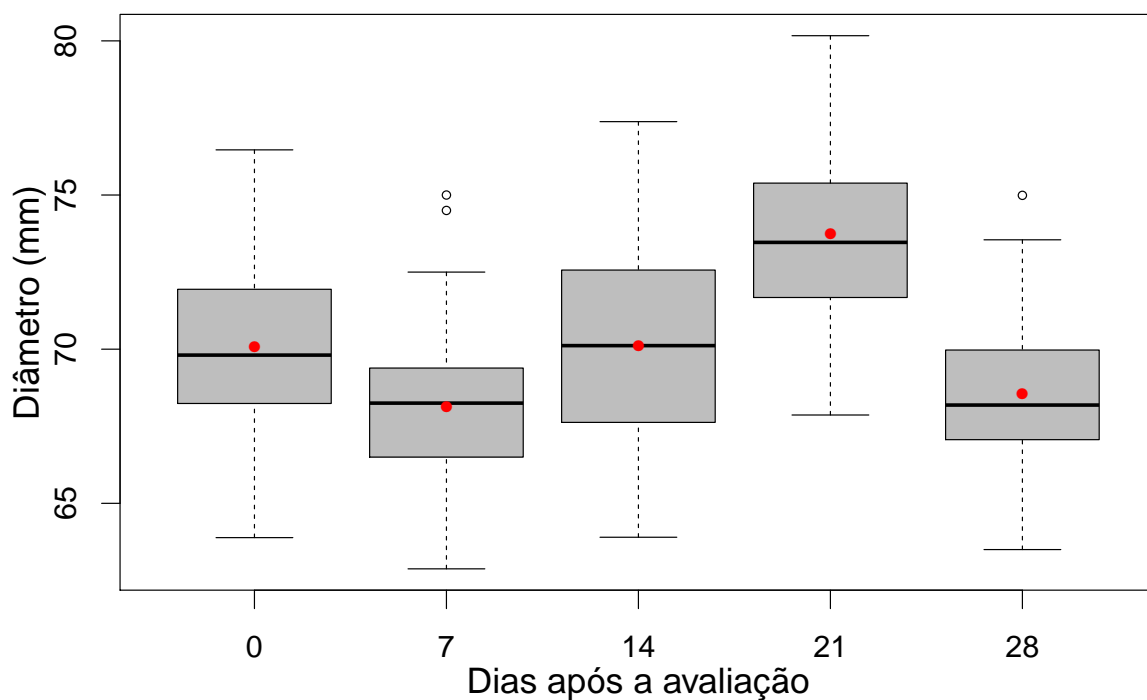


Figura 2: Boxplots do diâmetro conforme as avaliações (tempo).

Finalizando a análise descritiva, apresentamos na Figura 3 o gráfico da interação entre as avaliações (tempo) e as doses aplicadas. Pode-se notar que não existe fortes diferenças entre os boxplots dentro de cada tempo, por exemplo, fica evidente que 7 dias após a avaliação o comportamento dos boxplots são bastante similares. Outra vez, pode-se observar pela linha tracejada e os pontos, a falta de tendência da variável resposta ao longo do tempo, bem como a diferença da mesma a medida que as avaliações ocorrem.

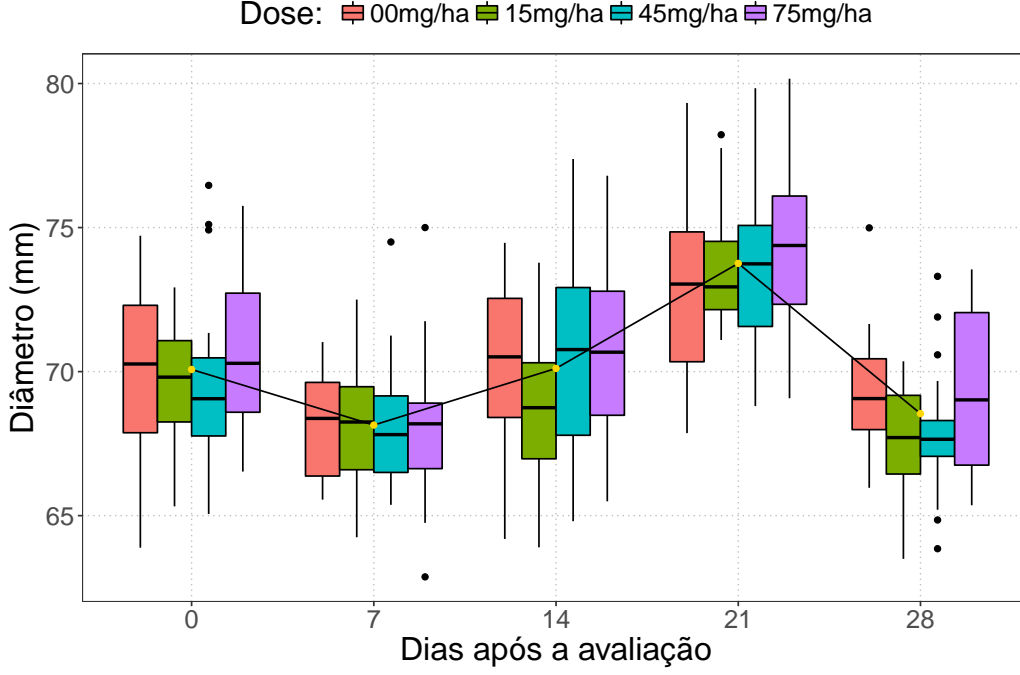


Figura 3: Boxplots do diâmetro conforme as avaliações (tempo) e as doses aplicadas.

Considerando a descrição dos dados exposta e a teoria dos modelos mistos discutido nas seções anteriores foi considerado um modelo misto afim de explicar a variabilidade dos dados e responder as perguntas do pesquisador. Algumas questões de interesse, entre outras, são: existe diferença entre as doses? O diâmetro do fruto sofre influencia do tempo? Existe alguma combinação entre tempo e tratamento que fornece maior(menor) resposta em média. Deve ser mencionado, que algumas dessas indagações podem ser vistas de forma empírica pela análise descritiva aqui apresentada. No entanto, um modelo estatístico que considerar erros aleatórios deve ser empregado afim de formalizar os resultados expostos na análise exploratória.

Seja Y_i o valor do diâmetro da i -ésima árvore o modelo considerado inicialmente pode ser escrito por:

$$Y_i = \mu + \alpha(\text{dose}_i) + \beta(\text{tempo}_i) + \gamma(\text{tempo}_i \text{dose}_i) + \delta(\text{arvore}_i) + \varepsilon_i \quad (8)$$

em que $i = 1, \dots, 400$, $\delta(\text{arvore}_i) \sim \mathcal{N}(0, \sigma_\delta^2)$ e $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ e todos δ 's e ε 's são independentes. Ressalta-se que o tempo (avaliação) foi considerada como covariável e não fator.

A estrutura de covariância deste modelo pode ser expressa por:

$$\text{Cov}(y_i, y_j) = \begin{cases} 0 & , \text{ se } \text{arvore}_i \neq \text{arvore}_j \text{ e } i \neq j \\ \sigma_\delta^2 & , \text{ se } \text{arvore}_i = \text{arvore}_j \text{ e } i \neq j \\ \sigma_\delta^2 + \sigma^2 & , \text{ se } i = j \end{cases}$$

Esta estrutura de variância, sugere que duas observações de arvores diferentes não estão correlacionadas, e duas observações da mesma arvores estão correlacionadas positivamente com o coeficiente de correlação dado por:

$$\rho = \frac{\sigma_{\delta}^2}{\sigma_{\delta}^2 + \sigma^2} \quad (9)$$

Os resultados da ANOVA para os efeitos fixos do modelo (8) são apresentados na Tabela 1. Pelos resultados da Tabela 1 observa-se que o efeito da dose e a interação entre dose e tempo (avaliação) não foi significativo. Esse resultado já era esperado, pois como apresentado nas análises descritivas, as doses assim como sua interação com o tempo não proporcionou grupos com fortes diferenças (ver Figura 1 e Figura 3).

Tabela 1: Resultados da ANOVA para os efeitos fixos.

Parâmetro	G.L. num	G.L. den	F	valor-p
Intercepto	1	316	119708.2262	0.0000
Dose	3	76	1.2644	0.2927
Tempo	1	316	5.4009	0.0208
Dose:Tempo	3	316	0.2474	0.8631

Como a interação não foi significativa, ela será retirada do modelo (8). Os resultados da ANOVA sem a interação estão expostos na Tabela 2. Pode-se notar que mesmo sem a interação o efeito da dose sobre o diâmetro do fruto de laranja não é significativo, em outras palavras dizemos que o comportamento do diâmetro da laranja, neste estudo, é independente da dose aplicada.

Tabela 2: Resultados da ANOVA para os efeitos fixos.

Parâmetro	G.L. num	G.L. den	F	valor-p
Intercepto	1	319	119708.2355	0.0000
Dose	3	76	1.2644	0.2927
Tempo	1	319	5.4394	0.0203

Tendo em vista os resultados discutidos e apresentados nas Tabelas 1 e 2 o modelo proposto se reduz a um modelo de regressão linear simples com efeito aleatório do sujeito (árvore). Sua especificação é dada por:

$$Y_i = \beta_0 + \beta_1 \text{tempo}_i + \delta(\text{arvore}_i) + \varepsilon_i \quad (10)$$

em que β_0 é o intercepto, β_1 o coeficiente angular devido ao tempo (avaliações), $\delta(\text{arvore})$

o efeito aleatório devido ao sujeito e ε_i são os erros aleatórios. Além disso, assumimos que $\delta(\text{arvore}_i) \sim \mathcal{N}(0, \sigma_\delta^2)$ e $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ e todos δ 's e ε 's são independentes.

A estrutura de covariância do modelo (10) é a mesma discutida anteriormente, ela é denominada na literatura como *Compound Symmetry* (VERBEKE; MOLENBERGHS, 1997). Nesta estrutura temos que duas medidas de um mesmo individuo estão correlacionadas não importa a distância (em tempo ou espaço) que elas foram medidas. Para os dados em estudo, uma estrutura intuitiva para a covariância dos erros é aquele onde a correlação entre duas medidas de um mesmo individuo (arvore) dependem de quão distante elas foram tomadas. Neste sentido, foi outras estruturas, denominadas de Exponencial e Gaussiana, também foram consideradas. Maiores detalhes sobre estas estruturas de correlação podem ser encontra em Diggle et al. (2002) e Verbeke e Molenberghs (1997).

Para comparar as estruturas *Compound Symmetry* (CS), Exponencial (Exp) e Gaussiana (Gaus) os critérios de informação AIC e BIC foram utilizados. Uma vez que modelos com as estruturas Exp e Gaus são modelos aninhados com a estrutura CS, o teste da razão de verossimilhança foi utilizado nestes casos.

Os critérios utilizados para comparar as estruturas de correção estão apresentados na Tabela 3. Os resultados indicam que não diferenças significativas entre as estruturas CS, Exp e Gaus. Na realidade, com base nos valores de AIC e BIC verifica-se que a estrutura CS é a mais adequada. Este fato é verificado também por meio do teste da razão de verossimilhança, o qual verifica se o parâmetro adicional das estruturas Exp e Gaus não são significativos para o modelo.

Tabela 3: Comparação entre as estruturas de correlação.

Estrutura	G.L.	AIC	BIC
CS	5	2097.613	2117.545
Exp	6	2099.613	2123.532
Gaus	6	2099.613	2123.532
Comparação		TRV	valor- p
CS vs Exp		$3.231 \cdot 10^{-7}$	0.9995
CS vs Gaus		$3.231 \cdot 10^{-7}$	0.9995

Tendo em vista os resultados discutidos o modelo especificado em (10) foi escolhido. As estimativas dos parâmetros e seus respectivos erros padrão estão apresentados na Tabela 4. No que se refere ao parâmetros referente a covariável tempo nota-se que sua estimativa foi de 0.0364 com intervalo de confiança variando entre 0.0058 e 0.0670, isto indica que a cada 7 dias após a avaliação espera-se um aumento entre 0.0058 a 0.0670 no diâmetro do fruto.

Em relação aos parâmetros σ^2 e σ_δ^2 referente na Tabela 4 estão apresentados suas estimativas e os intervalos de confiança baseados na função de verossimilhança perfilada. Dada as estimativas de σ^2 e σ_δ^2 podemos verificar que o coeficiente de correlação entre duas observações da mesma árvore é 12.87%.

Tabela 4: Resumo do ajuste considerando o modelo (10).

Parâmetro	Estimativa	EP	LI	LS	t	$P(T > t)$
β_0	69.6161	0.2987	69.0308	70.2013	233.0742	<0.0001
β_1	0.0364	0.0156	0.0058	0.0670	2.3322	0.0203
σ^2	9.5441	—	8.1807	11.1549	—	—
σ_δ^2	1.4106	—	0.4733	2.6686	—	—

Com o intuito de realizar uma critica sobre o modelo ajustado, isto uma análise de resíduo e diagnóstico, algumas ferramentas gráficas foram utilizadas para detectar possíveis afastamentos das suposições feitas pelo modelo definido em (10).

Na Figura 4 temos o conhecido *half-normal plot* para os resíduos marginais, note que nenhuma observação esteve fora do envelope simulado o que por diversos autores caracteriza um ajuste satisfatório do modelo de regressão estabelecido aos dados em estudo.

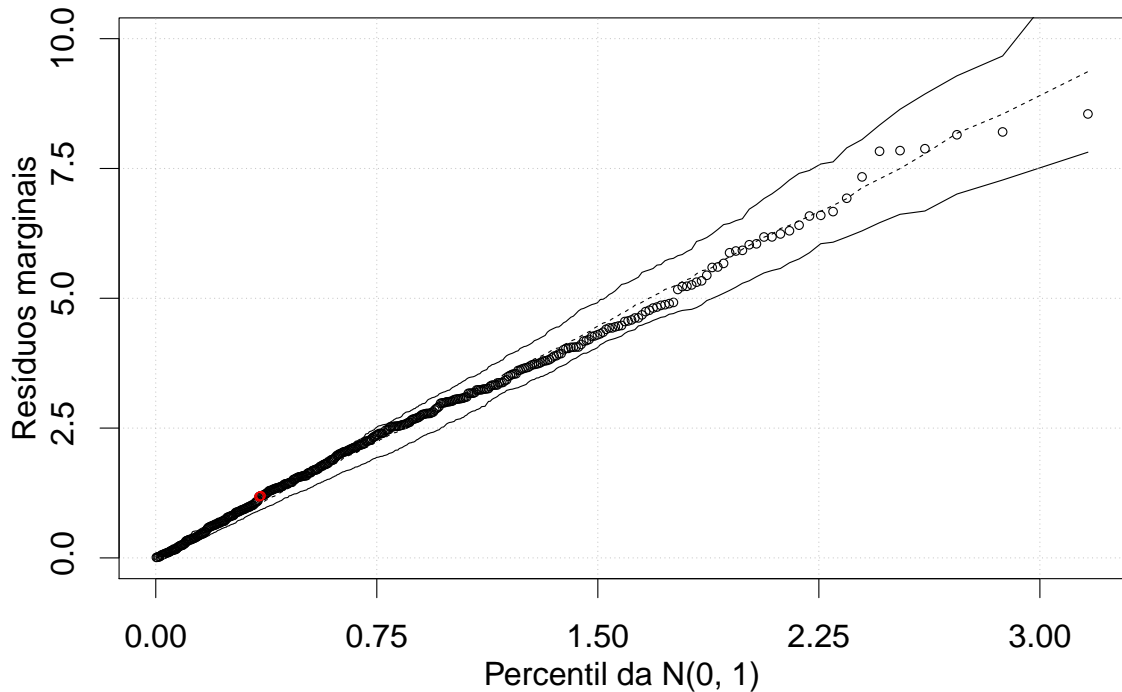


Figura 4: Gráfico meio normal de probabilidades com envelope simulado para os resíduos marginais.

A Figura 5 apresenta o gráfico dos resíduos marginais versus valores preditos. Verifica-se um comportamento aleatório em torno de zero, o que caracteriza um resíduo bem comportado.

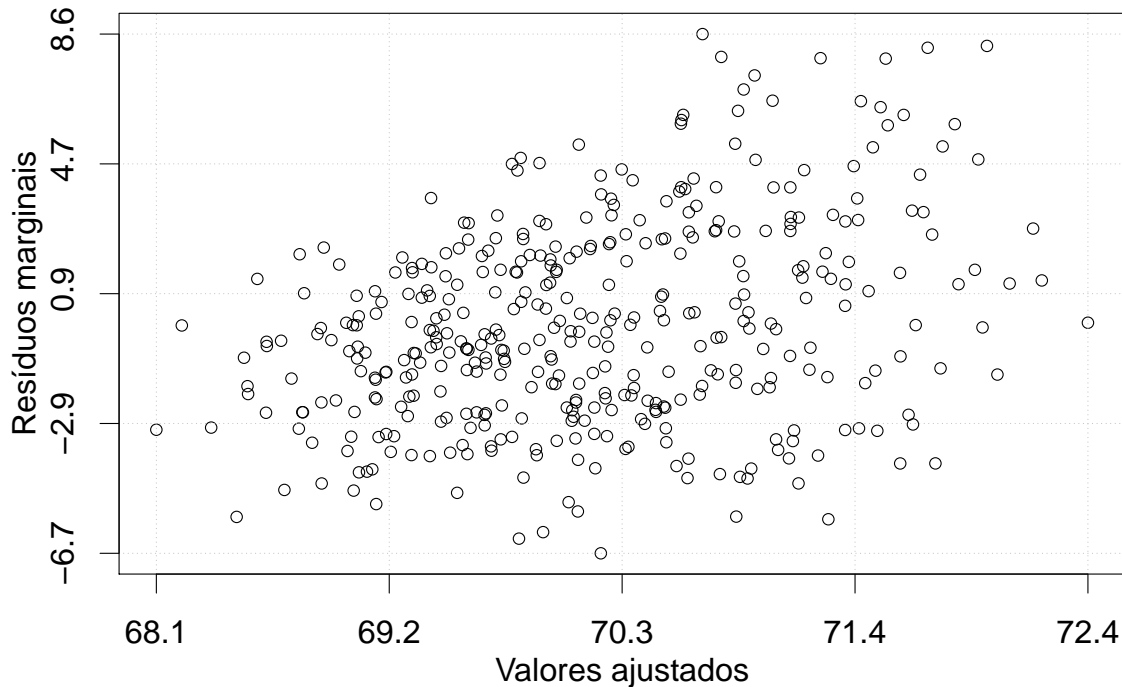


Figura 5: Gráfico dos resíduos marginais versus valores preditos.

No modelo (10) assumimos que o efeito da árvore segue uma distribuição normal. Assim, devemos investigar os BLUPs do efeito aleatório devido a árvore. Nesse sentido apresentamos a Figura 6, a qual compara os resíduos de efeito aleatório da árvore com o percentil teórico da distribuição Normal. Em geral, pode-se notar que não fortes afastamentos da distribuição dos efeitos aleatórios.

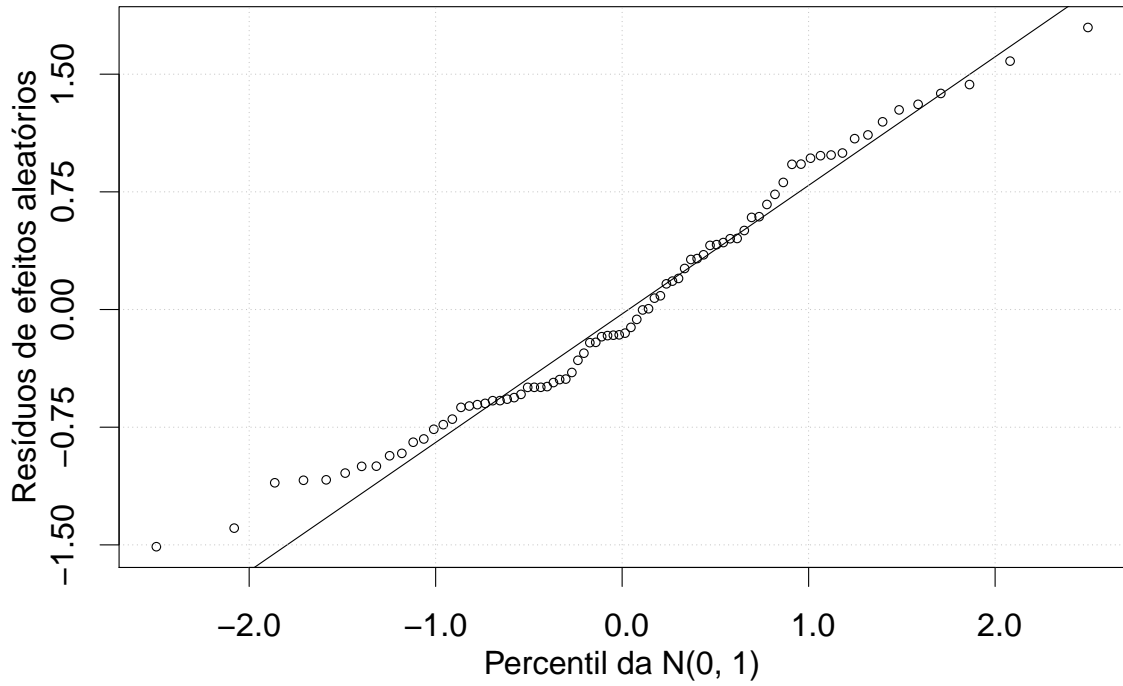


Figura 6: QQ-Plot para os resíduos de efeito aleatório da árvore.

Finalizando a análise de apresentamos a Figura 7 na qual apresenta o gráfico da distância de Cook versus os índices das observações. Nota-se que somente uma observação se destaca das demais com uma distância de Cook próxima de 0.015. Todavia, conforme critério proposto por Cook e Weisberg (1982) não temos problemas de observações influentes, uma vez que nenhuma observação apresentou distância maior que 1.

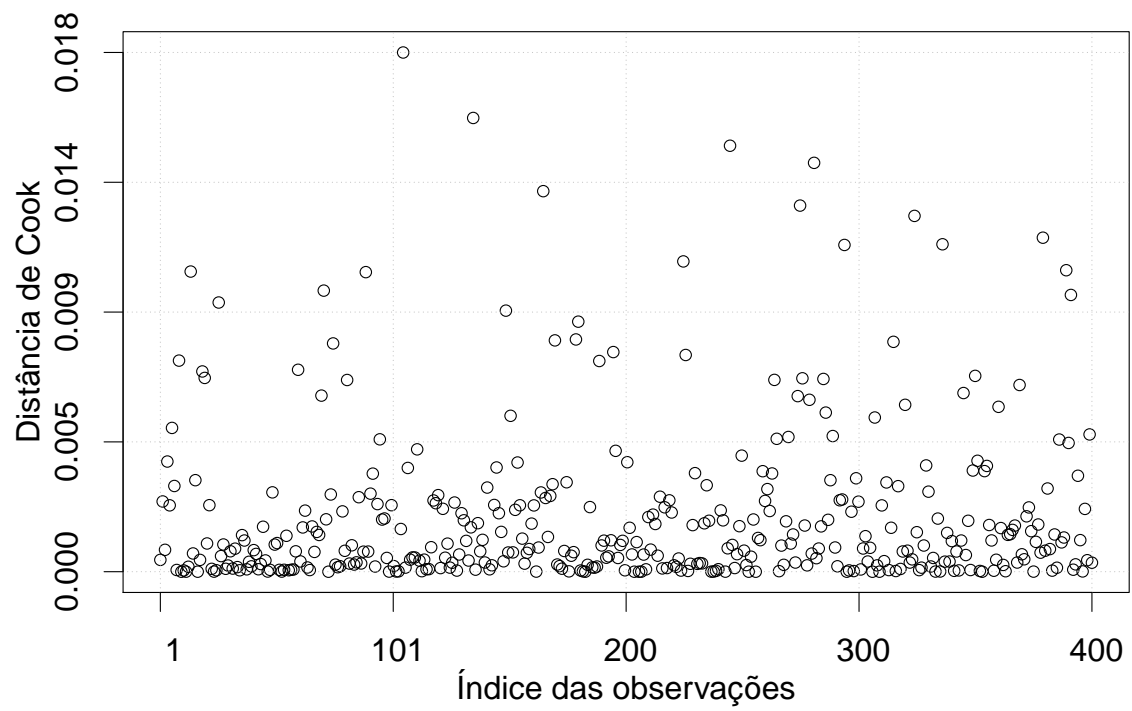


Figura 7: Gráfico da distância de Cook versus as observações.

4 Conclusão

Diante dos resultados apresentados e discutidos neste relatório conclui-se que:

- (i) não há efeito de tratamento sob o diâmetro do fruto de laranja, fato este verificado na análise descritiva e pelo modelo ajustado;
- (ii) após extensa comparação entre as estruturas de correção a *Compound Symmetry* se mostrou a mais adequada conforme o teste da razão de verossimilhanças e os critérios AIC e BIC;
- (iii) o modelo escolhido considerou apenas o tempo (avaliação) como significativo, sendo constatado que a cada 7 dias após a avaliação espera-se um aumento entre 0.0058 a 0.0670 no diâmetro do fruto;
- (iv) o modelo escolhido apresentou ajuste satisfatório, conforme foi verificado pela análise de resíduos (ver Figuras 4, 5, 6 e 7) .

Finalmente, deve-se destacar que os resultados apresentados na modelagem já haviam sido observado na análise exploratório dos dados (ver Figuras 1, 2 e 3). Portanto, este trabalho mostrou a importância da análise descritiva preliminar a metodologia estatística mais sofisticada.

Referências

- BATES, D.; MÄCHLER, M.; BOLKER, B.; WALKER, S. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, v. 67, n. 1, p. 1–48, 2015.
- BATES, D. M. *lme4: Mixed-effects modeling with R*. Springer, 2010. Disponível em: [<http://lme4.r-forge.r-project.org/book/>](http://lme4.r-forge.r-project.org/book/).
- COOK, R. D.; WEISBERG, S. *Residuals and influence in regression*. Chapman and Hall, 1982.
- DEMIDENKO, E. *Mixed Models Theory and Applications with R*. John Wiley & Sons, Inc., 2013.
- DIGGLE, P. J.; HEAGERTY, P. J.; LIANG, K.; L., Z. S. *Analysis of Longitudinal Data*. Oxford University Press, 2002.
- HENDERSON, C. R. *Applications of Linear Models in Animal Breeding*. University of Guelph, 1984.
- LITTELL, R. C.; MILLIKEN, G. A.; STROUP, W. W.; WOLFINGER, R. D.; SCHABENBERGER, O. *SAS[®] for Mixed Models*. Second. Cary, NC: SAS Institute Inc., 2006.
- PINHEIRO, J.; BATES, D.; DEBROY, S.; SARKAR, D.; R Core Team. *nlme: Linear and Nonlinear Mixed Effects Models*. 2016. R package version 3.1-128. Disponível em: [<http://CRAN.R-project.org/package=nlme>](http://CRAN.R-project.org/package=nlme).
- PINHEIRO, J. C. *Topics in Mixed Effects Models*. Tese (Doutorado) — University of Wisconsin – Madison, 1994.
- PINHEIRO, J. C.; BATES, D. M. *Mixed Effects Models in S and S-Plus*. Springer, 2000.
- R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2016. ISBN 3-900051-07-0.
- SAS Institute Inc. *The MIXED Procedure, SAS[®]/STAT User's Guide, Version 9.3*. Cary, NC: SAS Institute Inc., 2011. 4718–4909 p. 2605–2804.
- SINGER, J. M.; NOBRE, J. S.; ROCHA, F. M. M. *Análise de dados longitudinais*. Departamento de Estatística – Universidade de São Paulo, 2017.
- VERBEKE, G.; MOLENBERGHS, G. *Linear Mixed Models in Practice A SAS-Oriented Approach*. Springer, 1997.
- VERBEKE, G.; MOLENBERGHS, G. *Linear Mixed Models for Longitudinal Data*. Springer, 2000.

Apêndice

Neste apêndice disponibilizo os códigos em R utilizados para a análise dos dados.

```
1 # Definicoes gerais -----
2 rm(list = ls(all.names = TRUE))
3 bib <- c('lme4', 'lmerTest', 'lsmeans', 'hnp', 'dplyr', 'ggplot2', 'RLRsim', 'nlme', '
  xtable', 'influence.ME')
4 sapply(bib, require, character.only = T)
5 dados <- read.table(file = 'planta-final.txt', sep = ',', header = T)
6 dados$tempo_f <- factor(dados$tempo_f)
7 dados$arvore <- factor(dados$arvore)
8 head(dados)
9 str(dados)
10
11 # Descritiva -----
12
13 boxplot(diametro ~ trat, data = dados, xlab = '', ylab = '', cex = 0.6, col = 'gray')
14 mtext("Dose", side = 1, line = 2.0, cex = 1.8)
15 mtext("Diametro (mm)", side = 2, line = 2, cex = 1.8)
16
17 boxplot(diametro ~ tempo_f, data = dados, xlab = '', ylab = '', cex = 0.6, col = 'gray')
18 points(x = unique(dados$tempo_f), y = tapply(dados$diametro, dados$tempo_f, mean), pch =
  16, col = 'red', cex = 0.8)
19 mtext("Dias apos a avaliacao", side = 1, line = 2.0, cex = 1.8)
20 mtext("Diametro (mm)", side = 2, line = 2, cex = 1.8)
21
22 dados %>% ggplot(aes(x = tempo_f, y = diametro, group = interaction(tempo_f, trat))) +
23 geom_boxplot(aes(fill = factor(trat)), color = 'black') +
24 stat_summary(aes(group = 1), fun.y = mean, geom="line", color = 'black') +
25 stat_summary(aes(group = 1), fun.y = mean, geom="point", color = 'gold') +
26 labs(y = 'Diametro (mm)', x = 'Dias apos a avaliacao', fill = 'Dose:') +
27 theme_bw() +
28 theme(text = element_text(size=20), panel.grid.minor = element_blank(), legend.position="
  top",
29 panel.grid.major = element_line(size = 0.4, linetype = 'dotted', colour = 'gray'))
30
31 # Ajuste dos modelos -----
32
33 mod1 <- lme(fixed = diametro ~ trat + tempo_n + trat*tempo_n, data = dados, random = ~ 1
  | arvore)
34 mod2 <- lme(fixed = diametro ~ trat + tempo_n, data = dados, random = ~ 1 | arvore)
35 mod3 <- lme(fixed = diametro ~ tempo_n, data = dados, random = ~ 1 | arvore)
36 print(xtable(anova(mod2), digits = 4))
37
38 # Comparacoes das estrutura de correlacao -----
39
40 m.CS <- lme(fixed = diametro ~ trat + tempo_n + trat * tempo_n, data = dados, random = ~
  1 | arvore,
41 correlation = corCompSymm(form = ~tempo_n|arvore))
42 m.Exp <- lme(fixed = diametro ~ trat + tempo_n + trat * tempo_n, data = dados, random = ~
  1 | arvore,
43 correlation = corExp(form = ~tempo_n|arvore, nugget = T))
44 m.Gaus <- lme(fixed = diametro ~ trat + tempo_n + trat * tempo_n, data = dados, random =
  ~ 1 | arvore,
45 correlation = corGaus(form = ~tempo_n|arvore, nugget = T))
46 anova(m.CS, m.Exp)
47 anova(m.CS, m.Gaus)
48 anova(m.Exp, m.Gaus)
```

```

49 # Modelo escolhido -----
50
51 mod3 <- lme(fixed = diametro ~ tempo_n, data = dados, random = ~ 1 | arvore)
52 summary(mod3)
53 intervals(mod3)
54 mod3 <- lmer(diametro ~ tempo_n + (1 | arvore), data = dados)
55 summary(mod3)
56 confint(mod3)
57
58 # Analise de Residuos -----
59
60 ## Residuos marginais (erro aleatorio)
61 my.hnp <- hnp(mod3, halfnormal = T, how.many.out = T, paint.out = T, plot = T)
62 plot(my.hnp, xaxt = 'n', yaxt = 'n', xlab = '', ylab = '', cex = 0.6, ylim = c(0, 10))
63 mtext("Percentil da N(0, 1)", side = 1, line = 2.0, cex = 1.8)
64 mtext("Residuos marginais", side = 2, line = 2, cex = 1.8)
65 abline(h = seq(0, 10, l = 5), v = seq(0, 3, l = 5), col = "gray", lty = "dotted")
66 axis(1, seq(0, 3, l = 5))
67 axis(2, seq(0, 10, l = 5), FF(seq(0, 10, l = 5), 1))
68
69 ## Residuos de efeitos aleatorios
70 r2 <- random.effects(mod3)$arvore
71 qqnorm(r2[, 1], xaxt = 'n', yaxt = 'n', xlab = '', ylab = '', cex = 0.6, main = "");
72 qqline(r2[, 1])
73 mtext("Percentil da N(0, 1)", side = 1, line = 2.0, cex = 1.8)
74 mtext("Residuos de efeitos aleatorios", side = 2, line = 2, cex = 1.8)
75 abline(h = seq(-1.5, 1.5, l = 5), v = seq(-2, 2, l = 5), col = "gray", lty = "dotted")
76 axis(2, seq(-1.5, 1.5, l = 5))
77 axis(1, seq(-2, 2, l = 5), FF(seq(-2, 2, l = 5), 1))
78
79 ## Ajustado versus residuo
80 x = fitted(mod3); y = residuals(mod3); Rx = range(x); Ry = range(y)
81 plot(y ~ x, xlab = '', ylab = '', cex = 0.8, xaxt = 'n', yaxt = 'n')
82 mtext("Valores ajustados", side = 1, line = 2.0, cex = 1.8)
83 mtext("Residuos marginais", side = 2, line = 2, cex = 1.8)
84 abline(h = seq(Ry[1], Ry[2], l = 5), v = seq(Rx[1], Rx[2], l = 5), col = "gray", lty = "dotted")
85 axis(2, seq(Ry[1], Ry[2], l = 5), FF(seq(Ry[1], Ry[2], l = 5), 1))
86 axis(1, seq(Rx[1], Rx[2], l = 5), FF(seq(Rx[1], Rx[2], l = 5), 1))
87
88 ## Influencia
89 lmer3.infl <- influence(mod3, obs=TRUE)
90 cook <- cooks.distance(lmer3.infl)
91 x = 1:nrow(dados); y = cooks.distance(lmer3.infl); Rx = range(x); Ry = range(y)
92 plot(y, xlab = '', ylab = '', cex = 0.8, xaxt = 'n', yaxt = 'n')
93 mtext("Indice das observacoes", side = 1, line = 2.0, cex = 1.8)
94 mtext("Distancia de Cook", side = 2, line = 2, cex = 1.8)
95 abline(h = seq(Ry[1], Ry[2], l = 5), v = seq(Rx[1], Rx[2], l = 5), col = "gray", lty = "dotted")
96 axis(2, seq(Ry[1], Ry[2], l = 5), FF(seq(Ry[1], Ry[2], l = 5), 3))
97 axis(1, seq(Rx[1], Rx[2], l = 5), FF(seq(Rx[1], Rx[2], l = 5), 0))
98 graphics.off()

```