Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”

Modelos mistos para a análise da tonalidade da cor da casca de mamão (*Carica papaya* L.) cv. “Sunrise Solo”, avaliada ao longo do tempo por meio de um *scanner* e de um colorímetro

Thiago de Paula Oliveira

Dissertação apresentada para obtenção do título de Mestre em Ciências. Área de concentração: Estatística e Experimentação Agronômica

Piracicaba
2014

Thiago de Paula Oliveira
Engenheiro Agrônomo

**Modelos mistos para a análise da tonalidade da cor da casca de mamão
(*Carica papaya* L.) cv. “Sunrise Solo”, avaliada ao longo do tempo por meio
de um *scanner* e de um colorímetro**

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Orientador:
Prof. Dr. **SILVIO SANDOVAL ZOCCHI**

Dissertação apresentada para obtenção do título
de Mestre em Ciências. Área de concentração:
Estatística e Experimentação Agronômica

**Piracicaba
2014**

**Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA - ESALQ/USP**

Oliveira, Thiago de Paula

Modelos mistos para a análise da tonalidade da cor da casca de mamão (*Carica papaya* L.) cv. "Sunrise Solo", avaliada ao longo do tempo por meio de um *scanner* e de um colorímetro / Thiago de Paula Oliveira. - - versão revisada de acordo com a resolução CoPGr 6018 de 2011. - - Piracicaba, 2014.
155 p. : il.

Dissertação (Mestrado) - - Escola Superior de Agricultura "Luiz de Queiroz", 2013.
Bibliografia.

1. Imagens digitalizadas 2. Coloração desuniforme 3. Amostragem 4. Espaço de cores
CIE LCH 5. Software R I. Título

CDD 634.651
O48m

"Permitida a cópia total ou parcial deste documento, desde que citada a fonte – O autor"

DEDICATÓRIA

A minha família, amigos e ao meu orientador

por toda força, incentivo, companheirismo e amizade.

Dedico.

AGRADECIMENTOS

Agradeço, primeiramente, a Deus.

Aos meus pais, por me ensinarem a distinguir entre o que eles acreditam ser certo e errado, apoiando-me na maioria das minhas decisões. Também agradeço-os por lutarem para dar a mim a escolaridade que, infelizmente, eles não conseguiram obter.

Estendo meu agradecimento a toda minha família e especialmente a minha tia Maria e ao meu tio Arnaldo, os quais me ensinaram que fazer o bem as pessoas é fazer bem a si e que o amor é algo que deve ser sempre cultivado.

O meu muito obrigado ao Renan F. Alves e a sua família por me ajudarem em todos os momentos que precisei, sendo sempre muito atenciosos e amoráveis. Deixo também o meu muito obrigado ao Davi, “Maomé”, por todas as risadas, pelo companheirismo e amizade.

Agradeço ao meu “brother” Rafael Moral e a sua família pelo companheirismo e pela força que sempre me deram, aos meus queridos amigos Marcus (“Maverick”), Djair Durand (“Djavan”), Ricardo Klein, Maria Cristina, Bruna G. Wendpap, Andréia Erdmann, Célio Fernando, Isabela (“Tic Tac”), Edilan Quaresma, Fábio (“Xiri”), Livia M. D. Mella, Iuri Ferreira, Ana Júlia (“Porta”), Thiago Gentil, Rodrigo Pescim, Bárbara Navarro, Reginaldo Hilário, Renata Alcarde, Solange Sabadin.

Estendo meu agradecimento a Valéria P. Torquato por todos os ensinamentos, risadas e por trazer a minha história de vida momentos de muita alegria.

Agradeço ao meu querido orientador Silvio Sandoval Zocchi, uma pessoa maravilhosa que felizmente faz parte da minha vida. Um amigo que para mim tornou-se um irmão. O meu muito obrigado por todos os ensinamentos, pela paciência, por acreditar em mim, no meu potencial e por sempre estar disposto a me ajudar.

Agradeço a todos os alunos do programa, aos docentes e funcionários do Departamento de Ciências Exatas da ESALQ pelo aprendizado, amizade e risadas. Estendo meu agradecimento a Maria da Glória Eloi da Silva pela atenção.

Por fim, agradeço a CAPES pela concessão da bolsa de pesquisa e a FAPESP pelo financiamento do experimento.

Sonata de uma vida

Sempre de mãos dadas
Corríamos pelo campo a brincar
Assoprávamos as sementes do dente-de-leão
Germinando alegria nos corações alheios

Nossas palavras, eram os elementos de uma partitura
Nosso pensamento... o som de um piano
Uma música que me ensinou a ser humano
Uma música que me ensinou a amar!

Porém, em meio a tantas tempestades, tu fechastes os olhos!
Deixando inesquecíveis lembranças no tempo...
Mas eu continuarei escrevendo a nossa partitura!
Continuarei donde tu parastes

E saiba, minha querida, que muitos a ouvirão
E que a tocarão, alegrando o coração de outros mais
Pois tudo o que importa nessa tão breve vida
Tu me ensinastes!

Autor: Thiago de Paula Oliveira

Uma pequena homenagem a Maria Rodrigues da Costa (in memoriam).

SUMÁRIO

RESUMO	11
ABSTRACT	13
LISTA DE FIGURAS	15
LISTA DE TABELAS	19
LISTA DE ABREVIATURAS	23
LISTA DE SÍMBOLOS	25
1 INTRODUÇÃO	29
2 REVISÃO BIBLIOGRÁFICA	31
2.1 A importância e caracterização da cor	31
2.2 Modelos lineares de efeitos mistos	35
2.3 Estruturas de covariância	39
2.4 Estimação dos componentes de variância	41
2.5 Estimação dos termos fixos e predição dos termos aleatórios	47
2.5.1 Componentes de variância conhecidos	48
2.5.2 Variância desconhecida	48
2.5.3 Predição dos efeitos aleatórios	49
2.5.4 Estimação dos efeitos fixos e predição dos efeitos aleatórios simultaneamente	50
2.6 Seleção de modelos	52
2.6.1 Critérios de informação	52
2.6.2 Teste da razão de verossimilhanças	53
2.7 Diagnósticos	54
3 MATERIAL	59
4 MÉTODOS	63
5 RESULTADOS E DISCUSSÃO	75
5.1 Análise descritiva	75
5.2 Ajuste dos modelos	79
6 CONSIDERAÇÕES FINAIS	109
REFERÊNCIAS	111
APÊNDICES	117
ANEXOS	149

RESUMO

Modelos mistos para a análise da tonalidade da cor da casca de mamão (*Carica papaya* L.) cv. “Sunrise Solo”, avaliada ao longo do tempo por meio de um *scanner* e de um colorímetro

O mamão (*Carica papaya* L.) cv. “Sunrise Solo” é um fruto que apresenta mudança gradual e desuniforme da cor da casca, que vai de verde para amarela. Isso faz com que a metodologia instrumental para avaliação da cor, por meio de um colorímetro, seja subjetiva, devido ao número de pontos observados, bem como às localizações deles no fruto. Como alternativa, foi proposta a utilização de imagens digitalizadas de toda região da casca do fruto, obtidas por meio de um *scanner* de mesa. Para avaliar a precisão desses métodos, foi conduzido um experimento com 20 repetições. Cada repetição era constituída de um fruto de mamoeiro cv. “Sunrise Solo”, mantido sob temperatura e umidade relativa controladas. A cor da casca dos frutos foi avaliada, diariamente, utilizando um colorímetro e um *scanner*. Com o *scanner*, foram digitalizadas as duas faces do fruto e, com o colorímetro, foram observados quatro pontos equidistantes, na região equatorial do mesmo. Como a avaliação para cada fruto foi feita ao longo do tempo, os dados são classificados como longitudinais. Assim, utilizaram-se modelos lineares de efeitos mistos para estudar o comportamento da tonalidade média, pois essa técnica permite o uso de diferentes estruturas de variâncias e covariâncias para as matrizes dos efeitos aleatórios e dos erros. O processo de seleção do modelo foi realizado por meio do teste da razão de verossimilhanças e dos critérios de informação AIC e BIC, resultando no mesmo preditor linear e matrizes de variâncias e covariâncias para ambas as metodologias de quantificação da cor. O modelo final apresentou um preditor linear quadrático com efeitos aleatórios para o intercepto e para os termos linear e quadrático com matriz de variâncias e covariâncias dos efeitos aleatórios não estruturada e componentes de variância com heterocedasticidade para os erros. A utilização do *scanner* revelou dois grupos de maturação fisiológica distintos, que podem estar relacionados ao ponto de colheita do fruto, fato que não ficou evidente ao utilizar um colorímetro. De forma geral, o uso de um *scanner* possibilitou obter uma avaliação precisa da maturação do fruto, além de ser mais consistentes e eficiente do que o uso de um colorímetro para estudar a tonalidade média da casca de frutos que apresentam coloração desuniforme.

Palavras-chave: Imagens digitalizadas; Coloração desuniforme; Amostragem; Espaço de cores *CIELCH*; *Software R*

ABSTRACT

Mixed models for analysis of hue peel color of papaya (*Carica papaya* L.) cv. “Sunrise Solo”, measured along time by means of a scanner and a colorimeter

Papayas (*Carica papaya* L.) of “Sunrise solo” variety are fruits that present gradual and uneven changes in the peel color, which goes from green to yellow. As a consequence, when using a colorimeter to quantify their color, the results are subjective because of the number of observed points, as well as because of their position on the fruit. A proposed alternative was to use scanned images of the whole fruit peel to quantify color. To assess the precision of these methods, an experiment with 20 replicates was carried out. Each replicate consisted of a papaya fruit, kept under controlled temperature and relative humidity. The fruits’ peel colors were assessed, daily, using a colorimeter and a scanner. With the scanner, both sides of the fruit were scanned and, with the colorimeter, four equidistant points at the equatorial region of the fruit were observed. As the assessment was made through time for a same fruit, the data are classified as longitudinal. Therefore, linear mixed effect models were used to study the behavior of the average fruit color tonality through time, as this technique allows usage of different random effects and error covariance structures. Model selection was made using likelihood-ratio tests and the Akaike and Bayesian information criteria, which resulted in the selection of the same linear predictor and covariance matrices for both color quantification methods. The final model presented a quadratic linear predictor with random effects for the intercept, linear and quadratic terms with an unstructured variance-covariance matrix for the random effects and a variance components with heterogeneity matrix for the residuals. The use of a scanner revealed two distinct physiological maturation groups, which may be related to the harvesting time. This was not observed when using a colorimeter. In general, using a scanner made possible to obtain more consistent observations, which makes it a more efficient methodology to study the average fruit peel color tonality.

Keywords: Digitized image; Uneven coloration; Sampling; *CIELCH* color space; *R Software*

LISTA DE FIGURAS

Figura 1 - Região das cores pertencentes ao sistema de cores <i>CIELab</i> (ou <i>CIELCH</i>), com representação de duas cores: A, amarela, em que $a = 8$ e $b = 54$ (ou $C = 55$ e $H = 82^\circ$) e B, vermelha, em que $a = 50$ e $b = 22$ (ou $C = 55$ e $H = 24^\circ$)	34
Figura 2 - Representação das cores da carta de Munsell na região segundo o sistema <i>CIELCH</i>	34
Figura 3 - Frutos de mamoeiro estocados numa câmara de refrigeração em dias distintos	59
Figura 4 - (a) <i>Scanner</i> de mesa com um fruto e (b) sistema para a medição da cor de um fruto composto por um <i>scanner</i> de mesa com um fruto, coberto por uma caixa preta, interligado a um computador para o armazenamento da imagem	60
Figura 5 - Exemplo de 17 padrões da carta de cores de Munsell para tecidos vegetais	60
Figura 6 - Imagens digitalizadas da face A de um mesmo fruto aos 0, 5, 10, 15 e 17 dias após o início do experimento	61
Figura 7 - (a) Colorímetro tristímulus Minolta CR-300 (KONICA MINOLTA) e (b) exemplo de dois pontos para avaliação da cor por meio do colorímetro. . .	62
Figura 8 - Fluxograma do procedimento de seleção a partir do modelo completo M11. Teste de razão de verossimilhanças para verificar a necessidade de exclusão de um efeito aleatório (círculos em azul) e para verificar a necessidade de exclusão de um efeito fixo (círculos em laranja); retângulos em vermelho, verde ou amarelo indicam os modelos que podem ser comparados pelos critérios de informação AIC ou BIC	67
Figura 9 - Tonalidade média de frutos medidas ao longo do tempo por meio de um <i>scanner</i> (a) e de um colorímetro (b), bem como a escala de cores baseada na tonalidade, fixando $L = 50$ e $C = 50$	76
Figura 10 - Tonalidade média de frutos medidos ao longo do tempo por um colorímetro, desconsiderando os dados relativos aos dias 5 e 8, bem como a escala de cores baseada na tonalidade, fixando $L = 50$ e $C = 50$	77
Figura 11 - Gráficos de perfis das tonalidades médias das cascas de 20 frutos, medidas por meio de um <i>scanner</i> e de um colorímetro por um período de 16 dias, desconsiderando os dados relativos aos dias 5 e 8 para o colorímetro . . .	79

Figura 12 - Intervalos de 95% de confiança dos parâmetros de efeitos fixos do modelo polinomial quadrático (32) ajustados aos dados de tonalidade média da casca dos 20 frutos obtidos pelo <i>scanner</i>	80
Figura 13 - Intervalos de 95% de confiança dos parâmetros de efeitos fixos do modelo polinomial quadrático (32) ajustados aos dados de tonalidade média da casca dos 20 frutos obtidos pelo colorímetro	81
Figura 14 - Função de autocorrelação empírica (a) e autocorrelação parcial para a metodologia do <i>scanner</i> (b) e semivariograma em função da distância para a metodologia do colorímetro, com a utilização de um método de suavização LOESS para facilitar a visualização dos padrões (c)	84
Figura 15 - Gráfico dos resíduos condicionais padronizados em função do tempo após o início do experimento para o <i>scanner</i> (a) e para o colorímetro (b)	85
Figura 16 - Proposta de amostragem, por meio de um colorímetro, da tonalidade do fruto da cultivar “Sunrise Solo” com pontos em vermelho representando locais no fruto onde podem ser realizadas as observações	89
Figura 17 - Gráfico da tonalidade média estimada inicial da casca do fruto ($\hat{\beta}_0 + \tilde{b}_{0i}$) em função da velocidade de maturação aos 8 dias após o início do experimento (v_{8i}) para o <i>scanner</i> (a) e para o colorímetro (b)	92
Figura 18 - Gráficos de dispersão entre os efeitos aleatórios do modelo apresentado na expressão (38) e coeficientes de correlação linear entre os mesmos, considerando-se os dados obtidos por meio do <i>scanner</i> (a) e do colorímetro (b)	94
Figura 19 - Gráfico dos resíduos condicionais padronizados em função do tempo após o início do experimento para o <i>scanner</i> (a) e para o colorímetro (b)	95
Figura 20 - Gráfico das estimativas para os parâmetros $\delta_{s_{ij}}$ de variância relativos aos K níveis de estratificação (linha azul), bem como os limites superior e inferior do intervalo de 95% de confiança para os parâmetros $\delta_{s_{ij}}$ de variância relativos aos K níveis de estratificação (linha preta) para o <i>scanner</i> (a) e para o colorímetro (b), utilizando o fruto 2 como exemplo	96
Figura 21 - Gráfico de normalidade dos resíduos condicionais padronizados para o <i>scanner</i> (a) e para o colorímetro (b)	99

- Figura 22 -Histograma de normalidade dos resíduos condicionais padronizados, com adição de uma curva em vermelho indicando a distribuição amostral e uma curva em azul indicando a distribuição normal padrão 99
- Figura 23 -Histograma da distribuição empírica do coeficiente de assimetria e de curtose obtidas a partir da simulação de 1000 amostras aleatórias da distribuição normal padrão de tamanho 303, com adição de uma linha vermelha indicando a distribuição empírica e de uma linha azul indicando o valor calculado (V.C.) da assimetria e da curtose dos resíduos condicionais padronizados do modelo da expressão (38) para o *scanner* (a) e (b) e para o colorímetro (c) e (d) 100
- Figura 24 -Gráficos normais de probabilidade com envelope simulado para $\mathbf{b}_{0i}$ ((a) e (b)), $\mathbf{b}_{1i}$ ((c) e (d)) e $\mathbf{b}_{2i}$ ((e) e (f)) para o *scanner* e colorímetro, respectivamente 101
- Figura 25 -Gráfico dos resíduos condicionais padronizados em função dos valores preditos para o *scanner* (a) e para o colorímetro (b) 103
- Figura 26 -Resíduos condicionais padronizados individuais em função do tempo e curvas não paramétricas de regressão LOESS para o modelo apresentado na expressão (38 para o *scanner* (a) e para o colorímetro (b) 103
- Figura 27 -Gráfico dos valores ajustados (em graus) em função dos valores observados (em graus) para o *scanner* (a) e para o colorímetro (b) 104
- Figura 28 -Gráfico das curvas individuais ajustadas a partir do modelo (38) para a tonalidade média (em graus) em função do tempo (em dias) para o *scanner* (a) e para o colorímetro (b), bem como uma escala de cores baseada na tonalidade, fixando $L = 50$ e $C = 50$ 105
- Figura 29 -Gráfico dos ajustes individuais (curva rosa) para a tonalidade média (em graus) em função do tempo (em dias), com adição do modelo de efeitos fixos (curva azul) e das observações para a metodologia do *scanner* 106
- Figura 30 -Gráfico dos ajustes individuais (curva rosa) para a tonalidade média (em graus) em função do tempo (em dias), com adição do modelo de efeitos fixos (curva azul) e das observações para a metodologia do colorímetro . . 107

LISTA DE TABELAS

Tabela 1 - Modelos lineares de efeitos mistos construídos a partir de (1) considerando-se $p = 2$	66
Tabela 2 - Tonalidade média inicial, em graus, de 20 frutos avaliados por meio de um colorímetro e de um <i>scanner</i> ; números em negrito foram utilizados para destacar os frutos com estágio de maturação mais avançado (mais amarelados)	76
Tabela 3 - Estatística χ^2 para o teste da razão de verossimilhanças (TRV) entre o modelo M11 e os modelos aninhados subsequentes M10, M9 e M8; critérios de informação AIC e BIC para os modelos M1 a M11 ajustados aos dados de tonalidade média obtidos por meio do <i>scanner</i> e respectivos números de parâmetros do modelo (np), números de graus de liberdade (gl) e valores-p do teste	81
Tabela 4 - Estatística χ^2 para o teste da razão de verossimilhanças (TRV) entre o modelo M11 e os modelos aninhados subsequentes M10, M9 e M8; critérios de informação AIC e BIC para os modelos M1 a M11 ajustados aos dados de tonalidade média obtidos por meio do <i>scanner</i> e respectivos números de parâmetros do modelo (np), números de graus de liberdade (gl) e valores-p do teste	82
Tabela 5 - Estatística χ^2 para o teste da razão de verossimilhanças (TRV) para seleção da matriz de covariâncias $\mathbf{D}_*$ a partir do modelo M11, para o <i>scanner</i> , bem como os critérios de informação AIC e BIC	83
Tabela 6 - Estatística χ^2 para o teste da razão de verossimilhanças (TRV) para seleção da matriz de covariâncias $\mathbf{D}_*$ a partir do modelo M11, para o colorímetro, bem como os critérios de informação AIC e BIC	83
Tabela 7 - Estatística χ^2 para o teste da razão de verossimilhanças (TRV) para seleção da matriz de covariâncias $\mathbf{\Sigma}_i$ a partir do modelo M11, para o <i>scanner</i> , bem como os critérios de informação AIC e BIC	86
Tabela 8 - Estatística χ^2 para o teste da razão de verossimilhanças (TRV) para seleção da matriz de covariâncias $\mathbf{\Sigma}_i$ a partir do modelo M11, para o <i>scanner</i> , bem como os critérios de informação AIC e BIC	86

- Tabela 9 - Estatística χ^2 para o teste da razão de verossimilhanças (TRV) entre o modelo M11 e os modelos aninhados subsequentes M10, M9 e M8, assumindo estrutura NE para a matriz $\mathbf{D}_*$ e CV_H para Σ_i ; critérios de informação AIC e BIC para os modelos M1 a M11 ajustados aos dados de tonalidade média obtidos pelo *scanner* e respectivos números de parâmetros do modelo (np), números de graus de liberdade (gl) e valores-p do teste . . . 87
- Tabela 10 - Estatística χ^2 para o teste da razão de verossimilhanças (TRV) entre o modelo M11 e os modelos aninhados subsequentes M10, M9 e M8, assumindo estrutura NE para a matriz $\mathbf{D}_*$ e CV_H para Σ_i ; critérios de informação AIC e BIC para os modelos M1 a M11 ajustados aos dados de tonalidade média obtidos pelo colorímetro e respectivos números de parâmetros do modelo (np), números de graus de liberdade (gl) e valores-p do teste . . . 88
- Tabela 11 - Estimativas dos efeitos fixos do modelo (38) e os respectivos limites inferior (LI) e superior (LS) do intervalo de confiança aproximado, considerando um nível de 5% de significância, para o *scanner* 88
- Tabela 12 - Estimativas dos efeitos fixos do modelo (38) e os respectivos limites inferior (LI) e superior (LS) do intervalo de confiança aproximado, considerando um nível de 5% de significância, para o colorímetro 88
- Tabela 13 - Predições dos efeitos aleatórios do modelo apresentado na expressão (38) da tonalidade média de frutos medidos pelo *scanner* e velocidades de maturação para cada indivíduo após 8 dias do início do experimento (v_{8i}) em graus por dia 90
- Tabela 14 - Predições dos efeitos aleatórios do modelo apresentado na expressão (38) da tonalidade média de frutos medidos pelo colorímetro e velocidades de maturação para cada indivíduo após 8 dias do início do experimento (v_{8i}) 90
- Tabela 15 - Estimativas e limites inferiores (LI) e superiores (LS) dos intervalos de confiança, com nível de 5% de probabilidade, para o desvio padrão do intercepto ($\sigma_{\mathbf{b}_0}$), do efeito linear ($\sigma_{\mathbf{b}_1}$) e do efeito quadrático ($\sigma_{\mathbf{b}_2}$) da matriz de variâncias e covariâncias $\mathbf{D}_*$, bem como os intervalos de confiança, com nível de 5% de probabilidade, para o coeficiente de correlação entre os efeitos aleatórios $\mathbf{b}_0$ e $\mathbf{b}_1$ ($\rho(\mathbf{b}_0, \mathbf{b}_1)$), $\mathbf{b}_0$ e $\mathbf{b}_2$ ($\rho(\mathbf{b}_0, \mathbf{b}_2)$) e $\mathbf{b}_1$ e $\mathbf{b}_2$ ($\rho(\mathbf{b}_1, \mathbf{b}_2)$) para o *scanner* e para o colorímetro 93

Tabela 16 -Estimativas para os $\delta_{s_{2j}}$, bem como os respectivos limites inferior (LI) e superior (LS) do intervalo de confiança aproximado, considerando um nível de 5% de significância, para o parâmetro utilizando-se a metodologia do <i>scanner</i> ; o fruto 2 (i=2) foi utilizado como exemplo uma vez que apresenta maior número de medidas repedidas	96
Tabela 17 -Estimativas para os $\delta_{s_{2j}}$, bem como os respectivos limites inferior (LI) e superior (LS) do intervalo de confiança aproximado, considerando um nível de 5% de significância, para o parâmetro utilizando-se a metodologia do colorímetro; o fruto 2 (i=2) foi utilizado como exemplo uma vez que apresenta maior número de medidas repedidas	97
Tabela 18 -Dados de tonalidade média (em graus) dos frutos 1 ao 4 medidos ao longo do tempo (em dias) por meio de um <i>scanner</i> (Scan) e de um colorímetro (Col)	151
Tabela 19 -Dados de tonalidade média (em graus) dos frutos 5 ao 8 medidos ao longo do tempo (em dias) por meio de um <i>scanner</i> (Scan) e de um colorímetro (Col)	152
Tabela 20 -Dados de tonalidade média (em graus) dos frutos 9 ao 12 medidos ao longo do tempo (em dias) por meio de um <i>scanner</i> (Scan) e de um colorímetro (Col)	153
Tabela 21 -Dados de tonalidade média (em graus) dos frutos 13 ao 16 medidos ao longo do tempo (em dias) por meio de um <i>scanner</i> (Scan) e de um colorímetro (Col)	154
Tabela 22 -Dados de tonalidade média (em graus) dos frutos 17 ao 20 medidos ao longo do tempo (em dias) por meio de um <i>scanner</i> (Scan) e de um colorímetro (Col)	155

LISTA DE ABREVIATURAS

ACF – Função de auto correlação empírica

AIC – Critério de informação de Akaike

AR(1) – Auto regressiva de primeira ordem

AR_H(1) – Auto regressiva de primeira ordem com heterocedasticidade de variâncias

BIC – Critério de informação Bayesiano

BLUE – Melhor estimador não viesado

BLUP – Melhor preditor não viesado

CIE – Comissão Internacional de Iluminação

CV – Componentes de variâncias

CV_H – Componentes de variâncias com heterocedasticidade de variâncias

ESALQ – Escola Superior de Agricultura “Luiz de Queiroz”

EBLUP – Melhor preditor não viesado empírico

FLV – Função logaritmo da verossimilhança

GL – Graus de liberdade

LMEM – Modelo linear de efeitos misto

NE – Não estruturada

ML – Máxima verossimilhança

MLE – Estimativa de máxima verossimilhança

REML – Máxima verossimilhança restrita

SC – Simetria composta

SC_H – Simetria composta com heterocedasticidade de variâncias

LISTA DE SÍMBOLOS

$cov(.,.)$ – Covariância

$E(.)$ – Esperança matemática

$\mathbf{I}(.,.)$ – Matriz de informação esperada

i – Índice para indivíduos

j – Índice para repetições

L_{ML} – Função de máxima verossimilhança

l_{ML} – Logaritmo da função de máxima verossimilhança

L_R – Função de máxima verossimilhança restrita

l_R – Logaritmo da função de máxima verossimilhança restrita

l_{REML} – Logaritmo da função de máxima verossimilhança com respeito a todos os parâmetros $(\boldsymbol{\alpha}, \boldsymbol{\beta})$

k – Nível de estratificação

n – Número de repetições

$\mathcal{N}$ – Distribuição Normal

tr – Traço de uma matriz

$U(.)$ – Função escore de “Fisher”

$Var(.)$ – Variância

Y – Variável aleatória

$\mathbb{R}$ – Conjunto dos números reais

$\mathbf{W}(.)$ – representação da matriz $\mathbf{V}^{-1}(.)$

$\boldsymbol{\Theta}$ – Espaço paramétrico

σ^2 – Variância populacional

ϕ – componentes de variância associados a ε

$\hat{\beta}(\alpha)$ – Estimador de β

∂ – Derivada parcial

$\mathbf{b}$ – Vetor de efeitos aleatórios

$\mathbf{e}$ – Vetor de erros de $\mathbf{y} - \mathbf{X}\beta$

$\mathbf{y}$ – Vetor de respostas observadas

$\mathbf{0}$ – Vetor de zeros

κ – Vetor de elementos únicos da matriz simétrica $\mathbf{D}$

β – Vetor de efeitos fixos

ε – Vetor de erros

η – Vetor de erros

δ – Vetor de parâmetros de variância

θ – Vetor parâmetros desconhecidos

γ – Vetor de componentes de variância

$\mathbf{D}_*$ – Matriz de covariâncias não escalonada dos efeitos aleatórios

$\mathbf{D}$ – Matriz de covariâncias escalonada dos efeitos aleatórios

$\mathbf{I}$ – Matriz identidade

$\mathbf{L}$ – Matriz de contraste dos erros

$\mathbf{P}$ – Representa a matriz $\mathbf{V}_*^{-1} - \mathbf{V}_*^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{V}_*^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}_*^{-1}$

$\mathbf{V}_*(.)$ – Matriz de covariâncias não escalonada

$\mathbf{V}(.)$ – Matriz de covariâncias escalonada

$\mathbf{X}$ – Matriz de delineamento de variáveis explanatórias

$\mathbf{Z}$ – Matriz de delineamento de efeitos aleatórios

Φ – Matriz de zeros

Λ – Matriz de covariâncias escalonada do erro

Σ – Matriz de covariâncias do erro

1 INTRODUÇÃO

O consumo mundial de mamão (*Carica papaya* L.) apresentou grande crescimento nos últimos anos, sendo que o Brasil, em 2009, teve uma produção de aproximadamente 1,9 milhões de toneladas do fruto em uma área colhida de aproximadamente 34 mil hectares (EMBRAPA, 2010).

Independentemente da cultivar, a época de colheita é de fundamental importância, uma vez que frutos colhidos antes de atingirem a completa maturação fisiológica prejudicam o seu processo de amadurecimento e frutos colhidos totalmente maduros reduzem sua vida útil, gerando, em qualquer um dos casos, perdas quantitativas e qualitativas (CHITARRA; CHITARRA, 2005; JACOMINO; BRON; KLUGE, 2003). Segundo Fagundes e Yamanishi (2001), a qualidade do fruto depende do estágio de maturação e esse pode ser determinado por meio da cor da casca, que é um índice de maturidade considerado confiável, podendo também ser utilizado para determinar os pontos de colheita e de consumo de diversos frutos (KAYS, 1991).

A cor da casca é considerada como a primeira variável de qualidade avaliada pelos consumidores e está associada a gostos específicos ou usos que podem determinar sua aceitação ou rejeição no mercado (MENDOZA; AGUILERA, 2004; PICHA, 2006). De forma geral, há uma maior preferência por frutos com coloração amarela viva ou alaranjada brilhante em relação aos frutos com coloração esverdeada (FIORAVANÇO et al., 1994).

Assim, de forma a classificar os frutos de mamão cultivares “Golden” e “Sunrise Solo” segundo suas colorações, Balbino e Costa (2003) e a Associação Brasileira de Horticultura (2003), por meio do Programa Brasileiro para a Modernização da Horticultura, propuseram sistemas de classificação em subgrupos.

De modo a facilitar e padronizar o processo de medição da cor, técnicas objetivas e não destrutivas foram estudadas por Shahin e Symons (2001), Chen, Chao e Kim (2002) e por Darrigues et al. (2008) para a análise da cor, sendo uma das técnicas empregadas a análise de imagens digitalizadas, que superou as deficiências das técnicas visuais e instrumentais (colorímetro) e ofereceu uma medida objetiva da cor além de outros fatores físicos, como o formato do fruto (PAULUS; SCHREVEENS, 1999; SHAHIN; SYMONS, 2001; CHEN; CHAO; KIM, 2002).

A grande vantagem da utilização de imagens digitalizadas é de proporcionar

a construção de curvas de maturação baseadas na proporção de pixels da imagem em determinados intervalos de tonalidade da cor do fruto em função do tempo. Essa proposta abre novas possibilidades de aperfeiçoamento para a classificação dos frutos de acordo com a proporção de tonalidade desejada.

No entanto, como a cor é medida repetidamente em cada fruto ao longo do tempo, pode ocorrer dependência entre as observações que é uma característica inerente aos dados longitudinais. Assim, uma alternativa para a análise desses dados foi proposta inicialmente por Laird e Ware (1982) que utilizaram modelos mistos em dois estágios, possibilitando variações nas estruturas das matrizes de variâncias e covariâncias das observações entre as unidades e dentro das unidades experimentais. Atualmente, as técnicas para análise de dados longitudinais foram aperfeiçoadas, principalmente quanto aos métodos de estimação dos parâmetros do modelo (PINHEIRO; BATES, 1995; VERBEKE; MOLENBERGHS, 2009).

O mamão é um fruto que apresenta alta variabilidade da cor num mesmo fruto e entre frutos (OLIVEIRA; ZOCCHI, 2011). Deste modo, para incorporar tal variabilidade ao modelo, a inclusão de efeitos aleatórios pode ser necessária. Como consequência, a utilização de modelos mistos pode ser de fundamental importância para a análise de dados de cor. Esses modelos, por sua vez, são compostos por efeitos fixos, que descrevem o comportamento de toda a população ou de unidades associadas aos níveis de fatores experimentais, e por efeitos aleatórios, que englobam a variação entre os indivíduos (quando várias respostas são medidas por indivíduo), entre espécies, entre genótipos, entre períodos de tempo ou entre regiões (BOLKER et al., 2008).

Dessa forma, os modelos lineares de efeitos mistos foram utilizados aqui, como alternativa para incorporar a variabilidade entre os frutos e dentro de um mesmo fruto na análise da tonalidade média da casca de mamões cultivar “Sunrise Solo” medidos por meio de um *scanner* e de um colorímetro.

2 REVISÃO BIBLIOGRÁFICA

No item 2.1, é descrita a importância da cor da casca de mamões cv. “Sunrise Solo”, como caracterizá-la e determiná-la por meio de um *scanner* e de um colorímetro. No item 2.2, são descritos os modelos lineares de efeitos mistos (LMEM) aplicados a dados longitudinais. No item 2.3, são apresentadas as estruturas de variâncias e covariâncias mais comumente utilizadas para as matrizes $\mathbf{D}_*$ e Σ_i . Os métodos de estimação por máxima verossimilhança e por máxima verossimilhança restrita para os componentes de variância são descritos no item 2.4. No item 2.5, são descritos os métodos de estimação dos efeitos fixos e predição dos efeitos aleatórios e o método de estimação dos efeitos fixos e predição dos efeitos aleatórios simultaneamente. Os critérios de informação AIC e BIC e o teste da razão de verossimilhanças para seleção de modelos são descritos no item 2.6. Por fim, no item 2.7, são descritas as duas fontes diferentes de variabilidade dos modelos lineares de efeitos mistos e são apresentados os três tipos de perturbações que ocorrem nos LMEM.

2.1 A importância e caracterização da cor

No processo de amadurecimento, a coloração da casca é uma variável muito usada para classificação de frutos, estando diretamente relacionada ao estágio de maturação, ao ponto de colheita e à qualidade comestível dos mesmos (FAGUNDES; YAMANISHI, 2001; MENDOZA; AGUILERA, 2004).

Segundo Bleinroth (1995) e Jacomino, Bron e Kluge (2003), a grande importância em se conhecer esse ponto exato de colheita é para que o aroma e o sabor dos frutos se desenvolvam o melhor possível durante o amadurecimento. De acordo com os autores, os frutos da cultivar “Sunrise Solo” devem ser colhidos com no mínimo 11,5 °BRIX, o que corresponde a 6% de coloração amarelada na casca. Observam, ainda, o teor máximo de sólidos solúveis quando 33% da superfície do fruto estava amarelada.

A cor, neste caso, é considerada uma importante variável de qualidade, determinando o valor comercial dos frutos, além de ser responsável pelo impacto visual, que muitas vezes é marcante ao consumidor, sendo, portanto, a principal variável usada como critério de seleção (CHITARRA; CHITARRA, 2005; POLDERDIK et al., 1993; LIMA et al., 2004). Entretanto, apesar de haver limitações ao uso da coloração como meio de se prever o ponto de colheita ideal da cultura do mamão, essa é a variável rotineiramente

utilizada para as cultivares comercializadas, tanto no mercado interno quanto no mercado externo (BLEINROTH; SIGRIST, 1995).

Segundo Oliveira et al. (2002), o mamão tem por característica a mudança gradual e desuniforme da cor da casca de verde para amarela, formando inicialmente estrias amarelas partindo da região estilar em direção a inserção peduncular do fruto.

Para a quantificação da cor, por sua vez, podem-se utilizar escalas visuais pré-determinadas, aparelhos como os colorímetros ou, ainda, analisar imagens de objetos obtidas por meio de digitalização, fotos etc.

A caracterização das cores por escalas visuais pode ser feita por meio da comparação dos objetos com cores pertencentes a um catálogo de cores como o de Munsell. Essas, no entanto, são subjetivas, tendenciosas, demoradas e trabalhosas (VIEIRA et al., 1998; SHAHIN; SYMONS, 2001; BROSNAN; SUN, 2002; LI; WANG; GU, 2002; DARRIGUES et al., 2008). Além disso, a caracterização das cores por escalas visuais é um método que está sujeito a diferentes interpretações, uma vez que depende do treinamento do classificador, do tempo de trabalho e de sua fadiga. Assim, o mesmo fruto pode ter classificações distintas dependendo do classificador, o que pode levar à desuniformidade dos frutos numa caixa e/ou entre caixas (OLIVEIRA et al., 2002).

Por outro lado, para o emprego de colorímetros convencionais, são necessários frutos com coloração uniforme. Quando não é o caso, O'Sullivan et al. (2003) e Yam e Papadakis (2003) sugerem a observação de inúmeros pontos em cada fruto de modo a caracterizar a cor média do mesmo. No entanto, o número de pontos necessários, bem como a posição do ponto no fruto podem levar a medidas diferentes para um mesmo fruto.

De modo a reduzir as dificuldades de detecção de mudanças de cor em frutos, Darrigues et al. (2008) propuseram a utilização de *scanners* de mesa como dispositivos de medição de cores, por meio de imagens digitais, para tomates. Estendendo para outros produtos hortícolas, como a batata (*Solanum tuberosum* L.), o pepino (*Cucumis sativus* L.), a ameixa vermelha (*Prunus americana* Marsh.), o melão (*Cucumis melo* L.), a cenoura (*Daucus carota* L.) e o morango (*Fragaria ananassa* Duch.).

Convém observar que as imagens digitais geralmente utilizam o sistema de cores *RGB*, em que combinações de valores de cada componente da cor *R* (vermelho); *G* (verde) e *B* (azul), em cada pixel, reproduzem uma gama vasta de cores. Nesse sistema, a cor pertence ao espaço cúbico $[0; 1]^3$, em que cada eixo representa a variação de cada uma

das cores primárias ou componentes da cor, sendo a origem (0;0;0), a cor branca, tomada como ponto de referência do sistema (PASCALE, 2003).

Assim, sejam duas cores (R_1, G_1, B_1) e (R_2, G_2, B_2) pertencentes ao espaço de cores RGB , então a distância euclidiana entre as mesmas é dada por

$$\sqrt{(R_2 - R_1)^2 + (G_2 - G_1)^2 + (B_2 - B_1)^2}.$$

Considere, agora, que duas cores tenham distância fixa d . Dependendo da localização das cores no espaço RGB o olho humano pode ser capaz ou não de distingui-las. De modo a corrigir tal problema, os equipamentos usados para medição de cor geralmente utilizam sistemas de cores criados pela Comissão Internacional de Iluminação (CIE) propostos em 1976 (COMMISSION INTERNATIONALE DE L'ÉCLAIRAGE, 1978). Tais sistemas foram criados a partir de critérios baseados na percepção da cor pelo olho humano. No espaço de cor $CIELab$, por exemplo, o componente L indica a luminosidade da cor, variando de 0 (preto) a 100 (branco) e os componentes a e b pertencem ao intervalo de $[-60, 60]$ ou $[-100, 100]$. De forma geral, $a > 0$ e $b = 0$ corresponde a uma coloração próxima do vermelho e $a < 0$ e $b = 0$ corresponde a uma coloração próxima do verde, $b > 0$ e $a = 0$, coloração próxima do amarelo e $b < 0$ e $a = 0$, coloração próxima do azul.

Convém observar que quando $a = 0$ e $b = 0$, $L = 0$ ou $L = 100$, tem-se um ponto acromático e à medida que os componentes a e b da cor aumentam em módulo, a saturação ou intensidade da cor, dada por $C = \sqrt{a^2 + b^2}$, aumenta. A Figura 1 ilustra duas cores: amarelada (A) e avermelhada (B), com saturação próxima da máxima (60, no caso).

Além disso, a tonalidade da cor pode ser definida como um ângulo H , conforme ilustra a Figura 1. Neste caso, a região das cores formada pelos componentes L , C e H é denominada $CIELCH$.

Como ilustração, a Figura 2, apresenta cores da carta de Munsell na região de cores segundo o sistema $CIELCH$.

Para a calibração dos equipamentos envolvidos no processo de medição da cor, deve-se, segundo Darrigues et al. (2008); Rodríguez et al. (2010) e Oliveira e Zocchi (2010), escolher cores padrões baseadas na ampla gama de cores do objeto de interesse e realizar regressões lineares dos valores de luminosidade (L), de a e de b obtidos pelo

Figura 1 – Região das cores pertencentes ao sistema de cores *CIE Lab* (ou *CIE LCH*), com representação de duas cores: A, amarela, em que $a = 8$ e $b = 54$ (ou $C = 55$ e $H = 82^\circ$) e B, vermelha, em que $a = 50$ e $b = 22$ (ou $C = 55$ e $H = 24^\circ$)

Fonte: Modificado de ColorMetrix (2013)

Figura 2 – Representação das cores da carta de Munsell na região segundo o sistema *CIE LCH*

Fonte: Modificado de UNIVERSIDADE FEDERAL DE SANTA CATARINA (UFSC) (2013)

scanner, em função dos obtidos por meio de um colorímetro. Assim, as funções obtidas são utilizadas para a calibração da metodologia (BREWER et al., 2006; DARRIGUES et al., 2008; RODRÍGUEZ et al., 2010; OLIVEIRA; ZOCCHI, 2010).

De forma geral, pode-se utilizar as imagens digitalizadas para o cálculo, por exemplo, da tonalidade média. Por outro lado, também pode-se utilizá-las para a classifi-

cação de frutos, avaliando-se a proporção de pixels em um determinado intervalo para, por exemplo, os componentes L , C e H da cor. Adicionalmente, o emprego dessa técnica pode levar a uma melhor compreensão da tonalidade média dos frutos, comumente utilizada em pesquisas na área de fisiologia e tecnologia pós-colheita de produtos hortícolas. Além disso, o emprego de imagens digitalizadas também pode levar a uma melhor compreensão sobre a relação entre a porcentagem de pixels em um determinado intervalo de cor e a redução da clorofila e/ou aumento de carotenoides na superfície da casca do fruto (OLIVEIRA; ZOCCHI, 2011, 2013).

2.2 Modelos lineares de efeitos mistos

Nos experimentos para se avaliar a mudança da cor ao longo do tempo, cada fruto é avaliado repetidamente. De forma geral, quando são tomadas medidas repetidas a utilização de modelos lineares que considera apenas efeitos fixos é muito restrita (LINDSTROM; BATES, 1990), ocasionando análises estatísticas inadequadas.

Segundo Laird e Ware (1982) e Verbeke e Molenberghs (2009), a análise de dados longitudinais pode ser feito por meio de modelos multivariados com estrutura de variâncias e covariâncias irrestrita ou não estruturada. Esta, no entanto, é difícil, pois é comum esses conjuntos de dados serem incompletos ou desbalanceados ao longo do tempo.

Uma alternativa natural a essas técnicas de modelagem estatística é o uso de modelos mistos, que apresentam parâmetros de efeitos fixos e de efeitos aleatórios no modelo. Além do benefício da inclusão do efeito aleatório, os modelos mistos permitem a análise de dados longitudinais devido à flexibilidade de escolha das matrizes de variâncias e covariâncias dos efeitos aleatórios e dos erros, que podem assumir diferentes estruturas. Dentre elas têm-se, por exemplo, a de componentes de variância, simetria composta e autorregressiva de primeira ordem - AR(1), sendo que todas as estruturas mencionadas podem admitir ou não heteroscedasticidade de variâncias (ALCARDE, 2012).

Seja a resposta de interesse uma variável aleatória Y_{ij} , para o i -ésimo indivíduo, medida no tempo t_{ij} com $i = 1, 2, \dots, M$ e $j = 1, 2, \dots, n_i$, em que M é o número total de indivíduos e n_i é o número de observações realizadas para o i -ésimo indivíduo. Seja, ainda, $\mathbf{y}_i$ um vetor de dimensão n_i contendo as n_i medidas repetidas do indivíduo i , ou seja, $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{in_i})^T$. Então, para o caso em questão e de acordo com Demidenko (2004),

o LMEM para o indivíduo i pode ser escrito matricialmente da seguinte forma:

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

em que: $\mathbf{y}_i$ é o vetor de dimensão n_i de respostas do fruto i ; $\mathbf{X}_i$ é a matriz de delineamento de dimensões $n_i \times (p + 1)$ associados aos efeitos fixos; $\boldsymbol{\beta}$ é um vetor de dimensão $p + 1$ de parâmetros de efeitos fixos, também chamado de vetor dos coeficientes de efeitos fixos; $\mathbf{Z}_i$ é a matriz de delineamento de dimensões $n_i \times (q + 1)$ associada aos efeitos aleatórios; $\mathbf{b}_i$ é um vetor de parâmetros de efeitos aleatórios de dimensão $q + 1$ tal que $\mathbf{b}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{D}_*)$, sendo $\mathbf{D}_*$ a matriz de covariâncias dos efeitos aleatórios de dimensões $(q + 1) \times (q + 1)$ e $\boldsymbol{\varepsilon}_i$ é um vetor de erros de dimensão n_i tal que $\boldsymbol{\varepsilon}_i \sim \mathcal{N}(\mathbf{0}, \sigma^2 \boldsymbol{\Lambda}_i)$, em que $\boldsymbol{\Lambda}_i$ é uma matriz de variâncias e covariâncias escalonada com dimensões que dependem apenas do indivíduo i . Deve-se ressaltar, ainda, que o vetor de efeitos aleatórios $\mathbf{b}_i$ é independente do vetor de erros $\boldsymbol{\varepsilon}_i$.

Assim, no caso de uma regressão polinomial de grau $p \leq q$ com efeitos aleatórios, os vetores e as matrizes do modelo (1) apresentam a seguinte forma:

$$\mathbf{y}_i = \begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{in_i} \end{bmatrix}, \mathbf{X}_i = \begin{bmatrix} 1 & t_{i1} & t_{i1}^2 & \dots & t_{i1}^p \\ 1 & t_{i2} & t_{i2}^2 & \dots & t_{i2}^p \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & t_{in_i} & t_{in_i}^2 & \dots & t_{in_i}^p \end{bmatrix}, \mathbf{Z}_i = \begin{bmatrix} 1 & t_{i1} & t_{i1}^2 & \dots & t_{i1}^q \\ 1 & t_{i2} & t_{i2}^2 & \dots & t_{i2}^q \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & t_{in_i} & t_{in_i}^2 & \dots & t_{in_i}^q \end{bmatrix},$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}, \mathbf{b}_i = \begin{bmatrix} b_{0i} \\ b_{1i} \\ \vdots \\ b_{qi} \end{bmatrix} \text{ e } \boldsymbol{\varepsilon}_i = \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \vdots \\ \varepsilon_{in_i} \end{bmatrix}.$$

Frequentemente considera-se que $\boldsymbol{\Lambda}_i = \mathbf{I}_{n_i}$, sendo $\mathbf{I}_{n_i}$ a matriz identidade de ordem n_i . Entretanto, a adoção de matrizes diferentes da identidade permite a especificação de diferentes estruturas de covariância, como por exemplo a AR(1) (LINDSTROM; BATES, 1990).

Em termos de modelos hierárquicos, o modelo linear misto geral para o i -ésimo indivíduo pode ser reescrito da seguinte forma:

$$\begin{aligned}\mathbf{Y}_i|\mathbf{b}_i &\sim \mathcal{N}(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i, \sigma^2\boldsymbol{\Lambda}_i) \\ \mathbf{b}_i &\sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{D}).\end{aligned}\tag{2}$$

em que

$$\mathbf{D} = \frac{1}{\sigma^2}\mathbf{D}_* = \frac{1}{\sigma^2}\text{cov}(\mathbf{b}_i).$$

Assume-se, aqui, que os vetores aleatórios $\mathbf{b}_i$ e $\boldsymbol{\varepsilon}_i$ são independentes. Além disso, de acordo com Demidenko (2004), para que o LMEM seja identificável para $\boldsymbol{\beta}$, assume-se que a matriz $\sum_{i=1}^M \mathbf{X}_i^T \mathbf{X}_i$ é não singular e que $\sum_{i=1}^M n_i \geq (p+1)$, com $i = 1, 2, \dots, M$.

Adicionalmente, para fazer o modelo identificável para σ^2 e para $\mathbf{D}$, assume-se que pelo menos uma matriz $\mathbf{Z}_i^T \mathbf{Z}_i$ é positiva definida e que $\sum_{i=1}^M n_i \geq M(q+1)$.

Considerando os M indivíduos, pode-se, então, escrever o modelo:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\varepsilon},\tag{3}$$

em que

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_M \end{bmatrix}, \mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_M \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}, \mathbf{Z} = \begin{bmatrix} \mathbf{Z}_1 & \boldsymbol{\Phi} & \dots & \boldsymbol{\Phi} \\ \boldsymbol{\Phi} & \mathbf{Z}_2 & \dots & \boldsymbol{\Phi} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Phi} & \boldsymbol{\Phi} & \dots & \mathbf{Z}_M \end{bmatrix},$$

$$\mathbf{b} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_M \end{bmatrix} \text{ e } \boldsymbol{\varepsilon} = \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_M \end{bmatrix}$$

têm dimensões dadas por: $N_T \times 1$, $N_T \times (p+1)$, $(p+1) \times 1$, $N_T \times M(q+1)$, $M(q+1) \times 1$ e $N_T \times 1$, respectivamente, em que $N_T = \sum_{i=1}^M n_i$ é o número total de observações e $\boldsymbol{\Phi}$ são matrizes de zeros. Adicionalmente, considere que $\boldsymbol{\Sigma}$ é uma matriz $N_T \times N_T$ de variâncias e covariâncias não escalonada dos erros e que $\boldsymbol{\Lambda}$ é uma matriz de dimensões $N_T \times N_T$ de

variâncias e covariâncias escalonada dos erros, ou seja,

$$\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{\Lambda} = \sigma^2 \begin{bmatrix} \boldsymbol{\Lambda}_1 & \boldsymbol{\Phi} & \dots & \boldsymbol{\Phi} \\ \boldsymbol{\Phi} & \boldsymbol{\Lambda}_2 & \dots & \boldsymbol{\Phi} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Phi} & \boldsymbol{\Phi} & \dots & \boldsymbol{\Lambda}_M \end{bmatrix}.$$

Seja $\boldsymbol{\gamma}$ um vetor formado pelos $\frac{(q+1)(q+2)}{2}$ elementos únicos da matriz simétrica $\mathbf{D}_*$ e por todos os parâmetros da matriz $\boldsymbol{\Sigma}_i = \sigma^2 \boldsymbol{\Lambda}_i$. Considere, agora, que $\boldsymbol{\eta} = \mathbf{Z}\mathbf{b} + \boldsymbol{\varepsilon}$ pode-se reescrever o modelo (3) da seguinte forma:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\eta}, \quad (4)$$

em que

$$\boldsymbol{\eta} = \begin{bmatrix} \boldsymbol{\eta}_1 \\ \boldsymbol{\eta}_2 \\ \vdots \\ \boldsymbol{\eta}_M \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_1 \mathbf{b}_1 + \boldsymbol{\varepsilon}_1 \\ \mathbf{Z}_2 \mathbf{b}_2 + \boldsymbol{\varepsilon}_2 \\ \vdots \\ \mathbf{Z}_M \mathbf{b}_M + \boldsymbol{\varepsilon}_M \end{bmatrix}.$$

Pode-se provar que $E[\boldsymbol{\eta}] = \mathbf{0}$ e que a matriz de covariâncias de $\boldsymbol{\eta}$ é bloco diagonal na forma:

$$\begin{aligned} \mathbf{V}_*(\boldsymbol{\gamma}) &= \sigma^2 \begin{bmatrix} \mathbf{Z}_1 \mathbf{D} \mathbf{Z}_1^T + \boldsymbol{\Lambda}_1 & \boldsymbol{\Phi} & \boldsymbol{\Phi} & \boldsymbol{\Phi} \\ \boldsymbol{\Phi} & \mathbf{Z}_2 \mathbf{D} \mathbf{Z}_2^T + \boldsymbol{\Lambda}_2 & \boldsymbol{\Phi} & \boldsymbol{\Phi} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Phi} & \boldsymbol{\Phi} & \boldsymbol{\Phi} & \mathbf{Z}_M \mathbf{D} \mathbf{Z}_M^T + \boldsymbol{\Lambda}_M \end{bmatrix} = \sigma^2 \mathbf{V}(\boldsymbol{\gamma}) \\ &= \sigma^2 \left(\bigoplus_{i=1}^M \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T + \boldsymbol{\Lambda}_i \right) = \bigoplus_{i=1}^M \sigma^2 \left(\mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T + \boldsymbol{\Lambda}_i \right) = \bigoplus_{i=1}^M \mathbf{V}_{*i}(\boldsymbol{\gamma}), \end{aligned} \quad (5)$$

ou seja, $\boldsymbol{\eta} \stackrel{iid}{\sim} \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{V}(\boldsymbol{\gamma}))$ (ver Apêndice A) e $\mathbf{V}_{*i}(\boldsymbol{\gamma}) = \sigma^2 (\mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T + \boldsymbol{\Lambda}_i)$ é a matriz de variâncias e covariâncias para o indivíduo i .

Portanto, o LMEM com efeitos aleatórios normalmente distribuídos pode ser escrito marginalmente como:

$$\mathbf{Y}_i \stackrel{iid}{\sim} \mathcal{N}(\mathbf{X}_i \boldsymbol{\beta}, \sigma^2 (\mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T + \boldsymbol{\Lambda}_i)), \quad i = 1, 2, \dots, M. \quad (6)$$

O modelo marginal (6), entretanto, não evidencia qual modelo hierárquico foi ajustado, uma vez que modelos hierárquicos diferentes podem produzir o mesmo modelo marginal (VERBEKE; MOLENBERGHS, 2009).

2.3 Estruturas de covariância

Considere que $\Sigma_i = \sigma^2 \mathbf{\Lambda}_i$. Logo, tem-se que, para o modelo (1),

$$\varepsilon_i \stackrel{iid}{\sim} \mathcal{N}(\mathbf{0}, \Sigma_i) \text{ e } \mathbf{b}_i \stackrel{iid}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{D}_*).$$

Neste caso, as matrizes de variâncias e covariâncias Σ_i e $\mathbf{D}_*$ podem assumir diferentes estruturas, sendo que algumas delas, comumente utilizadas, são apresentadas a seguir:

(i) Não estruturada – NE:

$$\begin{bmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1K} \\ \sigma_{12} & \sigma_2^2 & \dots & \sigma_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1K} & \sigma_{2K} & \dots & \sigma_K^2 \end{bmatrix}.$$

Essa matriz contém $\frac{K(K+1)}{2}$ parâmetros, em que K denota a ordem da matriz, sendo que essa supõe heteroscedasticidade para os componentes.

(ii) Componentes de variâncias – CV:

$$\sigma^2 \mathbf{I}_K = \begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{bmatrix}.$$

Essa matriz contém apenas um parâmetro, σ^2 , e supõe independência e homoscedasticidade para os componentes.

(iii) Componentes de variâncias com heteroscedasticidade – CV_H:

$$\begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_K^2 \end{bmatrix}.$$

Essa matriz contém K parâmetros, em que K denota a ordem da matriz. Assim como a estrutura CV, a CV_H supõe independência, entretanto considera heteroscedasticidade para os componentes.

(iv) Simetria composta – SC:

$$\sigma^2 \mathbf{I}_K + \sigma_1^2 \mathbf{J}_K = \begin{bmatrix} \sigma_1^2 + \sigma^2 & \sigma_1^2 & \dots & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 + \sigma^2 & \dots & \sigma_1^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_1^2 & \sigma_1^2 & \dots & \sigma_1^2 + \sigma^2 \end{bmatrix},$$

em que $\mathbf{J}_K$ é uma matriz de uns, quadrada, de dimensões $K \times K$. Essa matriz contém dois parâmetros, admite homogeneidade de variâncias e covariâncias constantes.

(v) Simetria composta com heteroscedasticidade – SC_H :

$$\begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 & \dots & \rho\sigma_1\sigma_K \\ \rho\sigma_1\sigma_2 & \sigma_2^2 & \dots & \rho\sigma_2\sigma_K \\ \vdots & \vdots & \ddots & \vdots \\ \rho\sigma_1\sigma_K & \rho\sigma_2\sigma_K & \dots & \sigma_K^2 \end{bmatrix}.$$

Essa matriz contém $K + 1$ parâmetros, admitindo variâncias e covariâncias distintas, mas correlações, ρ , iguais.

(vi) Auto regressiva de primeira ordem – AR(1):

$$\sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^K \\ \rho & 1 & \rho & \dots & \rho^{K-1} \\ \rho^2 & \rho & 1 & \dots & \rho^{K-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^K & \rho^{K-1} & \rho^{K-2} & \dots & 1 \end{bmatrix}.$$

Essa matriz contém dois parâmetros, σ^2 e ρ , sendo uma alternativa para Σ_i . Nesse caso, a correlação entre as observações decresce em função do intervalo de tempo entre as mesmas. Essa estrutura, no entanto, só deve ser utilizada se quaisquer observações subsequentes forem tomadas em intervalos de tempo iguais.

(vii) Auto regressiva de primeira ordem com heteroscedasticidade – AR_H(1):

$$\sigma^2 \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 & \rho^2\sigma_1\sigma_3 & \dots & \rho^K\sigma_1\sigma_K \\ \rho\sigma_1\sigma_2 & \sigma_2^2 & \rho^2\sigma_2\sigma_3 & \dots & \rho^{K-1}\sigma_2\sigma_K \\ \rho\sigma_1\sigma_3 & \rho\sigma_2\sigma_3 & \sigma_3^2 & \dots & \rho^{K-2}\sigma_3\sigma_K \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^K\sigma_1\sigma_K & \rho^{K-1}\sigma_2\sigma_K & \rho^{K-2}\sigma_3\sigma_K & \dots & \sigma_K^2 \end{bmatrix}.$$

O número de parâmetros dessa matriz é igual a $K + 2$, sendo uma generalização da matriz AR(1), em que são consideradas correlações decrescentes com o tempo e variâncias e covariâncias distintas.

2.4 Estimação dos componentes de variância

Segundo Laird e Ware (1982), os modelos lineares de efeitos mistos (LMEM) para dados longitudinais podem ser ajustados em dois estágios, sendo que no primeiro estágio é ajustado um modelo de regressão para cada indivíduo e no segundo é ajustado um modelo de regressão multivariada para captar a variabilidade entre indivíduos. No entanto, com o avanço da informática foi possível o desenvolvimento de algoritmos mais eficientes na estimação dos parâmetros de efeito fixo e dos componentes de variâncias dos LMEM, como por exemplo, os estimadores de máxima verossimilhança e os estimadores de máxima verossimilhança restrita.

Segundo Harville (1977), os estimadores de máxima verossimilhança são funções de estatísticas suficientes, além de serem consistentes, assintoticamente normais e eficientes, sendo que, geralmente, são obtidos sob suposição de normalidade. Por outro lado, esses estimadores podem ser viesados para pequenas amostras, além de não considerarem a perda de graus de liberdade resultante da estimação dos efeitos fixos do modelo.

Assim, um método alternativo para redução do viés dos estimadores de máxima verossimilhança é o método da máxima verossimilhança restrita proposto por (HAR-

VILLE, 1977). Esse método é baseado na máxima verossimilhança, sendo desenvolvido inicialmente para alguns modelos balanceados e, em 1962, estendido para todos os modelos balanceados, utilizado a pressuposição de normalidade (HARVILLE, 1977). A partir de 1990 o método foi utilizado para conjunto de dados longitudinais, sendo que Verbyla et al. (1999) aplicaram a um conjunto de dados balanceados e Littell (2002) e Spilke, Piepho e Hu (2005) aplicaram a conjuntos de dados desbalanceados.

É importante ressaltar, ainda, que os estimadores dos componentes de variância, pelo método REML, têm sido amplamente adotados, pois leva em consideração os graus de liberdade envolvidos na estimação dos parâmetros fixos do modelo, ou seja, as estimativas de REML tendem a ser menos viesadas que as estimativas de ML (HARVILLE, 1977; GUILMOUR; THOMPSON; CULLIS, 1995; PINHEIRO; BATES, 1995; DEMIDENKO, 2004; VERBEKE; MOLENBERGHS, 2009).

Para ambos os métodos, as suposições, com relação às variáveis aleatórias $\mathbf{Y}_i$ e $\mathbf{b}_i$, são conforme apresentadas em (2), dadas por:

$$\begin{aligned}\mathbf{Y}_i|\mathbf{b}_i &\sim \mathcal{N}(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i, \sigma^2\boldsymbol{\Lambda}_i) \\ \mathbf{b}_i &\sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{D}).\end{aligned}$$

Dessa forma, marginalmente, $\mathbf{Y}_i \stackrel{iid}{\sim} \mathcal{N}(\mathbf{X}_i\boldsymbol{\beta}, \mathbf{V}_{*i}(\boldsymbol{\gamma}))$, em que $\mathbf{V}_{*i}(\boldsymbol{\gamma}) = \sigma^2(\mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^T + \boldsymbol{\Lambda}_i)$.

Como se assumiu que os $\mathbf{Y}_i$ são independentes e normalmente distribuídos e $f(\mathbf{Y}_i, \mathbf{b}_i) = f(\mathbf{Y}_i|\mathbf{b}_i)f(\mathbf{b}_i)$. Logo, pode-se escrever a função de densidade conjunta de $\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_M$ como um produto das densidades marginais:

$$f(\mathbf{y}, \mathbf{X}; \boldsymbol{\beta}, \mathbf{V}_{*i}(\boldsymbol{\gamma})) = \prod_{i=1}^M \frac{1}{\sqrt{2\pi|\mathbf{V}_{*i}(\boldsymbol{\gamma})|}} \exp\left\{-\frac{1}{2}[\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}]^T \mathbf{V}_{*i}^{-1}(\boldsymbol{\gamma}) [\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}]\right\}, \quad (7)$$

em que $|\mathbf{V}_{*i}(\boldsymbol{\gamma})|$ representa o determinante da matriz $\mathbf{V}_{*i}(\boldsymbol{\gamma})$ e $\boldsymbol{\gamma}$ é o vetor de componentes de variâncias.

Assim, uma aproximação clássica para inferência é baseada em estimadores

obtidos a partir da maximização da função de verossimilhança marginal L_{ML} dada por:

$$L_{ML}(\boldsymbol{\theta}|\mathbf{y}, \mathbf{X}) = \prod_{i=1}^M \frac{1}{\sqrt{2\pi\sigma^2 |\mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T + \boldsymbol{\Lambda}_i|}} \times \exp \left[-\frac{1}{2} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})^T \frac{1}{\sigma^2} (\mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T + \boldsymbol{\Lambda}_i)^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}) \right],$$

ou de seu logaritmo,

$$l_{ML}(\boldsymbol{\theta}|\mathbf{y}, \mathbf{X}) = -\frac{1}{2} \left\{ M [\ln(2\pi) + \ln \sigma^2] + \sum_{i=1}^M \left[\ln |\mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T + \boldsymbol{\Lambda}_i| + \frac{1}{\sigma^2} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})^T (\mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T + \boldsymbol{\Lambda}_i)^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}) \right] \right\}.$$

em que $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \gamma^T)^T$ é o vetor de parâmetros desconhecidos. Mas $-\frac{M}{2} \ln(2\pi)$ não depende de $\boldsymbol{\theta}$. Logo, pode-se escrever:

$$l_{ML}(\boldsymbol{\theta}|\mathbf{y}, \mathbf{X}) \propto -\frac{1}{2} \left\{ M \ln \sigma^2 + \sum_{i=1}^M \left[\ln |\mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T + \boldsymbol{\Lambda}_i| + \frac{1}{\sigma^2} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})^T (\mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T + \boldsymbol{\Lambda}_i)^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}) \right] \right\}. \quad (8)$$

Portanto, a dimensão total do vetor de parâmetros desconhecidos $\boldsymbol{\theta}$ é $p+1+\gamma^*$, em que γ^* é a soma dos números de parâmetros das matrizes $\mathbf{D}_*$ e $\boldsymbol{\Sigma}_i$. Assim, a Estimativa de Máxima Verossimilhança (MLE) de $\boldsymbol{\theta}$, que maximiza a função l_{ML} , pertence, no caso em questão, ao espaço paramétrico $\boldsymbol{\Theta}$ dado por:

$$\boldsymbol{\Theta} = \{ \boldsymbol{\theta} : \boldsymbol{\beta} \in \mathbb{R}^{p+1} | \boldsymbol{\Sigma}_i \text{ e } \mathbf{D}_* \text{ são positivas definidas} \}.$$

Sejam: $\mathbf{e}_i = \mathbf{e}_i(\boldsymbol{\beta}) = \mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}$ o vetor de dimensão n_i , de erros para o i -ésimo indivíduo, com $i = 1, 2, \dots, M$ e $\mathbf{W}_i(\gamma) = (\mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T + \boldsymbol{\Lambda}_i)^{-1}$. Tem-se, conseqüentemente, que o logaritmo da função de verossimilhança, desconsiderando o termo constante, dadas em (8) pode ser reescrito como:

$$l_{ML}(\boldsymbol{\theta}|\mathbf{y}, \mathbf{X}) \propto -\frac{1}{2} \left\{ M \ln \sigma^2 + \sum_{i=1}^M \left[\ln |\mathbf{W}_i^{-1}(\boldsymbol{\gamma})| + \frac{1}{\sigma^2} \mathbf{e}_i^T \mathbf{W}_i(\boldsymbol{\gamma}) \mathbf{e}_i \right] \right\}, \quad (9)$$

para mais detalhes consultar o Apêndice B.

Note que, segundo Demidenko (2004), existem outras parametrizações equivalentes para o logaritmo função da verossimilhança como, por exemplo, a parametrização de redução da dimensão (*dimension-reduction parameterization*), a parametrização inversa $\mathbf{D}$ ou matriz de precisão (*Inverse $\mathbf{D}$ parameterization*) ou o perfil de verossimilhança (*profile-likelihood*).

Para estimar os componentes de variâncias, por outro lado, utiliza-se uma função de máxima verossimilhança restrita (REML). Considere a matriz $\mathbf{L} = [\mathbf{L}_1 \ \mathbf{L}_2]$ em que $\mathbf{L}_1$ e $\mathbf{L}_2$ têm dimensões $(N_T \times (p+1))$ e $(N_T \times (N_T - p - 1))$, respectivamente e satisfazem $\mathbf{L}_1^T \mathbf{X} = \mathbf{I}_{p+1}$ e $\mathbf{L}_2^T \mathbf{X} = \mathbf{0}$. De acordo com Knight (2008), os dados podem então ser transformados de $\mathbf{y}$ para $\mathbf{L}^T \mathbf{y}$ ou seja

$$\mathbf{L}^T \mathbf{y} = [\mathbf{L}_1 \ \mathbf{L}_2]^T \mathbf{y} = \begin{bmatrix} \mathbf{L}_1^T \mathbf{y} \\ \mathbf{L}_2^T \mathbf{y} \end{bmatrix} = \begin{bmatrix} \mathbf{y}_1^* \\ \mathbf{y}_2^* \end{bmatrix}.$$

Neste caso, pode-se demonstrar que a variável aleatória $\mathbf{y}^T = [\mathbf{y}_1^* \ \mathbf{y}_2^*]^T$ tem distribuição com esperança e variância, respectivamente, dada por

$$\begin{aligned} E[\mathbf{Y}^*] &= E[\mathbf{L}_1 \ \mathbf{L}_2]^T \mathbf{y} = E \begin{bmatrix} \mathbf{L}_1^T \mathbf{y} \\ \mathbf{L}_2^T \mathbf{y} \end{bmatrix} = E \begin{bmatrix} \mathbf{L}_1^T (\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\varepsilon}) \\ \mathbf{L}_2^T (\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\varepsilon}) \end{bmatrix} \\ &= E \begin{bmatrix} \mathbf{L}_1^T \mathbf{X}\boldsymbol{\beta} + \mathbf{L}_1^T (\mathbf{Z}\mathbf{b} + \boldsymbol{\varepsilon}) \\ \mathbf{L}_2^T \mathbf{X}\boldsymbol{\beta} + \mathbf{L}_2^T (\mathbf{Z}\mathbf{b} + \boldsymbol{\varepsilon}) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\beta} \\ \mathbf{0} \end{bmatrix} \end{aligned} \quad (10)$$

e

$$Var[\mathbf{Y}^*] = Var \begin{bmatrix} \mathbf{L}_1^T \mathbf{X}\boldsymbol{\beta} + \mathbf{L}_1^T (\mathbf{Z}\mathbf{b} + \boldsymbol{\varepsilon}) \\ \mathbf{L}_2^T \mathbf{X}\boldsymbol{\beta} + \mathbf{L}_2^T (\mathbf{Z}\mathbf{b} + \boldsymbol{\varepsilon}) \end{bmatrix} = \begin{bmatrix} \mathbf{L}_1^T \mathbf{V}(\boldsymbol{\gamma}) \mathbf{L}_1 & \mathbf{L}_1^T \mathbf{V}(\boldsymbol{\gamma}) \mathbf{L}_2 \\ \mathbf{L}_2^T \mathbf{V}(\boldsymbol{\gamma}) \mathbf{L}_1 & \mathbf{L}_2^T \mathbf{V}(\boldsymbol{\gamma}) \mathbf{L}_2 \end{bmatrix}. \quad (11)$$

Assim, resumidamente, utilizando os resultados de (10) e (11) tem-se que

$$\mathbf{Y}^* = \begin{bmatrix} \mathbf{Y}_1^* \\ \mathbf{Y}_2^* \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \boldsymbol{\beta} \\ \mathbf{0} \end{bmatrix}, \sigma^2 \begin{bmatrix} \mathbf{L}_1^T \mathbf{V}(\gamma) \mathbf{L}_1 & \mathbf{L}_1^T \mathbf{V}(\gamma) \mathbf{L}_2 \\ \mathbf{L}_2^T \mathbf{V}(\gamma) \mathbf{L}_1 & \mathbf{L}_2^T \mathbf{V}(\gamma) \mathbf{L}_2 \end{bmatrix} \right). \quad (12)$$

e a partir da mesma, podem-se obter as distribuições marginal de $\mathbf{Y}_2^*$ e condicional de $\mathbf{Y}_1^*|\mathbf{Y}_2^*$. A primeira, segundo Knight (2008) e Alcarde (2012), é utilizada para a estimação dos componentes de variâncias em $\mathbf{V}(\gamma)$, tendo como base a verossimilhança restrita e a segunda, para a estimação dos $\boldsymbol{\beta}$.

Assim, como o vetor $\mathbf{Y}_2^*$ tem distribuição normal com vetor de médias zero e matriz de variâncias e covariâncias $\sigma^2 (\mathbf{L}_2^T \mathbf{V}(\gamma) \mathbf{L}_2)$, ou seja, $\mathbf{Y}_2^* \sim \mathcal{N}(\mathbf{0}, \sigma^2 (\mathbf{L}_2^T \mathbf{V}(\gamma) \mathbf{L}_2))$, então, $\mathbf{Y}_2^*$ não depende do vetor de parâmetros $\boldsymbol{\beta}$. Logo, a função de verossimilhanças da matriz de contrastes dos erros pode ser escrita, segundo Harville (1976), como:

$$L_R(\gamma) = (2\pi)^{-\frac{(N_T-p-1)}{2}} \left| \sum_{i=1}^M \mathbf{X}_i^T \mathbf{X}_i \right|^{\frac{1}{2}} \left| \sum_{i=1}^M \mathbf{X}_i^T \mathbf{V}_{*i}^{-1}(\gamma) \mathbf{X}_i \right|^{-\frac{1}{2}} \prod_{i=1}^M |\mathbf{V}_{*i}(\gamma)|^{-\frac{1}{2}} \\ \times \exp \left\{ -\frac{1}{2} \sum_{i=1}^M \left(\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}(\gamma) \right)^T \mathbf{V}_{*i}^{-1}(\gamma) \left(\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}(\gamma) \right) \right\}$$

em que $\hat{\boldsymbol{\beta}}$ é o estimador de $\boldsymbol{\beta}$ que é obtido maximizando (9), condicional a γ .

Por consequência, o logaritmo da função de verossimilhança restrita ($l_R(\gamma)$) tem como termo constante

$$-\frac{1}{2} \left[(N_T - p - 1) \ln(2\pi) + \ln \left| \sum_{i=1}^M \mathbf{X}_i^T \mathbf{X}_i \right| + \ln \frac{1}{\sigma^2} \right].$$

Logo, sabendo que $\mathbf{V}_i(\gamma) = \sigma^2 \mathbf{W}_i^{-1}(\gamma)$, tem-se que

$$l_R(\gamma) \propto -\frac{1}{2} \left\{ \ln \left| \sum_{i=1}^M \mathbf{X}_i^T \mathbf{W}_i(\gamma) \mathbf{X}_i \right| + M \ln \sigma^2 + \sum_{i=1}^M \left[\ln |\mathbf{W}_i^{-1}(\gamma)| \right. \right. \\ \left. \left. + \frac{1}{\sigma^2} \left(\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}(\gamma) \right)^T \mathbf{W}_i(\gamma) \left(\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}(\gamma) \right) \right] \right\}.$$

Finalmente, note que se pode escrever $l_R(\gamma)$ da seguinte forma:

$$l_R(\gamma) \propto -\frac{1}{2} \ln \left| \sum_{i=1}^M \mathbf{X}_i^T \mathbf{W}_i(\gamma) \mathbf{X}_i \right| + l_{ML}(\hat{\boldsymbol{\beta}}, \gamma) \quad (13)$$

em que $l_{ML}(\hat{\boldsymbol{\beta}}, \boldsymbol{\gamma}) = l_{ML}(\boldsymbol{\theta})$ é a função logaritmo da verossimilhança dada em (9). Segundo Laird e Ware (1982), como $\ln \left| \sum_{i=1}^M \mathbf{X}_i^T \mathbf{W}_i(\boldsymbol{\gamma}) \mathbf{X}_i \right|$ em (13) não depende de nenhum $\boldsymbol{\beta}$, tem-se que o estimador de RMLE para $\boldsymbol{\gamma}$ e $\boldsymbol{\beta}$ pode ser obtido maximizando o logaritmo da função de verossimilhança restrita

$$l_{REML}(\boldsymbol{\theta}) = -\frac{1}{2} \ln \left| \sum_{i=1}^M \mathbf{X}_i^T \mathbf{W}_i(\boldsymbol{\gamma}) \mathbf{X}_i \right| + l_{ML}(\boldsymbol{\theta}) \quad (14)$$

com respeito a todos os parâmetros ($\boldsymbol{\gamma}$ e $\boldsymbol{\beta}$) simultaneamente.

Seja $\boldsymbol{\gamma}^T = (\boldsymbol{\kappa}^T, \boldsymbol{\phi}^T)$ o vetor de componentes de variâncias, tal que $\boldsymbol{\kappa}$ contém os componentes de variâncias associados a $\mathbf{b}$ e $\boldsymbol{\phi}$ os componentes de variâncias associados a $\boldsymbol{\varepsilon}$, sua estimação é baseada em (14). Segundo Alcarde (2012), a estimativa REML de γ_l , com $l = 1, 2, \dots, L$, tal que $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \dots, \gamma_L)$, é obtida igualando-se a função escore

$$U(\gamma_l) = \frac{\partial l_{REML}}{\partial \gamma_l} = -\frac{1}{2} \left[\text{tr} \left(\mathbf{P} \frac{\partial \mathbf{V}_*}{\partial \gamma_l} \right) - \mathbf{y}^T \mathbf{P} \frac{\partial \mathbf{V}_*}{\partial \gamma_l} \mathbf{P} \mathbf{y} \right] \quad (15)$$

a zero, sendo que $\mathbf{P} = \mathbf{V}_*^{-1} - \mathbf{V}_*^{-1} \mathbf{X} (\mathbf{X}^T \mathbf{V}_*^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}_*^{-1}$.

Note que os elementos da matriz de informação observada são

$$\begin{aligned} -\frac{\partial^2 l_{REML}}{\partial \gamma_l \partial \gamma_k} &= \frac{1}{2} \text{tr} \left(\mathbf{P} \frac{\partial^2 \mathbf{V}_*}{\partial \gamma_l \partial \gamma_k} \right) - \frac{1}{2} \text{tr} \left(\mathbf{P} \frac{\partial \mathbf{V}_*}{\partial \gamma_l} \mathbf{P} \frac{\partial \mathbf{V}_*}{\partial \gamma_k} \right) + \\ &+ \mathbf{y}^T \mathbf{P} \frac{\partial \mathbf{V}_*}{\partial \gamma_l} \mathbf{P} \frac{\partial \mathbf{V}_*}{\partial \gamma_k} \mathbf{P} \mathbf{y} - \frac{1}{2} \mathbf{y}^T \mathbf{P} \frac{\partial^2 \mathbf{V}_*}{\partial \gamma_l \partial \gamma_k} \mathbf{P} \mathbf{y} \end{aligned} \quad (16)$$

e os elementos da matriz de informação esperada são

$$E \left(-\frac{\partial^2 l_{REML}}{\partial \gamma_l \partial \gamma_k} \right) = \frac{1}{2} \text{tr} \left(\mathbf{P} \frac{\partial \mathbf{V}_*}{\partial \gamma_l} \mathbf{P} \frac{\partial \mathbf{V}_*}{\partial \gamma_k} \right). \quad (17)$$

No entanto, a solução para $U(\gamma_l) = 0$ exige um algoritmo iterativo que utiliza uma estimativa inicial ($\boldsymbol{\gamma}^{(0)}$) para $\boldsymbol{\gamma}$ e uma atualização ($\boldsymbol{\gamma}^{(1)}$). Um desses algoritmos é o “Escore de Fisher”, em que:

$$\boldsymbol{\gamma}^{(1)} = \boldsymbol{\gamma}^{(0)} + \mathbf{I}(\boldsymbol{\gamma}^{(0)}, \boldsymbol{\gamma}^{(0)})^{-1} \mathbf{U}(\boldsymbol{\gamma}^{(0)}), \quad (18)$$

em que $\mathbf{U}(\boldsymbol{\gamma}^{(0)})$ é o vetor escore, apresentado na equação (15), e $\mathbf{I}(\boldsymbol{\gamma}^{(0)}, \boldsymbol{\gamma}^{(0)})$ representa a matriz de informação esperada de $\boldsymbol{\gamma}$, apresentada na equação (17), avaliada em

$\gamma^{(0)}$ (ALCARDE, 2012).

À medida em que as dimensões da matriz a ser invertida aumentam, a utilização do algoritmo “Escore de Fisher” pode levar a dificuldades computacionais, as quais são discutidas em Gilmour, Thompsom e Cullis (1995). Uma alternativa é a utilização do algoritmo AI que apresenta propriedades de convergência semelhantes ao algoritmo “Escore de Fisher”, evitando a sobrecarga computacional (GILMOUR; THOMPSON; CULLIS, 1995). Este, segundo os autores, é obtido substituindo a matriz de informação esperada $\mathbf{I}(\gamma_l, \gamma_k)$ por $\mathcal{I}_A(\gamma_l, \gamma_k)$, cujos elementos são dados por:

$$\mathcal{I}_A(\gamma_l, \gamma_k) = \frac{1}{2} \mathbf{y}^T \mathbf{P} \frac{\partial \mathbf{V}_*}{\partial \gamma_l} \mathbf{P} \frac{\partial \mathbf{V}_*}{\partial \gamma_k} \mathbf{P} \mathbf{y}$$

que é obtido pela média de (16) e de (17). Nos casos em que $\frac{\partial^2 \mathbf{V}_*}{\partial \gamma_l \partial \gamma_k} \neq 0$, deve-se, segundo Gilmour, Thompsom e Cullis (1995), aproximar $\mathbf{y}^T \mathbf{P} \frac{\partial^2 \mathbf{V}_*}{\partial \gamma_l \partial \gamma_k} \mathbf{P} \mathbf{y}$ por sua esperança, dada pelo $\text{tr} \left(\mathbf{P} \frac{\partial^2 \mathbf{V}_*}{\partial \gamma_l \partial \gamma_k} \right)$.

2.5 Estimação dos termos fixos e predição dos termos aleatórios

O melhor estimador linear não viesado (BLUE) do vetor de parâmetros de efeitos fixos (β) e o melhor preditor linear não viesado (BLUP) do vetor de parâmetros de efeitos aleatórios $\mathbf{b}_i$ são considerados “melhores” pois minimizam a variância amostral, “linear” por serem funções lineares de $\mathbf{y}$ e “não viesados” porque $E[\text{BLUE}(\beta)] = \beta$ e $E[\text{BLUP}(\mathbf{b})] = \mathbf{b}$ (WALSH, 2010).

No entanto, para que $E[\text{BLUE}(\beta)] = \beta$ e $E[\text{BLUP}(\mathbf{b})] = \mathbf{b}$ todos os parâmetros das matrizes de variâncias e covariâncias devem ser conhecidos, de forma que β é tratado como um vetor de efeitos fixos e $\mathbf{b}_i$ é tratado como um vetor de efeitos aleatórios para o indivíduo i , sendo as expressões para as estimativas dos efeitos populacionais, β , e individuais, $\mathbf{b}_i$, e respectivos erros padrões também conhecidos (LAIRD; WARE, 1982).

Assim, no item 2.5.1 é descrito o melhor estimador linear não viesado (BLUE) do vetor de parâmetros de efeitos fixos, enquanto que no item 2.5.2 é descrito o melhor estimador linear não viesado empírico (EBLUE). No item 2.5.3 são descritos o melhor preditor linear não viesado (BLUP) e o melhor preditor linear não viesado empírico (EBLUP). Por fim, no item 2.5.4 é descrito o processo de estimação dos efeitos fixos e predição dos efeitos aleatórios simultaneamente.

2.5.1 Componentes de variância conhecidos

Considerando o LMEM dado na expressão (1) têm-se que $E[\mathbf{b}_i] = \mathbf{0}$, $E[\boldsymbol{\varepsilon}_i] = \mathbf{0}$, $cov(\mathbf{b}_i) = \sigma^2 \mathbf{D}$, $cov(\boldsymbol{\varepsilon}_i) = \sigma^2 \boldsymbol{\Lambda}_i$ e $cov(\mathbf{b}_i, \boldsymbol{\varepsilon}_i) = \boldsymbol{\Phi}$ e escrevendo $\boldsymbol{\eta}_i = \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\varepsilon}_i$, o modelo linear de efeitos fixos para o indivíduo i é dado por:

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \boldsymbol{\eta}_i, \quad (19)$$

com $E[\boldsymbol{\eta}_i] = \mathbf{0}$ e $cov(\boldsymbol{\eta}_i) = cov(\mathbf{y}_i) = \sigma^2 (\mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T + \boldsymbol{\Lambda}_i)$.

Segundo Laird e Ware (1982) e Haslett e Puntanen (2011), assume-se que a matriz de covariâncias escalonada $\mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T + \boldsymbol{\Lambda}_i$ é positiva definida e que $\mathbf{X}_i$ possui posto coluna completo ($posto(\mathbf{X}_i) = p + 1 < n_i$). Adicionalmente, considerando a matriz de covariâncias do modelo apresentado na expressão (19), que é dada por $\mathbf{V}_{*i}(\boldsymbol{\gamma}) = \sigma^2 \mathbf{W}_i^{-1}(\boldsymbol{\gamma})$, tem-se que o melhor estimador linear não viesado (BLUE) de $\boldsymbol{\beta}$ é obtido maximizando (9) condicional a $\boldsymbol{\gamma}$, sendo dado por

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \left(\sum_{i=1}^M \mathbf{X}_i^T \mathbf{V}_{*i}^{-1}(\boldsymbol{\gamma}) \mathbf{X}_i \right)^{-1} \sum_{i=1}^M \mathbf{X}_i^T \mathbf{V}_{*i}^{-1}(\boldsymbol{\gamma}) \mathbf{y}_i \\ &= \left(\sum_{i=1}^M \mathbf{X}_i^T \mathbf{W}_i(\boldsymbol{\gamma}) \mathbf{X}_i \right)^{-1} \sum_{i=1}^M \mathbf{X}_i^T \mathbf{W}_i(\boldsymbol{\gamma}) \mathbf{y}_i. \end{aligned} \quad (20)$$

Assume-se aqui que é necessária a existência da matriz inversa. Segundo Laird e Ware (1982), fórmulas equivalentes utilizando inversa generalizada podem ser trabalhadas nos casos em que o posto da matriz não é completo.

A estimativa de máxima verossimilhança para o vetor de parâmetros de efeito fixo $\boldsymbol{\beta}$ é baseada na distribuição marginal dada em (7) e na estimativa não viesada de variância mínima.

2.5.2 Variância desconhecida

Para o caso em que os componentes de variância são desconhecidos, deve-se então, primeiramente, obter as estimativas para os componentes de variâncias por meio do seguinte estimador

$$\hat{\mathbf{V}}_{*i} = \mathbf{V}_{*i}(\hat{\boldsymbol{\gamma}}) = \hat{\sigma}^2 \mathbf{W}_i^{-1}(\hat{\boldsymbol{\gamma}}). \quad (21)$$

A partir da expressão (21), pode-se obter, então, o melhor estimador linear não viesado empírico (EBLUE) para os parâmetros de efeitos fixos, que é dado por:

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= \left(\sum_{i=1}^M \mathbf{X}_i^T \hat{\mathbf{V}}_{*i}^{-1} \mathbf{X}_i \right)^{-1} \sum_{i=1}^M \mathbf{X}_i^T \hat{\mathbf{V}}_{*i}^{-1} \mathbf{y}_i \\ &= \left(\sum_{i=1}^M \mathbf{X}_i^T \mathbf{W}_i(\hat{\boldsymbol{\gamma}}) \mathbf{X}_i \right)^{-1} \sum_{i=1}^M \mathbf{X}_i^T \mathbf{W}_i(\hat{\boldsymbol{\gamma}}) \mathbf{y}_i,\end{aligned}\quad (22)$$

sendo $Var(\hat{\boldsymbol{\beta}}) = \hat{\sigma}^2 \left[\sum_{i=1}^M \mathbf{X}_i^T \mathbf{W}_i(\hat{\boldsymbol{\gamma}}) \mathbf{X}_i \right]^{-1}$.

2.5.3 Predição dos efeitos aleatórios

Segundo Harville (1976) e Laird e Ware (1982), o preditor do vetor de efeitos aleatórios $\mathbf{b}_i$ não é obtido por máxima verossimilhança, mas pode ser derivado por meio de uma extensão do teorema de Gauss-Markov ou obtido por método bayesiano empírico a partir de informações a priori da média dos efeitos aleatórios. Assim, o melhor preditor linear não viesado (BLUP) é dado por

$$\tilde{\mathbf{b}}_i = \mathbf{DZ}_i^T \mathbf{W}_i(\boldsymbol{\gamma}) \left(\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}} \right). \quad (23)$$

Como $\hat{\boldsymbol{\beta}}$ e $\tilde{\mathbf{b}}_i$ são funções lineares de $\mathbf{y}$, segundo Laird e Ware (1982), as expressões para os respectivos erros padrões são:

$$Var(\hat{\boldsymbol{\beta}}) = \sigma^2 \left[\sum_{i=1}^M \mathbf{X}_i^T \mathbf{W}_i(\boldsymbol{\gamma}) \mathbf{X}_i \right]^{-1} \quad (24)$$

e

$$Var(\tilde{\mathbf{b}}_i) = \sigma^2 \mathbf{DZ}_i^T \left\{ \mathbf{W}_i(\boldsymbol{\gamma}) - \mathbf{W}_i(\boldsymbol{\gamma}) \mathbf{X}_i \left[\sum_{i=1}^M \mathbf{X}_i^T \mathbf{W}_i(\boldsymbol{\gamma}) \mathbf{X}_i \right]^{-1} \mathbf{X}_i^T \mathbf{W}_i(\boldsymbol{\gamma}) \right\} \mathbf{Z}_i \mathbf{D}. \quad (25)$$

Caso (25) seja utilizada para avaliar a estimativa do erro, Laird e Ware (1982) ressaltam que uma variação em $\tilde{\mathbf{b}}_i - \mathbf{b}_i$ pode ser subestimada porque essa expressão ignora

a variação de $\mathbf{b}_i$. Assim, a correção é dada por:

$$\begin{aligned} Var\left(\tilde{\mathbf{b}}_i - \mathbf{b}_i\right) &= \sigma^2 \left[\mathbf{D} - \mathbf{D}\mathbf{Z}_i^T \mathbf{W}_i(\gamma) \mathbf{Z}_i \mathbf{D} \right. \\ &\quad \left. + \mathbf{D}\mathbf{Z}_i^T \mathbf{W}_i(\gamma) \mathbf{X}_i \left(\sum_{i=1}^M \mathbf{X}_i^T \mathbf{W}_i(\gamma) \mathbf{X}_i \right)^{-1} \mathbf{X}_i^T \mathbf{W}_i(\gamma) \mathbf{Z}_i \mathbf{D} \right]. \end{aligned} \quad (26)$$

Segundo Harville (1976), quando $\mathbf{b}_i$ têm distribuição normal têm-se que

$$E\left[\mathbf{b}_i \mid \left(\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}\right)\right] = E\left(\mathbf{b}_i \mid \tilde{\mathbf{b}}_i\right) = \tilde{\mathbf{b}}_i$$

com probabilidade 1 e que

$$Var\left[\mathbf{b}_i \mid \left(\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}\right)\right] = Var\left(\mathbf{b}_i \mid \tilde{\mathbf{b}}_i\right) = Var\left(\tilde{\mathbf{b}}_i - \mathbf{b}_i\right).$$

Portanto, a distribuição do vetor de efeitos aleatórios $\mathbf{b}_i$ condicional ao vetor de resíduos $\left(\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}\right)$ tem distribuição idêntica ao vetor de efeitos aleatórios $\mathbf{b}_i$ condicional a q funções lineares dos resíduos que compreendem o BLUP de $\mathbf{b}_i$ (HARVILLE, 1976).

Quando, no entanto, os componentes de variâncias não são conhecidos podem ser utilizadas as estimativas de γ calculadas a partir do estimador dado na expressão (21), sendo que o preditor resultante é conhecido como EBLUP (melhor preditor linear não viesado empírico). Deve-se ressaltar, ainda, que o método da máxima verossimilhança restrita é mais indicado para estimar os componentes de variância (γ), uma vez que o viés é menor em relação ao método da máxima verossimilhança (PINHEIRO; BATES, 1995; VERBEKE; MOLENBERGHS, 2009).

2.5.4 Estimação dos efeitos fixos e predição dos efeitos aleatórios simultaneamente

Considerando os modelos mistos apresentados em (2), tem-se que a distribuição conjunta de $\mathbf{Y}_i$ e $\mathbf{b}_i$ condicional a γ é dada por:

$$\begin{aligned}
f(\mathbf{Y}, \mathbf{b}|\boldsymbol{\gamma}) &= \prod_{i=1}^M f(\mathbf{Y}_i|\mathbf{b}_i, \boldsymbol{\gamma}) \times f(\mathbf{b}_i|\boldsymbol{\gamma}) \\
&= \prod_{i=1}^M \left\{ (2\pi)^{-\frac{1}{2}} |\boldsymbol{\Sigma}_i|^{-\frac{1}{2}} \exp \left[-\frac{1}{2} (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i)^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i) \right] \right. \\
&\quad \times (2\pi)^{-\frac{1}{2}} |\mathbf{D}_*|^{-\frac{1}{2}} \exp \left[-\frac{1}{2} \mathbf{b}_i^T \mathbf{D}_*^{-1} \mathbf{b}_i \right] \left. \right\} \\
&= \prod_{i=1}^M \left\{ (2\pi)^{-1} |\boldsymbol{\Sigma}_i|^{-\frac{1}{2}} |\mathbf{D}_*|^{-\frac{1}{2}} \exp \left[-\frac{1}{2} \left((\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i)^T \times \right. \right. \right. \\
&\quad \times \boldsymbol{\Sigma}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i) + \mathbf{b}_i^T \mathbf{D}_*^{-1} \mathbf{b}_i \left. \left. \right) \right] \left. \right\}. \tag{27}
\end{aligned}$$

Aplicando o logaritmo na distribuição conjunta (27) têm-se que ela é proporcional a $l(\mathbf{Y}, \mathbf{b}|\boldsymbol{\gamma})$ e é dada por:

$$\begin{aligned}
l(\mathbf{Y}, \mathbf{b}|\boldsymbol{\gamma}) \propto & -\frac{1}{2} \left\{ M \ln |\mathbf{D}_*| + \sum_{i=1}^M \ln |\boldsymbol{\Sigma}_i| + \right. \\
& \left. + \sum_{i=1}^M \left[(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i)^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i) + \mathbf{b}_i^T \mathbf{D}_*^{-1} \mathbf{b}_i \right] \right\}
\end{aligned}$$

assim, as derivadas parciais, com relação a $\boldsymbol{\beta}$ e $\mathbf{b}$, são

$$\begin{bmatrix} \frac{\partial l}{\partial \boldsymbol{\beta}} \\ \frac{\partial l}{\partial \mathbf{b}} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^M (\mathbf{X}_i^T \boldsymbol{\Sigma}_i^{-1} \mathbf{y}_i - \mathbf{X}_i^T \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i \boldsymbol{\beta} - \mathbf{X}_i^T \boldsymbol{\Sigma}_i^{-1} \mathbf{Z}_i \mathbf{b}_i) \\ \sum_{i=1}^M (\mathbf{Z}_i^T \boldsymbol{\Sigma}_i^{-1} \mathbf{y}_i - \mathbf{Z}_i^T \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z}_i^T \boldsymbol{\Sigma}_i^{-1} \mathbf{Z}_i \mathbf{b}_i - \mathbf{D}_*^{-1} \mathbf{b}_i) \end{bmatrix} \tag{28}$$

Segundo Henderson (1953), quando as equações apresentadas em (28) são igualadas a zero, as equações de estimação do vetor de parâmetros $\boldsymbol{\beta}$ e predição do vetor de parâmetros $\mathbf{b}$ são dadas por:

$$\begin{bmatrix} \sum_{i=1}^M (\mathbf{X}_i^T \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i) & \sum_{i=1}^M (\mathbf{X}_i^T \boldsymbol{\Sigma}_i^{-1} \mathbf{Z}_i) \\ \sum_{i=1}^M (\mathbf{Z}_i^T \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i) & \sum_{i=1}^M (\mathbf{Z}_i^T \boldsymbol{\Sigma}_i^{-1} \mathbf{Z}_i + \mathbf{D}_*^{-1}) \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \tilde{\mathbf{b}} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^M (\mathbf{X}_i^T \boldsymbol{\Sigma}_i^{-1} \mathbf{y}_i) \\ \sum_{i=1}^M (\mathbf{Z}_i^T \boldsymbol{\Sigma}_i^{-1} \mathbf{y}_i) \end{bmatrix}$$

Dessa forma, as estimativas de $\boldsymbol{\beta}$ e predição de $\mathbf{b}_i$ são dadas, respectivamente,

por:

$$\begin{aligned} \hat{\beta} = & \left\{ \mathbf{X}^T \left[\Sigma^{-1} - \Sigma^{-1} \mathbf{Z} (\mathbf{Z}^T \Sigma^{-1} \mathbf{Z} + \mathbf{D}_*^{-1})^{-1} \mathbf{Z}^T \Sigma^{-1} \right] \mathbf{X} \right\} \times \\ & \times \mathbf{X}^T \left[\Sigma^{-1} - \Sigma^{-1} \mathbf{Z} (\mathbf{Z}^T \Sigma^{-1} \mathbf{Z} + \mathbf{D}_*^{-1})^{-1} \mathbf{Z}^T \Sigma^{-1} \right] \mathbf{y} \end{aligned} \quad (29)$$

e

$$\tilde{\mathbf{b}}_i = (\mathbf{Z}_i^T \Lambda_i^{-1} \mathbf{Z}_i + \mathbf{D}^{-1})^{-1} \mathbf{Z}_i^T \Lambda_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \hat{\beta}) \quad (30)$$

que exigem solução para $\gamma^T = (\kappa^T, \phi^T)$. Na prática, substituem-se κ^T e ϕ^T por suas estimativas REML, $\hat{\kappa}$ e $\hat{\phi}$.

O estimador (29) é equivalente ao estimador (20), que é o melhor estimador não viesado de β e o preditor (30) é equivalente ao (23) que é o melhor preditor não viesado de $\mathbf{b}_i$.

2.6 Seleção de modelos

O principal objetivo da seleção de modelos é buscar um modelo parcimonioso, ou seja, um modelo que ao mesmo tempo seja o que melhor prediz a variável resposta e o que possua o menor número de parâmetros. As duas principais opções para seleção de modelos nos casos de LMEM são os critérios de informação (Critério de Informação de Akaike (AIC) e Critério de Informação Bayesiano (BIC) e os testes de hipótese, como o teste da razão de verossimilhanças.

2.6.1 Critérios de informação

O AIC, proposto por Akaike (1974), é definido como duas vezes a diferença entre o valor da verossimilhança L , que é uma medida de ajuste, e o número de parâmetros em θ , que penaliza o modelo de acordo com sua complexidade. Já para o BIC, proposto por Schwarz (1978), leva em conta o número de observações, sendo mais indicado para conjuntos maiores de dados (DEMIDENKO, 2004).

Seja $\hat{\theta}$ o vetor de estimativas dos $\omega = p + 1 + \gamma^*$ parâmetros, N_T o número total de observações, L_{ML} a função de verossimilhança e L_{REML} a função de verossimilhança restrita.

Quando é utilizada a função de verossimilhança, as expressões para o cálculo dos critérios AIC e BIC são, respectivamente, dados por:

$$AIC = -2L_{ML}(\hat{\boldsymbol{\theta}}) + 2\omega$$

e

$$BIC = -2L_{ML}(\hat{\boldsymbol{\theta}}) + 2\omega \ln N_T$$

Por outro lado, quando se utiliza a função de verossimilhança restrita, as expressões são dadas por:

$$AIC = -2L_{REML}(\hat{\boldsymbol{\theta}}) + 2\omega$$

e

$$BIC = -2L_{REML}(\hat{\boldsymbol{\theta}}) + 2\omega \ln (N_T - \omega)$$

em que $N_T > \omega$. Dessa forma, tanto o AIC quanto o BIC baseados na REML não são comparáveis com os baseados na ML.

Para todos os casos, deve-se escolher o modelo que minimiza o critério de informação AIC ou BIC.

2.6.2 Teste da razão de verossimilhanças

O método da razão de verossimilhanças para testar uma determinada hipótese está associada ao estimador de máxima verossimilhança. Esse teste é baseado na comparação dos valores das funções de verossimilhança para dois modelos aninhados, tal que um modelo é mais simples (modelo aninhado) e o outro modelo é mais complexo (modelo de referência). Seja $\boldsymbol{\Theta}$ o espaço paramétrico do modelo de referência (maximização irrestrita) e $\boldsymbol{\Theta}_0$ o espaço paramétrico do modelo aninhado (maximização restrita), então tem-se que $\boldsymbol{\Theta}_0$ pertence a um subespaço de $\boldsymbol{\Theta}$. Assim, segundo Casella e Berger (2011), pode-se considerar que o estimador de máxima verossimilhança de $\boldsymbol{\theta}$ é obtido a partir da maximização irrestrita da função de verossimilhança, enquanto que o estimador de máxima verossimilhança de $\boldsymbol{\theta}_0$

é obtido a partir de uma maximização restrita. Para isso, assume-se que Θ_0 é o espaço paramétrico de θ_0 . Dessa forma, o teste de hipótese associado é dado por:

$$H_0 : \theta \in \Theta_0$$

$$H_1 : \theta \in \Theta_0^c$$

em que Θ_0^c é o espaço paramétrico complementar. Em outras palavras, $\hat{\theta}_0$ é o vetor de valores de $\theta \in \Theta_0$ que maximizam conjuntamente a função de verossimilhanças (CASELLA; BERGER, 2011). Assim, pode-se obter a estatística do teste da razão de verossimilhanças, que é dada por:

$$2 \log \left(\frac{L_2}{L_1} \right) = 2 [\log (L_2) - \log (L_1)]$$

em que L_1 representa a função de verossimilhança do modelo aninhado ou de maximização restrita e L_2 , a função de verossimilhança do modelo de referência ou de maximização irrestrita.

Sob determinadas condições de regularidade, a estatística teste de razão de verossimilhanças segue, assintoticamente, a distribuição χ^2 , em que o número de graus de liberdade (GL) é obtido subtraindo-se o número de parâmetros do modelo aninhado do número de parâmetros do modelo de referência. Vale ressaltar que a estatística teste da razão de verossimilhanças segue uma mistura de χ^2 quando os parâmetros encontram-se na fronteira do espaço paramétrico (ALCARDE, 2012 apud SELF; LIANG, 1987).

2.7 Diagnósticos

Diferentemente dos modelos de regressão linear que apresentam apenas uma fonte de variabilidade do resíduo, os LMEM possuem duas fontes diferentes de variabilidade, uma entre os indivíduos e outra dentro do mesmo indivíduo, tornando a análise mais complexa.

De acordo com Hilden-Minton (1995), existem três tipos de perturbações nos resíduos do modelo apresentado na expressão (1) que são:

i) Resíduo do modelo marginal

$$\hat{\boldsymbol{\eta}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{V}(\boldsymbol{\gamma}) \mathbf{Q}\mathbf{y},$$

em que $\mathbf{Q} = \mathbf{V}^{-1}(\boldsymbol{\gamma}) (\mathbf{I} - \mathbf{T})$, $\mathbf{T} = \mathbf{X} (\mathbf{X}^T \mathbf{V}^{-1}(\boldsymbol{\gamma}) \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1}(\boldsymbol{\gamma})$ e $\mathbf{V}(\boldsymbol{\gamma}) = (\mathbf{ZDZ}^T + \boldsymbol{\Lambda})$.

Assim, o resíduo marginal do modelo dado pela expressão (4) prediz o erro marginal, sendo $\boldsymbol{\eta} = \mathbf{y} - E[\mathbf{y}] = \mathbf{y} - \mathbf{X}\boldsymbol{\beta} = \mathbf{Zb} + \boldsymbol{\varepsilon}$. Mais detalhes são apresentados no Apêndice C;

ii) Resíduo do modelo condicional

$$\hat{\boldsymbol{\varepsilon}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{Z}\tilde{\mathbf{b}} = \boldsymbol{\Lambda}\mathbf{Q}\mathbf{y},$$

que predizem o erro condicional do modelo dado pela expressão (3), sendo $\boldsymbol{\varepsilon} = \mathbf{y} - E[\mathbf{y}|\mathbf{b}] = \mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Zb}$;

iii) Melhor preditor linear não viesado empírico (EBLUP), $\mathbf{Z}\tilde{\mathbf{b}}$, que predizem os efeitos aleatórios.

Segundo Nobre e Singer (2007), cada tipo de resíduo é útil para validar uma certa suposição do modelo apresentado na expressão (1). Para tanto, de acordo com Hilden-Minton (1995), resíduos que dependem de outros tipos de erros são chamados de resíduos confundidos, como é o caso do resíduo condicional e do $\mathbf{Z}\tilde{\mathbf{b}}$. Os autores consideram, ainda, que um resíduo é considerado puro quando ele depende apenas dos componentes fixos e do erro que é suposto para predizer. Com isso, considerando-se as expressões dadas em (3), (20) e (23), têm-se que

$$\hat{\boldsymbol{\eta}} = [\mathbf{I} - \mathbf{X} (\mathbf{X}^T \mathbf{V}^{-1}(\boldsymbol{\gamma}) \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1}(\boldsymbol{\gamma})] \boldsymbol{\eta}$$

é considerado um resíduo puro, enquanto que

$$\hat{\boldsymbol{\varepsilon}} = \boldsymbol{\Lambda}\mathbf{Q}\boldsymbol{\varepsilon} + \boldsymbol{\Lambda}\mathbf{Q}\mathbf{Zb}$$

e

$$\mathbf{Z}\tilde{\mathbf{b}} = \mathbf{ZDZ}^T \mathbf{Q}\mathbf{Zb} + \mathbf{ZDZ}^T \mathbf{Q}\boldsymbol{\varepsilon}$$

são considerados resíduos confundidos.

Dessa forma, tem-se que o resíduo $\widehat{\boldsymbol{\varepsilon}}$ e $\mathbf{Z}\widetilde{\mathbf{b}}$ são confundidos com $\mathbf{b}$ e $\boldsymbol{\varepsilon}$, respectivamente. Segundo Nobre e Singer (2007), isso implica, por exemplo, que $\widehat{\boldsymbol{\varepsilon}}$ pode não ser adequado para verificar a normalidade de $\boldsymbol{\varepsilon}$ quando $\mathbf{b}$ não possui normalidade, ou seja, $\widehat{\boldsymbol{\varepsilon}}$ pode não apresentar normalidade quando na verdade ele é normal.

De acordo com Nobre e Singer (2007), o resíduo marginal pode ser empregado para verificar a linearidade de $\mathbf{y}$ em relação a $\mathbf{X}$, para isso deve-se fazer o gráfico dos resíduos marginais estimados $\widehat{\boldsymbol{\eta}}$ *versus* às variáveis explanatórias em $\mathbf{X}$, sendo que um comportamento aleatório ao redor do zero é esperado quando existe uma relação linear. Adicionalmente, o resíduo marginal também pode ser utilizado para verificar a validade da estrutura de covariância dentro do grupo ou indivíduo, por exemplo, $\mathbf{V}_{*i}(\boldsymbol{\gamma}) = \sigma^2 (\mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T + \boldsymbol{\Lambda}_i)$.

Em relação ao resíduo condicional, Pinheiro e Bates (1995) consideram o gráfico dos elementos de $\widehat{\boldsymbol{\varepsilon}}/\widehat{\sigma}$, em que $\widehat{\sigma}$ é um estimador de σ , em função de $\widehat{\mathbf{y}} = \mathbf{X}\widehat{\boldsymbol{\beta}} + \mathbf{Z}\widetilde{\mathbf{b}}$ para verificar a homoscedasticidade e o gráfico dos quantis (*Q-Q plot*) para verificar se há indícios de normalidade do resíduo condicional. Já Nobre e Singer (2007) sugerem padronizar o erro condicional

$$\widehat{\boldsymbol{\varepsilon}}_{\mathcal{L}}^* = \frac{\widehat{\boldsymbol{\varepsilon}}_{\mathcal{L}}}{\widehat{\sigma} \sqrt{\widehat{p}_{\mathcal{L}\mathcal{L}}}} \quad (31)$$

em que $\widehat{p}_{\mathcal{L}\mathcal{L}}$ é uma estimativa de $p_{\mathcal{L}\mathcal{L}}$ e $p_{\mathcal{L}\mathcal{L}}$ representa os $\mathcal{L}$ elementos da diagonal principal de $\boldsymbol{\Lambda} \mathbf{Q} \boldsymbol{\Lambda}$, com $\mathcal{L} = 1, 2, \dots, N_T$, de tal forma que este resíduo constitui uma generalização dos resíduos padronizados, pois quando $\boldsymbol{\Lambda} = \mathbf{I}_{N_T}$, sob as pressuposições de independência e homoscedasticidade, o resíduo condicional padrão apresentado na expressão (31) é reduzido para

$$\widehat{\boldsymbol{\varepsilon}}_{\mathcal{L}}^* = \frac{\widehat{\boldsymbol{\varepsilon}}_{\mathcal{L}}}{\widehat{\sigma} \sqrt{\widehat{q}_{\mathcal{L}\mathcal{L}}}}$$

com $q_{\mathcal{L}\mathcal{L}}$ denotando uma estimativa dos $\mathcal{L}$ elementos da diagonal principal de $\mathbf{Q}$. Segundo Nobre e Singer (2007), o resíduo condicional padrão, dado pela expressão (31), é útil para identificar observações com grande influência na estimativa de σ^2 e para verificar a homoscedasticidade do erro condicional (quando $\boldsymbol{\Lambda} = \mathbf{I}_{N_T}$), considerando o gráfico do resíduo condicional padrão em função dos valores ajustados ($\widehat{\mathbf{y}}$). Já para verificar a normalidade deve-se considerar o confundimento presente em $\boldsymbol{\varepsilon}$, que é discutido nos trabalhos de Hilden-

Minton (1995) e de Nobre e Singer (2007).

No modelo apresentado na expressão (1), o termo $\mathbf{Z}_i \tilde{\mathbf{b}}_i$ reflete a diferença entre a resposta predita para o i -ésimo indivíduo e a média populacional, o que, segundo Nobre e Singer (2007), pode ser usado verificar se o i -ésimo indivíduo é um *outlier*. Uma alternativa, proposta por Pinheiro e Bates (1995) foi a construção do gráfico dos elementos de $\tilde{\mathbf{b}}_i$ em função dos índices do indivíduo, que são úteis para encontrar possíveis *outliers* considerando o i -ésimo elemento do respectivo EBLUP. Para avaliar quais indivíduos são sensíveis à homogeneidade das matrizes de covariâncias dos efeitos aleatórios deve-se construir um gráfico de dispersão da matriz dos efeitos aleatórios preditos.

3 MATERIAL

Para obtenção dos dados de tonalidade média da casca do mamão cv. “Sunrise Solo” foi conduzido um experimento no laboratório do Departamento de Produção Vegetal da Escola Superior de Agricultura “Luiz de Queiroz”, no município de Piracicaba, SP, localizada a 22°41′ de latitude sul, 47°38′ de longitude oeste e 573 metros de altitude, durante o ano agrícola 2010/2011. Esse experimento foi financiado pela Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) por meio do programa de Iniciação Científica (processo 2010/16955-1) e pela Companhia de Entrepósitos e Armazéns Gerais de São Paulo (CEAGESP) unidade de Piracicaba-SP.

Para o experimento foram utilizados 20 frutos de mamoeiro cultivar “Sunrise Solo” colhidos em um mesmo dia e pertencentes a um mesmo lote, sendo que todos os frutos pertencentes a esse lote foram previamente classificados pela metodologia visual proposta por Balbino (1997), sendo classificados como pertencentes ao estágio 1 de maturação (até 10% de coloração amarela na casca). Os frutos, então, foram enumerados de 1 a 20 e estocados em câmara de refrigeração à 18 °C com 80% ± 5% de umidade relativa, como ilustra a Figura 3.

Figura 3 – Frutos de mamoeiro estocados numa câmara de refrigeração em dias distintos

Os equipamentos utilizados para a medição da cor da casca foram um *scanner* do tipo *flat-bed* (Scanner HP *Scanjet* G2410), utilizando uma resolução de 200 pixels por polegada (Figura 4(a)), e um colorímetro tristímulo Minolta CR-300 (KONICA MINOLTA).

Cada fruto foi avaliado diariamente e nos mesmos horários por meio de ambos os equipamentos, implicando, portanto, em um estudo longitudinal. Adicionalmente, após o término das avaliações os frutos que apresentavam desenvolvimento de doenças, como

Figura 4 – (a) *Scanner* de mesa com um fruto e (b) sistema para a medição da cor de um fruto composto por um *scanner* de mesa com um fruto, coberto por uma caixa preta, interligado a um computador para o armazenamento da imagem

antracnose (*Colletotrichum gloeosporioides* Penz.) ou podridão peduncular, eram descartados, conseqüentemente o número de avaliações por fruto foi diferente, implicando em um experimento desbalanceado. De forma geral, os períodos de avaliações variaram de 11 a 19 dias conforme o fruto.

Para o *scanner*, como descrito no item 2.1, foi necessário realizar a calibração antes do início do experimento. Assim, para a calibração foram utilizados 297 padrões de cor (Figura 5), da carta de cores de Munsell para tecidos vegetais, MUNSELL COLOR COMPANY (1952), sendo a metodologia para a calibração descrita por Oliveira e Zocchi (2010). Deve-se ressaltar que o *scanner* precisou apenas de uma única calibração durante todo o experimento.

Figura 5 – Exemplo de 17 padrões da carta de cores de Munsell para tecidos vegetais

Para cada digitalização da casca do fruto o *scanner* foi coberto com uma caixa

de papelão revestida com tecido de coloração preta, visando minimizar os efeitos de sombra e fornecer um fundo escuro à imagem, como ilustra a Figura 4(b). Essa imagem foi salva como um arquivo em formato Tiff (.tif), que apesar de possuir um tamanho de arquivo maior, preserva a imagem como ela foi originalmente digitalizada, além de ser recomendado para trabalhar com objetos escuros (RODRÍGUEZ et al., 2010).

Para cada avaliação foram digitalizadas duas faces de cada um dos 20 frutos, sendo que as faces foram delimitadas em cada fruto de forma que uma era oposta a outra e foram denominadas face A e face B. Ressalta-se que a digitalização das duas faces possibilitou a observação de praticamente toda a superfície do fruto. A Figura 6 mostra, como ilustração, as imagens digitalizadas da face A de um mesmo fruto nos dias 0, 5, 10, 15 e 17 após o início do experimento.

Figura 6 – Imagens digitalizadas da face A de um mesmo fruto aos 0, 5, 10, 15 e 17 dias após o início do experimento

Para obter a variável tonalidade média da cor (H) para o *scanner*, cada imagem teve seus pixels convertidos do sistema de cores *RGB* para o sistema de cores *CIELCH*, utilizando fórmulas específicas apresentadas por Asmare, Asirvadam e Iznita (2009).

Em seguida, foi calculada a tonalidade média, considerando as duas faces de cada fruto, a partir da metodologia apresentada por Oliveira e Zocchi (2013).

Para o colorímetro (Figura 7(a)) foram realizadas calibrações diárias, antes do início das avaliações, utilizando a metodologia descrita no manual do equipamento. Em seguida, para cada fruto foram observados quatro pontos equidistantes na região equatorial em locais pré-determinados como ilustra a Figura 7(b). Da mesma forma que para o *scanner*, a tonalidade média, para o colorímetro, foi obtida a partir da metodologia apresentada por Oliveira e Zocchi (2013).

É importante ressaltar que a proposta adotada para a avaliação da tonalidade

Figura 7 – (a) Colorímetro tristimulus Minolta CR-300 (KONICA MINOLTA) e (b) exemplo de dois pontos para avaliação da cor por meio do colorímetro.

média da casca do mamão utilizando um colorímetro é comumente utilizada no setor de pesquisa, podendo-se encontrar inúmeros exemplos dessa prática na literatura, como em Oliveira et al. (2002), Almeida et al. (2006) e Pinto et al. (2011). O mesmo não se pode dizer em relação à técnica de análise de imagens digitais, sendo, portanto, uma importante inovação tecnológica para o cálculo da tonalidade média da casca do mamão.

Uma parte dos dados de tonalidade média obtidos, diariamente, por meio desse experimento, para ambas as metodologias de análise da cor, são apresentados nas Tabelas 18 a 20 do Anexo A. Adicionalmente, no Apêndice F são apresentados os códigos utilizados no ambiente R para a análise dos dados de tonalidade média.

4 MÉTODOS

Como passo inicial da análise, sugere-se construir os gráficos de perfis com o objetivo de se detectarem as tendências individuais e conhecer um pouco a respeito da variabilidade entre os perfis individuais. Além disso, medidas de posição, como média e mediana, além de medidas de dispersão, como variância e desvio padrão, podem auxiliar na descrição do conjunto de dados.

No caso de haver uma variável regressora x , pode-se utilizar, como modelo, o polinomial de grau p dado por

$$y_{ij} = \beta_{0i} + \beta_{1i}x_{ij} + \beta_{2i}x_{ij}^2 + \cdots + \beta_{pi}x_{ij}^p, \quad (32)$$

que deve ser ajustado para o i -ésimo indivíduo, com $i = 1, 2, \dots, M$.

Os polinômios a serem utilizados, entretanto, não devem ter grau elevado, por exemplo, acima de grau três, de modo a evitar o aumento excessivo de complexidade do modelo.

Em seguida, para verificar se há necessidade de inclusão de efeitos aleatórios no modelo, Pinheiro e Bates (1995) sugerem construir, para cada parâmetro de efeito fixo $\beta_{0i}, \beta_{1i}, \dots, \beta_{pi}$, um intervalo de confiança baseado em um nível de significância α pré especificado para cada indivíduo.

Isso é feito da seguinte forma, para cada parâmetro de efeito fixo, são observados os M intervalos de confiança relativos aos M indivíduos. Se ao menos dois intervalos não se sobrepuserem, há um indício de que se deve adicionar um efeito aleatório ao respectivo parâmetro.

O procedimento sugerido será detalhado a seguir. Considere, para isso,

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i.$$

em que $\boldsymbol{\beta}_i = (\beta_{0i}, \beta_{1i}, \dots, \beta_{pi})^T$ é o vetor de parâmetros de dimensão $p + 1$. Então, para estimar os parâmetros pode-se utilizar o método de mínimos quadrados ordinários, cujas estimativas minimizam a soma de quadrados dos desvios das n_i observações do indivíduo i

em relação aos valores preditos pelo modelo, dada por

$$\begin{aligned}\boldsymbol{\varepsilon}_i^T \boldsymbol{\varepsilon}_i &= (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_i)^T (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_i) \\ &= (\mathbf{y}_i^T - \boldsymbol{\beta}_i^T \mathbf{X}_i^T) (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_i) \\ &= \mathbf{y}_i^T \mathbf{y}_i - 2\boldsymbol{\beta}_i^T \mathbf{X}_i^T \mathbf{y}_i + \boldsymbol{\beta}_i^T \mathbf{X}_i^T \mathbf{X}_i \boldsymbol{\beta}_i.\end{aligned}$$

Para isso, deve-se derivar $\boldsymbol{\varepsilon}_i^T \boldsymbol{\varepsilon}_i$ em relação a $\boldsymbol{\beta}_i$, ou seja,

$$\frac{\partial \boldsymbol{\varepsilon}_i^T \boldsymbol{\varepsilon}_i}{\partial \boldsymbol{\beta}_i} = -2\mathbf{X}_i^T \mathbf{y}_i + 2\mathbf{X}_i^T \mathbf{X}_i \boldsymbol{\beta}_i$$

e, em seguida, igualá-lo a um vetor de zeros, obtendo-se, assim, o sistema de equações normais dado por

$$\mathbf{X}_i^T \mathbf{X}_i \hat{\boldsymbol{\beta}}_i = \mathbf{X}_i^T \mathbf{y}_i$$

cujas soluções, uma vez que a matriz $\mathbf{X}_i^T \mathbf{X}_i$ tem posto completo, é única e é dada por

$$\hat{\boldsymbol{\beta}}_i = (\mathbf{X}_i^T \mathbf{X}_i)^{-1} \mathbf{X}_i^T \mathbf{y}_i. \quad (33)$$

Uma vez que a derivada segunda de $\boldsymbol{\varepsilon}_i^T \boldsymbol{\varepsilon}_i$ em relação a $\boldsymbol{\beta}_i$ é $2\mathbf{X}_i^T \mathbf{X}_i$ e que $\mathbf{X}_i^T \mathbf{X}_i$ é positiva definida, isto é, tem seus autovalores todos positivos não nulos, então o vetor $\hat{\boldsymbol{\beta}}_i$, que anula a derivada primeira, minimiza a soma de quadrados dos desvios $\boldsymbol{\varepsilon}_i^T \boldsymbol{\varepsilon}_i$.

Pode-se, ainda, demonstrar facilmente que $E(\hat{\boldsymbol{\beta}}_i) = \boldsymbol{\beta}_i$ e que a $Cov(\hat{\boldsymbol{\beta}}_i) = \sigma^2 (\mathbf{X}_i^T \mathbf{X}_i)^{-1}$. Detalhes sobre a demonstração são apresentados no Apêndice D. Além disso, $\boldsymbol{\varepsilon}_i \stackrel{iid}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{I}\sigma^2)$, tem-se que, sob a hipótese $H_0 : \beta_{Pi} = 0$,

$$t_p = \frac{\hat{\beta}_{Pi} - \beta_{Pi}}{s_i \sqrt{g_{PPi}}}$$

tem distribuição t central, com $n_i - (p+1) - 1$ graus de liberdade, em que $P = 0, 1, 2, \dots, p$, g_{PPi} é o P -ésimo elemento da diagonal da matriz $(\mathbf{X}_i^T \mathbf{X}_i)^{-1}$ e

$$s_i^2 = \frac{\mathbf{y}_i^T \mathbf{y}_i - \hat{\boldsymbol{\beta}}_i^T \mathbf{X}_i^T \mathbf{y}_i}{n_i - p - 2}.$$

é um estimador não viesado de σ_i^2 . Logo, demonstra-se que:

$$P \left[-t_{(\frac{\alpha}{2}, n_i - p - 2)} < \frac{\hat{\beta}_{Pi} - \beta_{Pi}}{s_i \sqrt{g_{PPi}}} < t_{(\frac{\alpha}{2}, n_i - p - 2)} \right] = 1 - \alpha$$

ou seja, o intervalo de confiança de $100(1 - \alpha)\%$ para β_{Pi} é dado por

$$\hat{\beta}_{Pi} \pm t_{(\frac{\alpha}{2}, n_i - p - 2)} s_i \sqrt{g_{PPi}}.$$

em que $t_{(\frac{\alpha}{2}, n_i - p - 2)}$ é o quantil de ordem $(1 - \frac{\alpha}{2})$ da distribuição t com $(n_i - p - 2)$ graus de liberdade.

Na prática, se ao menos dois intervalos de confiança para β_{Pi} não se sobrepuserem, há um indício de que se deve adicionar um efeito aleatório ao respectivo parâmetro.

Computacionalmente, as estimativas dos parâmetros de efeitos fixos do modelo dado na expressão (32) e respectivos intervalos de confiança para todos os indivíduos, podem ser obtidos no ambiente computacional **R** versão 2.15.2 (BATES et al., 2013) por meio da função **lmList** da biblioteca de funções **nlme** versão 3.1-108 (PINHEIRO et al., 2013).

Uma vez decidido de modo empírico quais parâmetros são afetados pelos diferentes indivíduos, ou seja, a quais parâmetros devem-se acrescentar efeitos aleatórios, chega-se, para $p = 2$, a um modelo dentre os modelos candidatos apresentados na Tabela 1. Estes, por sua vez, podem ser ajustados utilizando-se a metodologia descrita no item 2.2, implementada no ambiente **R** por Pinheiro e Bates (1995), por meio da função **lme** da biblioteca de funções **nlme** versão 3.1-108 (PINHEIRO et al., 2013).

Como alternativa para a seleção do modelo, pode-se partir de um modelo completo, no caso para $p = 2$, um modelo polinomial quadrático com efeitos aleatórios sobre β_0, β_1 e β_2 , retirando-se, gradativamente, parâmetros que não são relevantes na descrição do processo, como ilustra a Figura 8.

Assim, para a seleção dos modelos, deve-se seguir o seguinte procedimento:

- i) Assumir, como ponto de partida, as estruturas de variâncias e covariâncias $\Sigma_i = \sigma^2 \mathbf{I}$ e não estruturada (NE) para a matriz de variâncias e covariâncias $\mathbf{D}_*$ dos modelos M1 a M11;
- ii) Quando o teste da razão de verossimilhanças for para verificar a exclusão de um efeito

Tabela 1 – Modelos lineares de efeitos mistos construídos a partir de (1) considerando-se $p = 2$

Efeitos Fixos	Efeitos aleatórios	Modelo	Expressão
β_0	b_{0i}	M1	$\beta_0 + b_{0i}$
	b_{0i}	M2	$\beta_0 + b_{0i} + \beta_1 x_i$
β_0 e β_1	b_{1i}	M3	$\beta_0 + (\beta_1 + b_{1i}) x_i$
	b_{0i} e b_{1i}	M4	$\beta_0 + b_{0i} + (\beta_1 + b_{1i}) x_i$
	b_{0i}	M5	$\beta_0 + b_{0i} + \beta_1 x_i + \beta_2 x_i^2$
	b_{1i}	M6	$\beta_0 + (\beta_1 + b_{1i}) x_i + \beta_2 x_i^2$
	b_{2i}	M7	$\beta_0 + \beta_1 x_i + (\beta_2 + b_{2i}) x_i^2$
β_0, β_1 e β_2	b_{0i} e b_{1i}	M8	$\beta_0 + b_{0i} + (\beta_1 + b_{1i}) x_i + \beta_2 x_i^2$
	b_{0i} e b_{2i}	M9	$\beta_0 + b_{0i} + \beta_1 x_i + (\beta_2 + b_{2i}) x_i^2$
	b_{1i} e b_{2i}	M10	$\beta_0 + (\beta_1 + b_{1i}) x_i + (\beta_2 + b_{2i}) x_i^2$
	b_{0i}, b_{1i} e b_{2i}	M11	$\beta_0 + b_{0i} + (\beta_1 + b_{1i}) x_i + (\beta_2 + b_{2i}) x_i^2$

aleatório do modelo, ou seja, testar se o componente de variância associado a esse efeito aleatório é igual a zero, deve-se utilizar o método da máxima verossimilhança restrita para estimar os efeitos fixos e os componentes de variância dos modelos a serem comparados;

- iii) Quando o teste da razão de verossimilhanças for para verificar a exclusão de um efeito fixo do modelo, ou seja, testar se esse efeito fixo é igual a zero, deve-se utilizar o método da máxima verossimilhança para estimar os efeitos fixos e os componentes de variâncias dos modelos a serem comparados;
- iv) Comparar os modelos M8 a M10 com o modelo de referência M11 por meio do teste da razão de verossimilhanças, descrito no item 2.6.2, sendo que:
 - 1) Se o modelo escolhido for o M11, encerra-se o processo de seleção;
 - 2) Se um dentre os modelos M8 a M10 for escolhido, o processo de seleção continua a

Figura 8 – Fluxograma do procedimento de seleção a partir do modelo completo M11. Teste de razão de verossimilhanças para verificar a necessidade de exclusão de um efeito aleatório (círculos em azul) e para verificar a necessidade de exclusão de um efeito fixo (círculos em laranja); retângulos em vermelho, verde ou amarelo indicam os modelos que podem ser comparados pelos critérios de informação AIC ou BIC

partir desse modelo;

- 3) Se pelo menos 2 modelos forem escolhidos dentre os modelos M8 a M10 deve-se comparar os modelos por meio do critério de informação AIC ou BIC, descritos no item 2.6.1, selecionando o modelo que apresentar menor valor de AIC ou BIC.

Segundo Demidenko (2004), o BIC deve ser utilizado para conjunto grande de dados, caso contrário, o modelo pode ser escolhido a partir do AIC;

v) A partir do modelo escolhido no item iv, com exceção do modelo M11, deve-se prosseguir a seleção a partir de um dos três casos:

1) Caso M8: comparar os modelos M4, M5 e M6 com o modelo M8 por meio do teste da razão de verossimilhanças, sendo que:

- a) Se o modelo escolhido for o M8 encerra-se o processo de seleção;
- b) Se um dentre os modelos M4 a M6 for escolhido, o processo de seleção continua a partir desse modelo;
- c) Se pelo menos 2 modelos foram escolhidos dentre os modelos M4 a M6 deve-se comparar os modelos por meio do critério de informação AIC ou BIC selecionando o modelo que apresentar menor valor de AIC ou BIC. É importante ressaltar que para comparar o modelo M4 com os modelos M5 ou M6 por meio dos critérios AIC ou BIC deve-se, antes, estimar os efeitos fixos e os componentes de variância do modelo M4 pelo método da máxima verossimilhança restrita;

2) Caso M9: comparar os modelos M5 e M7 com o modelo M9 por meio do teste da razão de verossimilhanças, sendo que:

- a) Se o modelo escolhido for o M9 encerra-se o processo de seleção;
- b) Se um dentre os modelos M5 e M7 for escolhido, o processo de seleção continua a partir desse modelo;
- c) Se ambos os modelos foram escolhidos deve-se compará-los por meio do critério de informação AIC ou BIC selecionando o modelo que apresentar menor valor de AIC ou BIC;
- d) Se o modelo escolhido for o M7 encerra-se o processo de seleção;

3) Caso M10: comparar os modelos M6 e M7 com o modelo M10 por meio do teste da razão de verossimilhanças, sendo que:

- a) Se o modelo escolhido for o M10 encerra-se o processo de seleção;
- b) Se um dentre os modelos M6 e M7 for escolhido, o processo de seleção continua a partir desse modelo;

- c) Se ambos os modelos foram escolhidos deve-se compará-los por meio do critério de informação AIC ou BIC selecionando o modelo que apresentar menor valor de AIC ou BIC;
 - d) Se o modelo escolhido for o M7 encerra-se o processo de seleção;
- vi) A partir do modelo escolhido no item v, com exceção dos modelo M11 a M7, deve-se prosseguir a seleção a partir de um dos três casos:
- 1) Caso M4: comparar os modelos M3 e M2 com o modelo M4 por meio do teste da razão de verossimilhanças, sendo que:
 - a) Se o modelo escolhido for o M3 ou o M4 encerra-se o processo de seleção;
 - b) Se o modelo escolhido for o M2, o processo de seleção continua a partir desse modelo;
 - 2) Caso M5: comparar o modelo M2 com o modelo M5 por meio do teste da razão de verossimilhanças, sendo que:
 - a) Se o modelo escolhido for o M5 encerra-se o processo de seleção;
 - b) Se o modelo escolhido for o M2, o processo de seleção continua a partir desse modelo;
 - 3) Caso M6: comparar o modelo M3 com o modelo M6 por meio do teste da razão de verossimilhanças. Independentemente do modelo selecionado, encerra-se o processo de seleção;
- vii) Comparar, caso necessário, os modelos M2 e M1 por meio do teste da razão de verossimilhanças. Independentemente do modelo selecionado, encerra-se o processo de seleção.

Em seguida, parte-se para a seleção da estrutura da matriz de variâncias e covariâncias dos efeitos aleatórios $\mathbf{D}_*$ a partir do teste da razão de verossimilhanças para modelos aninhados. Nesse ponto, deve-se ressaltar que os parâmetros de efeitos fixos e os componentes de variância dos modelos devem ser estimados pelo método da máxima verossimilhança restrita. Computacionalmente, podem ser construídas algumas das estruturas descritas no item 2.3 para a matriz de variâncias e covariâncias $\mathbf{D}_*$ utilizando-se o argumento *pdMat* da função *lme*, do ambiente *R*.

Ressalta-se, ainda, que caso a estrutura de variâncias e covariâncias selecionada para a matriz $\mathbf{D}_*$ não for a estrutura assumida previamente (NE), deve-se seguir os seguintes passos:

- i) Assumir a nova estrutura selecionada para a matriz de variâncias e covariâncias dos efeitos aleatórios ($\mathbf{D}_*$) e manter a estrutura da matriz de variâncias e covariâncias dos erros ($\Sigma_i = \sigma^2 \mathbf{I}$) nos modelos apresentados na Tabela 1;
- ii) Realizar uma nova seleção desses modelos a partir do item ii da seleção de modelos, descrita anteriormente nessa seção.

Selecionada a matriz $\mathbf{D}_*$, pode-se calcular a autocorrelação empírica (ACF) usando os pares de resíduos dentro de um mesmo indivíduo para dados observados ao longo do tempo em intervalos de tempo equidistantes. Para tanto, pode-se utilizar a função de autocorrelação (ACF) do pacote *nlme* versão 3.1-108 (PINHEIRO et al., 2013).

Por outro lado, quando os dados não são observados em intervalos iguais de tempo, pode-se utilizar o gráfico das estimativas do semivariograma amostral correspondente aos resíduos condicionais padronizados, com adição de uma curva de suavização LOESS para facilitar a visualização dos padrões no semivariograma. Computacionalmente, as estimativas do semivariograma amostral podem ser obtidas por meio da função *Vario-gram* da biblioteca de funções *nlme* versão 3.1-108 (PINHEIRO et al., 2013).

Após verificar a correlação temporal, parte-se para a seleção da estrutura de variâncias e covariâncias da matriz Σ_i . Nesse ponto, podem-se assumir estruturas do tipo CV, CV_H e AR(1), descritas no item 2.3. Adicionalmente, para modelar a heterocedasticidade de variâncias para a estrutura de covariâncias AR(1) pode-se utilizar uma função de variância exponencial para a matriz Σ , proposta por Pinheiro e Bates (1995), dada por:

$$Var(\varepsilon_{ij}|\mathbf{b}_i) = \sigma^2 \exp(\delta v_{ij}), \quad i = 1, 2, \dots, M, \quad j = 1, 2, \dots, n_i \quad (34)$$

em que v_{ij} é uma covariável de variância e δ é um parâmetro de variância, sendo que δ é irrestrito. Já para construir a estrutura CV_H pode-se utilizar uma função de variância dada por:

$$Var(\varepsilon_{ij}) = \sigma^2 \delta_{k_{ij}}^2, \quad (35)$$

em que δ é um parâmetro de variância e k representa cada nível de estratificação (no caso em questão, cada estrato é dado pelo j -ésimo dia, $j = 1, 2, \dots, n_i$, com $k = 1, 2, \dots, K$). Como o modelo apresentado na expressão (35) utiliza $K + 1$ parâmetros para representar K variâncias deve-se, segundo Pinheiro e Bates (1995), colocar uma restrição no vetor de parâmetros de variância $\boldsymbol{\delta}$. A restrição utilizada foi $\delta_{1_{ij}} = 1$, de modo que $\delta_{s_{ij}}$, $s = 2, \dots, K$, representa a razão entre o desvio padrão do s -ésimo estrato e o primeiro estrato. Por definição, Pinheiro e Bates (1995), utilizaram $\delta_s > 0$. Para maiores informações consultar Pinheiro e Bates (1995, p. 209).

Dessa forma, a seleção da estrutura de variâncias e covariâncias para a matriz $\boldsymbol{\Sigma}_i$ pode ser realizada por meio do teste da razão de verossimilhança para modelos aninhados.

Novamente, ressalta-se que caso a estrutura selecionada para a matriz de variâncias e covariâncias $\boldsymbol{\Sigma}_i$ não for a estrutura previamente assumida ($\boldsymbol{\Sigma}_i = \sigma^2 \mathbf{I}$), deve-se seguir os seguintes passos:

- i) Assumir a nova estrutura selecionada para a matriz de variâncias e covariâncias dos erros ($\boldsymbol{\Sigma}_i$) e manter a estrutura selecionada para a matriz de variâncias e covariâncias dos efeitos aleatórios ($\mathbf{D}_*$) nos modelos apresentados na Tabela 1;
- ii) Realizar uma nova seleção desses modelos a partir do item ii da seleção de modelos, descrita anteriormente nessa seção.

Selecionada as estruturas mais adequadas para $\mathbf{D}_*$ e $\boldsymbol{\Sigma}_i$, parte-se para a análise de diagnóstico, que, de forma geral, pode ser realizada observando os gráficos de dispersão e de normalidade para o resíduo condicional e para os efeitos aleatórios.

Assim, para verificar a homoscedasticidade pode-se utilizar o gráfico dos elementos de $\widehat{\boldsymbol{\varepsilon}}/\widehat{\sigma}$, em que $\widehat{\sigma}$ é um estimador de σ , em função de $\widehat{\mathbf{y}} = \mathbf{X}\widehat{\boldsymbol{\beta}} + \mathbf{Z}\widetilde{\mathbf{b}}$.

Já para verificar a normalidade do resíduo condicional pode-se utilizar o gráfico dos resíduos condicionais padronizados em função dos quantis da distribuição normal padrão com adição de um envelope simulado¹. Adicionalmente, pode-se realizar o teste de normalidade de Shapiro-Wilk para testar a hipótese de normalidade dos resíduos condicionais padronizados, sendo que o teste pode ser realizado utilizando a função *shapiro.test* da biblioteca de funções *stats* (R Core Team and contributors worldwide, 2010).

¹MORAL, R. A.; HINDE, J.; DEMÉTRIO, C. G. B. Modelagem estatística e ecológica de relações tróficas em pragas e inimigos naturais. 2014. Dissertação (Mestrado em Estatística e Experimentação Agrônômica) – Escola Superior de Agricultura “Luiz de Queiroz”, Piracicaba, 2014, 173p.

Em relação a normalidade dos efeitos aleatórios, podem-se utilizar os gráficos normais de probabilidade com envelope simulado para cada efeito aleatório. Computacionalmente, pode-se utilizar a função **qqPlot** da biblioteca de funções **car** versão 2.0-19 (FOX; WEISBERG, 2011).

Em seguida, podem-se calcular os intervalos de confiança aproximados para os componentes de variância e covariância e para os efeitos fixos utilizando distribuições aproximadas para as estimativas de máxima verossimilhança e de máxima verossimilhança restrita e o teste t condicional.

Segundo Pinheiro e Bates (1995) as distribuições aproximadas das estimativas de máxima verossimilhança $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_Q, \sigma^2)$ em um LMEM com Q níveis de aninhamento são

$$\hat{\boldsymbol{\beta}} \sim \mathcal{N}(\boldsymbol{\beta}, \sigma^2 [\mathbf{R}^{-1} \mathbf{R}^{-T}])$$

e

$$\begin{bmatrix} \hat{\boldsymbol{\theta}}_1 \\ \hat{\boldsymbol{\theta}}_2 \\ \vdots \\ \hat{\boldsymbol{\theta}}_Q \\ \log \hat{\sigma} \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \boldsymbol{\theta}_1 \\ \boldsymbol{\theta}_2 \\ \vdots \\ \boldsymbol{\theta}_Q \\ \log \sigma \end{bmatrix}, \mathcal{I}^{-1}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_Q, \sigma) \right) \quad (36)$$

em que

$$\mathcal{I}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_Q, \sigma) = - \begin{bmatrix} \frac{\partial^2 l}{\partial \boldsymbol{\theta}_1 \partial \boldsymbol{\theta}_1^T} & \frac{\partial^2 l}{\partial \boldsymbol{\theta}_2 \partial \boldsymbol{\theta}_1^T} & \frac{\partial^2 l}{\partial \boldsymbol{\theta}_3 \partial \boldsymbol{\theta}_1^T} & \cdots & \frac{\partial^2 l}{\partial \log \sigma \partial \boldsymbol{\theta}_1^T} \\ \frac{\partial^2 l}{\partial \boldsymbol{\theta}_1 \partial \boldsymbol{\theta}_2^T} & \frac{\partial^2 l}{\partial \boldsymbol{\theta}_2 \partial \boldsymbol{\theta}_2^T} & \frac{\partial^2 l}{\partial \boldsymbol{\theta}_3 \partial \boldsymbol{\theta}_2^T} & \cdots & \frac{\partial^2 l}{\partial \log \sigma \partial \boldsymbol{\theta}_2^T} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 l}{\partial \boldsymbol{\theta}_1 \partial \log \sigma} & \frac{\partial^2 l}{\partial \boldsymbol{\theta}_2 \partial \log \sigma} & \frac{\partial^2 l}{\partial \boldsymbol{\theta}_3 \partial \log \sigma} & \cdots & \frac{\partial^2 l}{\partial^2 \log \sigma} \end{bmatrix}$$

é a matriz de informação empírica e

$$\mathbf{R} = \sum_{i=1}^M \frac{\mathbf{X}_i^T \sigma^2 \mathbf{W}_i(\gamma) \mathbf{X}_i}{2}.$$

Seja gl_p o número de graus de liberdade do teste t condicional para o p -ésimo

efeito fixo, segundo Pinheiro e Bates (1995), um intervalo de confiança aproximado com nível de $1 - \alpha$ de significância para β_p é

$$\hat{\beta}_p \pm t_{gl_p, (1-\frac{\alpha}{2})} \hat{\sigma}_R \sqrt{[\mathbf{R}^{-1}\mathbf{R}^{-T}]_{pp}} \quad (37)$$

em que

$$\hat{\sigma}_R^2(\boldsymbol{\theta}) = \frac{N_T \hat{\sigma}^2(\boldsymbol{\theta})}{N_T - p - 1},$$

é o estimador de máxima verossimilhança restrita de σ^2 e $\hat{\sigma}^2$ é o estimador de máxima verossimilhança de σ^2 , sendo que $\hat{\sigma}^2$ é a soma de quadrado dos resíduos dividida por N_T e $[\mathbf{R}^{-1}\mathbf{R}^{-T}]_{pp}$ representa os elementos da diagonal principal da matriz $\mathbf{R}^{-1}\mathbf{R}^{-T}$.

Já o intervalo de confiança para o desvio padrão σ dentro de cada indivíduo é obtido, segundo Pinheiro e Bates (1995), a partir da expressão (36). Seja $[\mathcal{I}]_{\sigma\sigma}$ os elementos da diagonal principal da inversa da matriz de informação empírica, apresentada na expressão (36), então, um intervalo de confiança aproximado de nível $1 - \alpha$ de significância para σ é dado por

$$\left[\hat{\sigma} \exp \left(-z_{(1-\frac{\alpha}{2})} \sqrt{[\mathcal{I}^{-1}]_{\sigma\sigma}} \right), \hat{\sigma} \exp \left(z_{(1-\frac{\alpha}{2})} \sqrt{[\mathcal{I}^{-1}]_{\sigma\sigma}} \right) \right],$$

em que $z_{(1-\frac{\alpha}{2})}$ representa o quantil $1 - \frac{\alpha}{2}$ da distribuição normal padrão.

Seja $\boldsymbol{\kappa}$ o vetor de parâmetros da matriz de variâncias e covariâncias $\mathbf{D}_*$. Agora, considere que κ_j corresponda ao logaritmo do desvio padrão em $\mathbf{D}_*$ e fazendo $[\mathcal{I}]_{jj}$ representando os elementos da diagonal associados à inversa da matriz de informação empírica, um intervalo de confiança aproximado de nível $1 - \alpha$ para o correspondente desvio padrão, segundo Pinheiro e Bates (1995), é dado por

$$\left[\exp \left(\hat{\kappa}_j - z_{(1-\frac{\alpha}{2})} \sqrt{[\mathcal{I}^{-1}]_{jj}} \right), \exp \left(\hat{\kappa}_j + z_{(1-\frac{\alpha}{2})} \sqrt{[\mathcal{I}^{-1}]_{jj}} \right) \right],$$

Computacionalmente, os intervalos de confiança para os efeitos fixos $\boldsymbol{\beta}$ e para os componentes de variância $\boldsymbol{\gamma}$ são obtidos a partir da função **intervals** da biblioteca de funções **nlme** versão 3.1-108 (PINHEIRO et al., 2013).

5 RESULTADOS E DISCUSSÃO

Inicialmente, no item 5.1 são apresentados, para o *scanner* e para o colorímetro, os resultados e discussões da análise descritiva para dados descritos na seção 4 e apresentados nas Tabelas 18 a 22 do Anexo A, realizada por meio de gráficos de perfis individuais e por medidas de posição. Já no item 5.2, são apresentados, para ambas as metodologias de análise da cor, os resultados e discussões a respeito do procedimento de seleção dos modelos, dos gráficos e teste de normalidade para os resíduos, bem como dos gráficos de normalidade dos efeitos aleatórios. São apresentados, ainda, os gráficos dos ajustes individuais e os intervalos de 95 % de confiança para os parâmetros de efeito fixo e para os componentes de variância das matrizes $\mathbf{D}_*$ e Σ_i .

5.1 Análise descritiva

A Figura 9 apresenta o gráfico da tonalidade média em função do tempo, em dias após o início das avaliações, para ambas as metodologias de análise da cor. De forma geral, independentemente da metodologia, a tonalidade média dos frutos ao longo do experimento mudou de verde (tonalidade entre 110 e 120 graus) para amarelo alaranjado (tonalidade próxima a 75 graus). Por outro lado, apesar de os frutos terem sido escolhidos de modo a terem colorações semelhantes no início do experimento, ao final apresentaram colorações bem mais variáveis, levantando a hipótese de que ocorreram diferentes formas de maturação.

Verifica-se, também, que a tonalidade média inicial de frutos avaliados pelo colorímetro foi sempre maior do que quando avaliados pelo *scanner*. Considerando a Tabela 2 e a Figura 9(a), verificou-se que os frutos 14, 15 e 18 estavam em um estágio de maturação inicial mais avançado (mais amarelados), indicando que o lote em questão, apesar de ter sido colhido em um mesmo dia e classificado como homogêneo por meio de classificação visual da cor, apresentou uma variabilidade inicial que deve ser incorporada ao modelo. Por outro lado, ao observar a Tabela 2 e a Figura 9(b), notou-se que os frutos 12 e 18 se destacaram por apresentarem estágio de maturação, aparentemente, mais avançado (mais amarelados), divergindo dos resultados apresentados pelo *scanner*.

Figura 9 – Tonalidade média de frutos medidas ao longo do tempo por meio de um *scanner* (a) e de um colorímetro (b), bem como a escala de cores baseada na tonalidade, fixando $L = 50$ e $C = 50$.

Tabela 2 – Tonalidade média inicial, em graus, de 20 frutos avaliados por meio de um colorímetro e de um *scanner*; números em negrito foram utilizados para destacar os frutos com estágio de maturação mais avançado (mais amarelados)

Fruto	Colorímetro	<i>Scanner</i>	Fruto	Colorímetro	<i>Scanner</i>
1	116,54	111,55	11	115,56	112,06
2	115,41	113,83	12	112,96	111,16
3	116,98	113,75	13	115,48	112,31
4	115,88	112,77	14	115,05	109,77
5	115,71	112,09	15	114,38	109,26
6	116,14	111,55	16	114,32	111,48
7	114,44	112,16	17	115,75	111,65
8	116,10	112,40	18	113,15	107,88
9	115,18	111,51	19	116,59	111,24
10	116,56	113,43	20	116,41	111,42

Observou-se, assim, que as duas metodologias não medem a tonalidade média inicial de um fruto da mesma forma, ou seja, não são equivalentes. Isto pois, na forma usual, o colorímetro é utilizado em quatro pontos da região equatorial do fruto, não detectando as mudanças iniciais que começam na região estilar do mesmo. Por outro lado, na metodologia relacionada ao *scanner*, os pontos são observados em toda região da casca do fruto,

implicando, aparentemente, em uma tonalidade média menor em relação a obtida quando utiliza-se um colorímetro.

Adicionalmente, a Figura 9(a) revela a possível existência de dois grupos de mamões que apresentam maturidades fisiológicas distintas, sendo que um grupo apresenta maturação mais lenta e menos indivíduos em relação ao outro grupo. Segundo Fonseca et al. (2007), os frutos colhidos no estágio fisiológico 1 (com até 15% da superfície da casca amarela), apresentam riscos grandes de não amadurecerem. Assim, o grupo de amadurecimento mais lento, detectado pela metodologia do *scanner*, pode, aparentemente, estar relacionado ao ponto de colheita precoce dos frutos.

Por outro lado, quando se avaliam os frutos medidos pelo colorímetro, (Figura 9(b)), esses grupos não são tão evidentes, podendo indicar que somente quatro pontos, bem como a posição dos mesmos no fruto podem não ser suficientes para avaliar sua tonalidade média do fruto como um todo.

Na Figura 9(b) observa-se que em dois dias ocorreram possíveis erros sistemáticos nas medições, uma vez que essas medidas estão muito acima de medidas anteriores e posteriores. Este erro, está associado a uma provável calibração errada do colorímetro e, assim, os dados relativos aos dias 5 e 8 foram retirados da análise. A Figura 10 apresenta o gráfico de perfis para a metodologia de análise da cor por meio de um colorímetro após a remoção dos dados correspondentes aos dias 5 e 8.

Figura 10 – Tonalidade média de frutos medidos ao longo do tempo por um colorímetro, desconsiderando os dados relativos aos dias 5 e 8, bem como a escala de cores baseada na tonalidade, fixando $L = 50$ e $C = 50$.

A Figura 11 apresenta os gráficos de perfis das tonalidades médias das cascas de 20 frutos, medidas por meio de um *scanner* e de um colorímetro. Analisando-a, nota-se que, de forma geral, as tonalidades médias dos frutos medidas por meio de um *scanner* foram, aparentemente, inferiores às medidas por meio de um colorímetro durante o período de tempo de observação dos frutos. Isso é provavelmente relativo a região amostrada e ao número de pontos observados no fruto serem diferentes em cada método de análise da cor. Na metodologia relacionada ao colorímetro, geralmente observam-se 4 pontos na região equatorial do fruto, enquanto que na metodologia relacionada ao *scanner* observam-se cerca de 250.000 pontos em toda superfície do mesmo, considerando os lados opostos denominados, aqui, como face A e face B. Isso mostra que o estudo da cor baseada em imagens digitalizadas possibilita uma melhor compreensão da evolução da cor no fruto e, conseqüentemente, um estudo de maturação mais preciso.

Além disso, observando o gráfico de perfis individuais (Figura 11), as tonalidades médias dos frutos 2, 5, 7, 8, 9, 10, 12, 14, 15, 16, 18, 19 e 20 medidas por meio de um *scanner* e as tonalidades médias dos frutos 2, 4, 5, 8, 9, 12, 16, 18 e 19 medidas por meio de um colorímetro, ao longo do tempo, apresentam, aparentemente, comportamento linear. Já as tonalidades médias dos demais frutos medidas por meio de um *scanner* e de um colorímetro, ao longo do tempo, apresentam, aparentemente, comportamento quadrático.

Além disso, para explicar a variabilidade ao longo do tempo entre os frutos e num mesmo fruto, serão utilizados, na próxima subseção, os LMEM, uma vez que permitem incorporar a variabilidade entre os indivíduos e dentro do mesmo indivíduo. Isto devido à flexibilidade das matrizes de variâncias e covariâncias, sendo indicados por Pinheiro e Bates (1995) e por Verbeke e Molenberghs (2009) para a análise de dados longitudinais.

Figura 11 – Gráficos de perfis das tonalidades médias das cascas de 20 frutos, medidas por meio de um *scanner* e de um colorímetro por um período de 16 dias, desconsiderando os dados relativos aos dias 5 e 8 para o colorímetro

5.2 Ajuste dos modelos

Para verificar a necessidade da inclusão de efeitos aleatórios no modelo, ajustou-se, inicialmente, o modelo polinomial quadrático apresentado na expressão (32) para cada fruto, obtendo-se para cada um, os intervalos de 95% de confiança dos parâmetros, apresentados graficamente nas Figuras 12 e 13. Esses gráficos também são amplamente utilizados por Pinheiro e Bates (1995), de modo a decidir a qual parâmetro de efeito fixo há necessidade de incluir um parâmetro de efeito aleatório, ou seja, auxiliam o pesquisador a saber qual será o LMEM provavelmente mais adequado.

De forma geral, para ambas as metodologias, nota-se pelas Figuras 12 e 13, que, aparentemente, há necessidade de incluir efeito aleatório em todos os parâmetros de efeito fixo do modelo polinomial quadrático, uma vez que, existem pelo menos dois intervalos de confiança para β_{0i} , para β_{1i} e para β_{2i} que não estão sobrepostos, respectivamente. Note que, a necessidade do efeito quadrático está relacionada, aparentemente, aos frutos 1, 3, 6

e 11 que apresentam maturação fisiológica aparentemente mais lenta. Além disso, os frutos que não necessitam do efeito quadrático no LMEM devem apresentar $\beta_2 + b_{2i} \approx 0$.

Figura 12 – Intervalos de 95% de confiança dos parâmetros de efeitos fixos do modelo polinomial quadrático (32) ajustados aos dados de tonalidade média da casca dos 20 frutos obtidos pelo *scanner*

Assim, para selecionar o modelo que melhor explica a variabilidade da tonalidade média ao longo do tempo, foram ajustados os 11 modelos descritos na seção 4. Para a seleção dos modelos utilizou-se o teste da razão de verossimilhanças (TRV) e os critérios AIC ou BIC. Os resultados são apresentados nas Tabela 3 e 4 para o *scanner* e para o colorímetro, respectivamente, e mostram que, para ambas as metodologias, o ajuste do modelo M11 difere dos modelos M10, M9 e M8, a 5% de significância, sendo, dentre os modelos apresentados, o modelo mais adequado para esse conjunto de dados.

Em seguida, a partir do modelo M11, foram testadas diferentes estruturas de variâncias e covariâncias para a matriz de efeitos aleatórios $\mathbf{D}_*$, cujo resultado é apresentado na Tabela 5 para o *scanner* e na Tabela 6 para o colorímetro. A partir das mesmas, verifica-se que a matriz de variâncias e covariâncias dos efeitos aleatórios dos modelos ajustados não necessita de modificação, ou seja, a estrutura NE, assumida previamente, deve permanecer nos respectivos modelos. Isso indica que os efeitos aleatórios não são independentes, ou seja, que a covariância entre dois efeitos aleatórios distintos é diferente de zero e que os efeitos aleatórios possuem variâncias distintas.

Figura 13 – Intervalos de 95% de confiança dos parâmetros de efeitos fixos do modelo polinomial quadrático (32) ajustados aos dados de tonalidade média da casca dos 20 frutos obtidos pelo colorímetro

Tabela 3 – Estatística χ^2 para o teste da razão de verossimilhanças (TRV) entre o modelo M11 e os modelos aninhados subsequentes M10, M9 e M8; critérios de informação AIC e BIC para os modelos M1 a M11 ajustados aos dados de tonalidade média obtidos por meio do *scanner* e respectivos números de parâmetros do modelo (np), números de graus de liberdade (gl) e valores-p do teste

Modelo	AIC	BIC	np	χ^2	gl	valor-p
M11	867,9	905,0	10			
M10	920,1	946,1	7	58,2	3	<0,001
M9	1026,6	1052,6	7	164,7	3	<0,001
M8	973,6	999,6	7	111,7	3	<0,001
M7	1297,8	1316,4	5			
M6	1123,1	1141,7	5			
M5	1181,8	1200,4	5			
M4	1024,3	1046,9	6			
M3	1145,3	1160,2	4			
M2	1194,3	1209,2	4			
M1	2180,4	2191,6	3			

Tabela 4 – Estatística χ^2 para o teste da razão de verossimilhanças (TRV) entre o modelo M11 e os modelos aninhados subsequentes M10, M9 e M8; critérios de informação AIC e BIC para os modelos M1 a M11 ajustados aos dados de tonalidade média obtidos por meio do *scanner* e respectivos números de parâmetros do modelo (np), números de graus de liberdade (gl) e valores-p do teste

Modelo	AIC	BIC	np	χ^2	gl	valor-p
M11	805,6	841,33	10			
M10	848,0	873,0	7	48,4	3	0,002
M9	939,2	964,3	7	139,7	3	<0,001
M8	870,5	895,5	7	70,9	3	<0,001
M7	1159,3	1177,2	5			
M6	983,2	1001,1	5			
M5	1168,4	1186,3	5			
M4	1068,0	1089,5	6			
M3	1112,3	1126,6	4			
M2	1222,4	1236,7	4			
M1	2001,3	2012,0	3			

Na prática, a estrutura NE indica que a tonalidade média da casca, medida ao longo do tempo, apresenta comportamentos variáveis de acordo com o indivíduo, possibilitando, por exemplo, o ajuste de curvas concorrentes. Dessa forma, a variabilidade natural existente entre os frutos, apresentada na Figura 9 da subseção 5.1, pode ser captada com a utilização dessa estrutura para a matriz de variâncias e covariâncias $\mathbf{D}_*$. Esses comportamentos variáveis entre os frutos de mamoeiro também foram observados por Oliveira et al. (2002), que utilizaram o método instrumental de análise da cor por meio de um colorímetro, e por Fagundes e Yamanishi (2001) e Fonseca et al. (2007), que utilizaram o método visual de análise da cor.

Em seguida, como para o *scanner* a tonalidade média da casca dos frutos são observadas em intervalos iguais de tempo, foi calculada a função de autocorrelação empírica (ACF) usando os pares de resíduos dentro de um mesmo indivíduo. A partir da Figura 14(a) e da Figura 14(b), nota-se que não há autocorrelação entre os pares de resíduos, indicando que a estrutura de covariâncias auto-regressiva de primeira ordem não é necessária para explicar a variabilidade dos dados.

Tabela 5 – Estatística χ^2 para o teste da razão de verossimilhanças (TRV) para seleção da matriz de covariâncias $\mathbf{D}_*$ a partir do modelo M11, para o *scanner*, bem como os critérios de informação AIC e BIC

Estrutura para $\mathbf{D}_*$	AIC	BIC	χ^2	gl	valor-p
CV	972,1	990,6			
SC	1288,2	1310,4	314,1	1	<0,001
CV _H	873,1	899,0	103,0	2	<0,001
NE	855,5	892,6	126,6	5	<0,001
SC	1288,2	1310,4			
CV _H	873,0	898,9	417,1	1	<0,001
NE	855,5	892,6	440,7	4	<0,001
CV _H	873,0	898,9			
NE	855,5	892,6	23,6	3	<0,001

Tabela 6 – Estatística χ^2 para o teste da razão de verossimilhanças (TRV) para seleção da matriz de covariâncias $\mathbf{D}_*$ a partir do modelo M11, para o colorímetro, bem como os critérios de informação AIC e BIC

Estrutura para $\mathbf{D}_*$	AIC	BIC	χ^2	gl	valor-p
CV	929,7	947,6			
SC	1157,4	1178,7	225,6	1	<0,001
CV _H	829,5	854,5	104,2	2	<0,001
NE	818,0	853,7	121,7	5	<0,001
SC	1157,4	1178,7			
CV _H	829,5	854,5	329,9	1	<0,001
NE	818,0	853,7	347,3	4	<0,001
CV _H	829,5	854,5			
NE	818,0	853,7	17,5	3	<0,001

Já para a metodologia do colorímetro, a tonalidade média da casca dos frutos não são observados em intervalos iguais de tempo, pois foram desconsiderados dois dias na análise. Assim, optou-se por utilizar o semivariograma em função da distância para explicar a correlação no tempo. A Figura 14(c) mostra que não há correlação temporal

entre os pares de resíduos padronizados, apesar da autocorrelação positiva no início do experimento. Dessa forma, não há necessidade de considerar uma estrutura de correlação temporal no modelo.

Figura 14 – Função de autocorrelação empírica (a) e autocorrelação parcial para a metodologia do *scanner* (b) e semivariograma em função da distância para a metodologia do colorímetro, com a utilização de um método de suavização LOESS para facilitar a visualização dos padrões (c)

Portanto, as Figuras 14(a), 14(b) para o *scanner* e a Figura 14(c) para o colorímetro mostram que, aparentemente, não há necessidade de incluir uma estrutura autorregressiva, apesar dos dados serem observados ao longo do tempo.

Para verificar a homoscedasticidade do resíduo condicional padronizado utilizou-se o gráfico dos resíduos padronizados em função do tempo, sendo apresentado na Figura

15(a) e 15(b) para o *scanner* e para o colorímetro, respectivamente. De forma geral, a Figura 15 mostra que a variabilidade dos resíduos condicionais padronizados para o colorímetro (Figura 15(b)) é maior do que para o *scanner* (Figura 15(a)). Nota-se também que os resíduos condicionais padronizados, tanto para o *scanner* quanto para o colorímetro, apresentam heterocedasticidade, sendo mais evidente no caso do colorímetro.

Figura 15 – Gráfico dos resíduos condicionais padronizados em função do tempo após o início do experimento para o *scanner* (a) e para o colorímetro (b)

Assim, a partir do modelo M11 e da estrutura selecionada para a matriz $\mathbf{D}_*$, foram testadas diferentes estruturas para a matriz Σ_i . Embora os gráficos de autocorrelação (Figuras 14(a), 14(b)) indiquem que não há necessidade de estruturas auto regressivas, optou-se por considerar as estruturas AR(1) e a AR(1) com uma função de variância exponencial para a matriz Σ_i , para descartar, por meio de testes formais, a necessidade dessas estruturas. No caso do colorímetro, essas estruturas não foram consideradas para a matriz Σ_i , uma vez que a tonalidade média da casca dos frutos não foram observadas em intervalos iguais de tempo.

Os resultados do teste da razão de verossimilhanças (TRV) para seleção da matriz de variâncias e covariâncias Σ_i são apresentados nas Tabelas 7 e 8 para o *scanner* e para o colorímetro, respectivamente. De forma geral, para ambas as metodologias de análise da cor, observa-se que foi necessário admitir a estrutura CV_H para incorporar, aos respectivos modelos, a heterocedasticidade dos resíduos condicionais, discutida anteriormente para cada metodologia.

Como, para ambas as metodologias, a estrutura da matriz Σ_i escolhida não

Tabela 7 – Estatística χ^2 para o teste da razão de verossimilhanças (TRV) para seleção da matriz de covariâncias Σ_i a partir do modelo M11, para o *scanner*, bem como os critérios de informação AIC e BIC

Estrutura para Σ_i	AIC	BIC	χ^2	gl	valor-p
CV	855,5	892,6			
AR(1)	849,7	890,5	7,4	1	0,006
AR(1) _{Exp} [*]	849,5	893,9	10,0	2	0,007
CV _H	835,7	935,7	53,7	17	< 0,001
AR(1)	849,7	890,5			
AR(1) _{Exp} [*]	849,5	893,9	2,7	1	0,1019
CV _H	835,7	935,7	46,4	16	< 0,001

* estrutura de covariâncias AR(1) com uma função de variância exponencial

Tabela 8 – Estatística χ^2 para o teste da razão de verossimilhanças (TRV) para seleção da matriz de covariâncias Σ_i a partir do modelo M11, para o *scanner*, bem como os critérios de informação AIC e BIC

Estrutura para Σ_i	AIC	BIC	χ^2	gl	valor-p
CV	818,0	853,7			
CV _H	770,9	860,0	77,2	15	< 0,001

foi a estrutura assumida inicialmente, $\Sigma_i = \mathbf{I}\sigma^2$, realizou-se uma nova seleção dos modelos apresentados na Tabela 1, assumindo a estrutura selecionada para a matriz de variâncias e covariâncias $\mathbf{D}_*$ e Σ_i . Os resultados da nova seleção são apresentados nas Tabelas 9 e 10 para o *scanner* e colorímetro, respectivamente. Em ambos os casos, verificou-se que o modelo M11 deve ser mantido, ou seja, mesmo modificando a estrutura de variâncias e covariâncias da matriz Σ_i para uma estrutura mais complexa, o modelo polinomial de segundo grau, com efeito aleatório adicionado a todos os parâmetros de efeitos fixos, ainda é o que melhor explica a variabilidade dos dados de tonalidade média da casca do fruto.

Portanto, o mesmo preditor linear e matrizes de variâncias e covariâncias foram utilizadas, indicando que, independentemente da metodologia de análise da cor, os indivíduos apresentam comportamentos distintos.

Assim, o modelo resultante tanto para o *scanner* quanto para o colorímetro

Tabela 9 – Estatística χ^2 para o teste da razão de verossimilhanças (TRV) entre o modelo M11 e os modelos aninhados subsequentes M10, M9 e M8, assumindo estrutura NE para a matriz $\mathbf{D}_*$ e $\text{CV}_{\mathbf{H}}$ para Σ_i ; critérios de informação AIC e BIC para os modelos M1 a M11 ajustados aos dados de tonalidade média obtidos pelo *scanner* e respectivos números de parâmetros do modelo (np), números de graus de liberdade (gl) e valores-p do teste

Modelo	AIC	BIC	np	χ^2	gl	valor-p
M11	857,83	957,92	27			
M10	886,6	975,52	24	34,7	3	0,001
M9	978,6	1067,5	24	126,7	3	<0,001
M8	934,5	1023,5	24	82,7	3	<0,001
M7	1284,6	1366,2	22			
M6	1081,7	1163,3	22			
M5	1156,6	1238,2	22			
M4	969,1	1054,4	23			
M3	1100,9	1178,8	21			
M2	1149,9	1227,8	21			
M1	2017,2	2091,5	20			

é dado por

$$\hat{\mathbf{y}}_i = \hat{\beta}_0 + \tilde{\mathbf{b}}_{0i} + (\hat{\beta}_1 + \tilde{\mathbf{b}}_{1i}) \mathbf{x}_i + (\hat{\beta}_2 + \tilde{\mathbf{b}}_{2i}) \mathbf{x}_i^2,$$

$$\mathbf{D}_* = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 \end{bmatrix} \text{ e } \Sigma_i = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{n_i}^2 \end{bmatrix}. \quad (38)$$

Em seguida, foram calculadas, por meio da expressão (37), os intervalos de confiança dos parâmetros considerando um nível de 5% de significância. Os resultados são apresentados na Tabela 11 para o *scanner* e na Tabela 12 para o colorímetro.

Uma vez que o modelo selecionado para a tonalidade média da casca de frutos avaliados por meio de um *scanner* é o mesmo para frutos avaliados por meio de um colorímetro, podem-se considerar que o intervalo de confiança para β_0 e β_2 do modelo

Tabela 10 – Estatística χ^2 para o teste da razão de verossimilhanças (TRV) entre o modelo M11 e os modelos aninhados subsequentes M10, M9 e M8, assumindo estrutura NE para a matriz $\mathbf{D}_*$ e $\mathbf{CV}_H$ para $\mathbf{\Sigma}_i$; critérios de informação AIC e BIC para os modelos M1 a M11 ajustados aos dados de tonalidade média obtidos pelo colorímetro e respectivos números de parâmetros do modelo (np), números de graus de liberdade (gl) e valores-p do teste

Modelo	AIC	BIC	np	χ^2	gl	valor-p
M11	770,9	860,0	25			
M10	848,8	927,2	22	48,4	3	0,002
M9	871,7	950,1	22	139,7	3	<0,001
M8	809,5	887,9	22	70,9	3	<0,001
M7	1133,4	1204,7	20			
M6	966,8	1038,1	20			
M5	1155,4	1226,7	20			
M4	980,9	1055,8	21			
M3	1072,1	1139,9	19			
M2	1198,8	1266,6	19			
M1	1867,3	1931,6	18			

Tabela 11 – Estimativas dos efeitos fixos do modelo (38) e os respectivos limites inferior (LI) e superior (LS) do intervalo de confiança aproximado, considerando um nível de 5% de significância, para o *scanner*

Efeito Fixo	LI	Estimativa	LS
β_0	111,0935	111,6256	112,1577
β_1	-1,6590	-1,4073	-1,1555
β_2	-0,0405	-0,0279	-0,0153

Tabela 12 – Estimativas dos efeitos fixos do modelo (38) e os respectivos limites inferior (LI) e superior (LS) do intervalo de confiança aproximado, considerando um nível de 5% de significância, para o colorímetro

Efeito Fixo	LI	Estimativa	LS
β_0	115,4669	116,0095	116,5522
β_1	-1,5474	-1,3192	-1,0909
β_2	-0,0672	-0,0565	-0,0457

associado ao *scanner* não está sobreposto ao intervalo de confiança para β_0 e β_2 do modelo associado ao colorímetro (Tabela 12). Por outro lado, o mesmo não ocorre com o efeito fixo β_1 .

A não sobreposição dos intervalos de confiança para β_0 , entre os modelos para

o *scanner* e para o colorímetro, sugere que a tonalidade média inicial da casca de frutos avaliados por meio de um *scanner* seja diferente da tonalidade média inicial da casca de frutos avaliados por meio de um colorímetro. Isso indica que quando o fruto foi avaliado em seu estágio inicial de maturação por meio de um colorímetro, a tonalidade média resultante a partir dessa metodologia para esse fruto é diferente da tonalidade média resultante da metodologia utilizando-se um *scanner*. De forma geral, quando utilizou-se um *scanner* a tonalidade média inicial resultante foi menos esverdeada do que a tonalidade média inicial resultante da metodologia utilizando-se um colorímetro.

Vale ressaltar que quando utilizou-se a metodologia do colorímetro a região amostrada foi a região equatorial da casca do fruto, enquanto que com a metodologia do *scanner* a região amostrada foi, praticamente, toda a superfície da casca do mesmo. Dessa forma, teoricamente as metodologias não são comparáveis, no entanto, na prática é comum a utilização dos pontos observados na região equatorial da casca do fruto para descrever a tonalidade média de toda superfície da casca do mesmo, implicando, assim, em um provável viés de amostragem, principalmente nos estádios de maturação iniciais do fruto.

Na prática, uma possível alternativa para tentar reduzir o problema observado com o colorímetro seria subdividir os frutos em 3 regiões, sendo elas, a região próxima a inserção peduncular, a região equatorial e a região estilar do fruto, de forma que em cada região, devem ser observados 4 pontos equidistantes ao longo dessas secções transversais, como ilustra a Figura 16.

Figura 16 – Proposta de amostragem, por meio de um colorímetro, da tonalidade do fruto da cultivar “Sunrise Solo” com pontos em vermelho representando locais no fruto onde podem ser realizadas as observações

As predições dos efeitos aleatórios são apresentadas na Tabela 13 para o *scanner* e na Tabela 14 para o colorímetro. Vale ressaltar que, em ambas as metodologias, quanto menores os valores preditos $\tilde{b}_{1i}$ e $\tilde{b}_{2i}$ mais rápida é a maturação do fruto e quanto maiores os valores preditos mais lento é o processo de maturação do mesmo. Dessa forma pode-se associar os parâmetros de efeito aleatório $\tilde{b}_{1i}$ e $\tilde{b}_{2i}$ a uma classificação de grupos de maturação em “lento” e “normal”. Já em relação ao parâmetro de efeito aleatório $\tilde{b}_{0i}$, quanto menor o valor predito mais maduro o fruto estava no primeiro dia de análise. Quanto a esse aspecto, os frutos 12 e 18 se destacam por serem mais maduros segundo a metodologia de análise de imagens digitalizadas por meio de um *scanner*, o que confirma a análise exploratória gráfica apresentada no item 5.1 dessa seção. Por meio do colorímetro, somente o fruto 18 se destaca dos demais por apresentar tonalidade média inicial mais amarelada.

Tabela 13 – Predições dos efeitos aleatórios do modelo apresentado na expressão (38) da tonalidade média de frutos medidos pelo *scanner* e velocidades de maturação para cada indivíduo após 8 dias do início do experimento (v_{8_i}) em graus por dia

Fruto	$\tilde{b}_{0i}$	$\tilde{b}_{1i}$	$\tilde{b}_{2i}$	v_{8_i}	Fruto	$\tilde{b}_{0i}$	$\tilde{b}_{1i}$	$\tilde{b}_{2i}$	v_{8_i}
1	-1,4994	0,5524	-0,0085	1,46	11	1,5638	0,8957	-0,0332	1,48
2	0,0668	-0,9035	0,0505	1,97	12	-0,5247	-0,6779	0,0387	1,92
3	1,3004	1,0409	-0,0428	1,52	13	0,7480	0,0488	-0,0131	2,02
4	1,7664	-0,3247	-0,0132	2,44	14	-0,3336	-0,3279	0,0016	2,18
5	-0,3440	0,0517	0,0016	1,80	15	-0,7740	-0,3735	0,0126	2,03
6	0,8976	1,0984	-0,0333	1,30	16	-0,2246	-0,2178	0,0109	1,88
7	-0,8254	0,1549	-0,0142	1,96	17	0,7479	0,0464	-0,0045	1,88
8	-0,6327	0,0563	0,0054	1,74	18	-1,9680	-0,4478	0,0307	1,82
9	-0,6579	-0,4063	0,0252	1,87	19	0,8081	-0,1194	-0,0310	2,50
10	-1,2669	-0,1101	0,0233	1,61	20	1,1523	-0,0366	-0,0068	2,00

Tabela 14 – Predições dos efeitos aleatórios do modelo apresentado na expressão (38) da tonalidade média de frutos medidos pelo colorímetro e velocidades de maturação para cada indivíduo após 8 dias do início do experimento (v_{8_i})

Fruto	$\tilde{b}_{0i}$	$\tilde{b}_{1i}$	$\tilde{b}_{2i}$	v_{8_i}	Fruto	$\tilde{b}_{0i}$	$\tilde{b}_{1i}$	$\tilde{b}_{2i}$	v_{8_i}
1	-1,0346	0,3265	0,0075	1,78	11	1,5424	0,8900	-0,0108	1,51
2	0,5018	-0,5693	0,0229	2,43	12	-2,7550	-0,7297	0,0466	2,21
3	1,6009	0,7920	-0,0198	1,75	13	0,7735	-0,1037	-0,0097	2,48
4	0,7811	-0,1707	-0,0279	2,84	14	-0,7582	-0,4502	-0,0008	2,68
5	-0,2859	-0,1649	0,0208	2,06	15	-0,5860	-0,1714	-0,0008	2,41
6	1,1610	0,7842	-0,0045	1,51	16	-0,9278	-0,2285	0,0013	2,43
7	-0,2907	0,0736	-0,0269	2,58	17	0,2312	0,3730	-0,0121	2,04
8	-0,2831	0,0844	0,0080	2,01	18	-2,1328	-0,6852	0,0384	2,29
9	-0,5403	-0,3163	0,0130	2,33	19	0,7827	-0,4029	-0,0186	2,92
10	0,8840	0,4912	-0,0115	1,92	20	1,3357	0,1778	-0,0151	2,29

As Tabelas 13 e 14 apresentam também as velocidades de maturação dos frutos aos 8 dias após o início do experimento, obtidas por meio da derivada primeira dada, para cada indivíduo, por:

$$\begin{aligned}\hat{v}_{8_i} &= -\frac{d\hat{y}_i}{dx_i}\Big|_{x_i=8} = -\left[\hat{\beta}_1 + \tilde{b}_{1i} + 2\left(\hat{\beta}_2 + \tilde{b}_{2i}\right)8\right] \\ &= -\left[\hat{\beta}_1 + \tilde{b}_{1i} + 16\left(\hat{\beta}_2 + \tilde{b}_{2i}\right)\right].\end{aligned}$$

Convém ressaltar que foi escolhido o oitavo dia por ser a metade do tempo de duração do experimento e que a velocidade foi multiplicada por menos um de modo a traduzir a ideia de velocidade de maturação, ou seja, quanto mais negativa a taxa de variação instantânea da tonalidade, maior a velocidade de maturação.

Assim, analisando as Tabelas 13 e 14, observa-se, tanto para o *scanner* quanto para o colorímetro, que os frutos com menores velocidades de maturação foram o 1, 3, 6 e 11. Isso indica que, independentemente da região observada, é possível detectar os frutos que apresentam maturação fisiológica lenta, apesar de a metodologia do colorímetro não ter levado a uma distinção clara entre esses grupos na análise descritiva apresentada na subseção 5.1.

Para se ter uma ideia da eventual existência de relação entre a velocidade de maturação do fruto aos 8 dias e a tonalidade média inicial das cascas dos mesmo, foram construídos os gráficos da Figura 17. A partir dessa Figura, nota-se que tanto para a metodologia do *scanner* (Figura 17(a)) quanto para a metodologia do colorímetro (Figura 17(b)) não existe, aparentemente, uma relação entre o ponto de colheita dos frutos e o estágio fisiológico dos mesmos, indo contra os resultados apresentados por Fonseca et al. (2007). Note, entretanto, que se tivessem sido selecionados frutos com tonalidades médias iniciais de casca mais diferentes para o experimento, talvez tal relação pudesse ter sido observada.

A partir das Tabelas 13 e 14 e da Figura 17, é importante notar que, independentemente da metodologia de análise da cor, existem frutos que apresentam, ao mesmo tempo, velocidade de maturação e tonalidade média inicial altas, como por exemplo os frutos 4 e 20 e frutos que apresentam, ao mesmo tempo, velocidade de maturação baixa e tonalidade média inicial alta, como por exemplo os frutos 3, 6 e 11. Dessa forma, aparentemente, os frutos que apresentam maturação fisiológica “lenta” (1, 3, 6 e 11) não estão

Figura 17 – Gráfico da tonalidade média estimada inicial da casca do fruto ($\hat{\beta}_0 + \tilde{b}_{0i}$) em função da velocidade de maturação aos 8 dias após o início do experimento (v_{8i}) para o *scanner* (a) e para o colorímetro (b)

relacionados ao ponto de colheita dos mesmos. Isso indica a provável existência de outros fatores fisiológicos que diretamente ou indiretamente interferem na velocidade de maturação dos frutos.

A Tabela 15 apresenta, para ambas as metodologias, os intervalos de 95% de confiança para os desvios padrões σ_{b_0} , σ_{b_1} e σ_{b_2} e para as correlações entre os efeitos aleatórios b_0 e b_1 , b_0 e b_2 e b_1 e b_2 . De forma geral, os intervalos de confiança para $\hat{\sigma}_{b_0}$, $\hat{\sigma}_{b_1}$ e $\hat{\sigma}_{b_2}$ são próximos para ambas as metodologias, indicando que independentemente da região amostrada na casca do fruto as variabilidades entre a tonalidade média da casca, medida por meio de um *scanner* e de um colorímetro, são próximas.

A partir da matriz de covariâncias dos efeitos aleatórios do modelo dado na expressão (38) para o *scanner*

$$\mathbf{D}_{*S} = \begin{bmatrix} 1,2862 & 0,1793 & -0,0145 \\ 0,1793 & 0,3056 & -0,0124 \\ -0,0145 & -0,0124 & 0,0266 \end{bmatrix}$$

nota-se que as variâncias dos efeitos aleatório são diferentes e existe uma relação positiva entre os efeitos aleatórios b_0 e b_1 e negativa entre os efeitos aleatórios b_0 e b_2 e entre b_1 e b_2 . Da mesma forma, a matriz de covariâncias dos efeitos aleatórios do modelo dado na

Tabela 15 – Estimativas e limites inferiores (LI) e superiores (LS) dos intervalos de confiança, com nível de 5% de probabilidade, para o desvio padrão do intercepto ($\sigma_{\mathbf{b}_0}$), do efeito linear ($\sigma_{\mathbf{b}_1}$) e do efeito quadrático ($\sigma_{\mathbf{b}_2}$) da matriz de variâncias e covariâncias $\mathbf{D}_*$, bem como os intervalos de confiança, com nível de 5% de probabilidade, para o coeficiente de correlação entre os efeitos aleatórios $\mathbf{b}_0$ e $\mathbf{b}_1$ ($\rho(\mathbf{b}_0, \mathbf{b}_1)$), $\mathbf{b}_0$ e $\mathbf{b}_2$ ($\rho(\mathbf{b}_0, \mathbf{b}_2)$) e $\mathbf{b}_1$ e $\mathbf{b}_2$ ($\rho(\mathbf{b}_1, \mathbf{b}_2)$) para o *scanner* e para o colorímetro

Termo	<i>Scanner</i>			Colorímetro		
	LI	Estimativa	LS	LI	Estimativa	LS
$\sigma_{\mathbf{b}_0}$	0,7698	1,1341	1,6708	0,8267	1,2019	1,7473
$\sigma_{\mathbf{b}_1}$	0,3922	0,5528	0,7793	0,3482	0,4985	0,7136
$\sigma_{\mathbf{b}_2}$	0,0186	0,0266	0,0381	0,0144	0,0220	0,0336
$\rho_1(\mathbf{b}_0, \mathbf{b}_1)$	-0,2478	0,2860	0,6866	0,1926	0,6243	0,8535
$\rho_2(\mathbf{b}_0, \mathbf{b}_2)$	-0,8088	-0,4790	0,0801	-0,8909	-0,6438	-0,1025
$\rho_3(\mathbf{b}_1, \mathbf{b}_2)$	-0,9404	-0,8432	-0,6187	-0,8421	-0,5982	-0,1514

expressão (38) para o colorímetro

$$\mathbf{D}_{*C} = \begin{bmatrix} 1,4445 & 0,3740 & -0,0170 \\ 0,3740 & 0,2484 & -0,0066 \\ -0,0170 & -0,0066 & 0,0005 \end{bmatrix}$$

possui as mesmas relações entre os efeitos aleatórios preditos para o *scanner*.

Para avaliar essas relações entre os efeitos aleatórios utilizou-se o coeficiente de correlação linear de Pearson entre os mesmos, cujo gráfico de dispersão são apresentadas na Figura 18(a) para o *scanner* e na Figura 18(b) para o colorímetro. Os intervalos de confiança para as correlações entre os efeitos aleatórios foram apresentados na Tabela 15 para ambas as metodologias.

Os efeitos aleatórios $\mathbf{b}_0$ e $\mathbf{b}_1$ apresentam correlação linear positiva, porém, como o intervalo de confiança para $\rho(\mathbf{b}_0, \mathbf{b}_1)$ inclui o zero, a correlação entre eles não é significativa (Tabela 15). Já a correlação linear entre os efeitos aleatórios $\mathbf{b}_0$ e $\mathbf{b}_2$ é negativa, porém, o intervalo de confiança para $\rho(\mathbf{b}_0, \mathbf{b}_2)$ também inclui o zero, indicando que a correlação entre eles não é significativa (Tabela 15). Por outro lado, a correlação linear entre os efeitos aleatórios $\mathbf{b}_1$ e $\mathbf{b}_2$ é negativa e alta (considerada, aqui, $|\rho| > 0,8$), além disso, o intervalo de confiança para $\rho(\mathbf{b}_1, \mathbf{b}_2)$ não inclui o zero, assim, a correlação entre $\mathbf{b}_1$ e $\mathbf{b}_2$ é significativa a 5%.

Apesar do intervalo de confiança para $\rho(\mathbf{b}_1, \mathbf{b}_2)$ não incluir o zero, quando avaliou-se a necessidade de remoção do efeito aleatório $\mathbf{b}_2$ do modelo para o *scanner*, o valor-p, do teste razão de verossimilhanças, foi menor do que 0,001, indicando que ambos

Figura 18 – Gráficos de dispersão entre os efeitos aleatórios do modelo apresentado na expressão (38) e coeficientes de correlação linear entre os mesmos, considerando-se os dados obtidos por meio do *scanner* (a) e do colorímetro (b)

os efeitos aleatórios devem permanecer no modelo.

Já para o colorímetro, a Figura 18(b) e a Tabela 15 mostram que a correlação entre os efeitos aleatórios $\mathbf{b}_0$ e $\mathbf{b}_1$ é positiva e o intervalo de confiança para $\rho(\mathbf{b}_0, \mathbf{b}_1)$ não inclui o zero, indicando que a correlação entre $\mathbf{b}_0$ e $\mathbf{b}_1$ é significativa. Por outro lado, a correlação entre $\mathbf{b}_0$ e $\mathbf{b}_2$ e entre $\mathbf{b}_1$ e $\mathbf{b}_2$ são negativas e os respectivos intervalos de confiança para $\rho(\mathbf{b}_0, \mathbf{b}_2)$ e $\rho(\mathbf{b}_1, \mathbf{b}_2)$ não incluem o zero, indicando que ambas as correlações são significativas.

Da mesma forma que para o *scanner*, ao avaliar a necessidade de remoção de qualquer efeito aleatório do modelo para o colorímetro, o valor-p, do teste razão de verossimilhanças, sempre foi menor do que 0,001, indicando que os efeitos aleatórios também devem permanecer no modelo, apesar da correlação relativamente alta entre eles.

Um caso semelhante foi apresentado por Pinheiro e Bates (1995, p. 195), no qual a correlação entre os efeitos aleatórios foi $-0,96$ e, ao realizar o teste razão de verossimilhanças para avaliar a necessidade de retirar um dos efeitos aleatórios, o valor-p também foi pequeno, mantendo-se no modelo ambos os efeitos aleatórios, apesar da correlação alta entre eles. Uma alternativa proposta pelos autores foi a de ajustar um modelo não-linear de efeitos mistos aos dados. Assim, em um trabalho futuro, pode-se considerar o ajuste de um modelo não-linear de efeitos mistos aos dados do *scanner* e do colorímetro como alternativa ao modelo apresentado na expressão (38).

Para verificar se a estrutura CV_H incorporou ao modelo a heterogeneidade de

variâncias para o *scanner* e para o colorímetro, utilizou-se o gráfico dos resíduos condicionais padronizados em função do tempo, apresentados na Figura 19 para o *scanner* e para o colorímetro. De forma geral, para ambas as metodologias, nota-se que aparentemente a estrutura CV_H incorporou ao modelo a heterocedasticidade.

Figura 19 – Gráfico dos resíduos condicionais padronizados em função do tempo após o início do experimento para o *scanner* (a) e para o colorímetro (b)

A partir da função de variâncias dada pela expressão (35), utilizada para a construção da estrutura de variâncias e covariâncias CV_H , calcularam-se os intervalos de 95% de confiança para os parâmetros $\delta_{s_{ij}}$ de variância relativos aos K níveis de estratificação (no caso, número de dias após o início do experimento) para a metodologia do *scanner* e do colorímetro. Note que utilizou-se a restrição de que $\delta_{1_{ij}} = 1$, assim, $\delta_{s_{ij}}$, $s = 2, \dots, 18$, representa a razão entre o desvio padrão do s -ésimo estrato e o primeiro estrato.

Em relação ao *scanner*, a Tabela 16 e a Figura 20(a) apresentam as estimativas para os parâmetros $\delta_{s_{ij}}$ de variância relativos aos K níveis de estratificação, utilizando o fruto 2 ($i = 2$) como exemplo, bem como os limites inferior e superior do intervalo de 95% de confiança para $\delta_{s_{ij}}$. É importante notar que, de forma geral, as estimativas do parâmetro $\delta_{s_{ij}}$ variaram de 0,3 a 0,7, indicando que utilização do parâmetro $\delta_{s_{ij}}$ foi de fundamental importância para corrigir a heterocedasticidade ao longo do tempo. Além disso, a amplitude dos intervalos de 95% de confiança para o parâmetro $\delta_{s_{ij}}$ aumentou nos dias 14, 15 e 17, podendo estar diretamente relacionado à redução do número de frutos avaliados de 12 para 4 a partir do dia 14.

Já em relação ao colorímetro, a Tabela 17 e a Figura 20(b) apresentam,

Tabela 16 – Estimativas para os $\delta_{s_{2j}}$, bem como os respectivos limites inferior (LI) e superior (LS) do intervalo de confiança aproximado, considerando um nível de 5% de significância, para o parâmetro utilizando-se a metodologia do *scanner*; o fruto 2 ($i=2$) foi utilizado como exemplo uma vez que apresenta maior número de medidas repedidas

s	Dia	LI	Estimativa	LS	s	Dia	LI	Estimativa	LS
-	0	-	1,0000	-	10	9	0,3158	0,5250	0,8727
2	1	0,2344	0,4262	0,7748	11	10	0,0904	0,1887	0,3939
3	2	0,2197	0,4002	0,7288	12	11	0,2618	0,4377	0,7317
4	3	0,3536	0,5769	0,9414	13	12	0,3002	0,5066	0,8550
5	4	0,1672	0,2978	0,5306	14	13	0,1032	0,2295	0,5105
6	5	0,2663	0,4453	0,7449	15	14	0,3766	0,6671	1,1816
7	6	0,2905	0,4801	0,7934	16	15	0,3063	0,6145	1,2327
8	7	0,1790	0,3077	0,5289	17	16	0,0009	0,0019	0,0036
9	8	0,2809	0,4667	0,7753	18	17	0,1704	0,3769	0,8338

(a) *Scanner*

(b) Colorímetro

Figura 20 – Gráfico das estimativas para os parâmetros $\delta_{s_{ij}}$ de variância relativos aos K níveis de estratificação (linha azul), bem como os limites superior e inferior do intervalo de 95% de confiança para os parâmetros $\delta_{s_{ij}}$ de variância relativos aos K níveis de estratificação (linha preta) para o *scanner* (a) e para o colorímetro (b), utilizando o fruto 2 como exemplo

utilizando o fruto 2 como exemplo, as estimativas do parâmetro $\delta_{s_{ij}}$, bem como os limites inferior e superior do intervalo de 95% de confiança de $\delta_{s_{ij}}$ para o colorímetro. Da mesma forma que para o *scanner*, foi de fundamental importância a utilização do parâmetro $\delta_{s_{ij}}$ para corrigir a heterocedasticidade ao longo do tempo, uma vez que, de forma geral, as estimativas desse parâmetro variaram de 0,2 a 1,0. Nota-se, também, que a amplitude dos intervalos de 95% de confiança para o parâmetro $\delta_{s_{ij}}$ aumentou nos dias 15, 16 e 17, podendo estar diretamente relacionado à redução do número de frutos avaliados de 12 para 4 a partir do dia 14.

Tabela 17 – Estimativas para os $\delta_{s_{2j}}$, bem como os respectivos limites inferior (LI) e superior (LS) do intervalo de confiança aproximado, considerando um nível de 5% de significância, para o parâmetro utilizando-se a metodologia do colorímetro; o fruto 2 ($i=2$) foi utilizado como exemplo uma vez que apresenta maior número de medidas repedidas

s	Dia	LI	Estimativa	LS	s	Dia	LI	Estimativa	LS
-	0	-	1,0000	-	9	10	0,4032	0,6305	0,9861
2	1	0,0828	0,2081	0,5231	10	11	0,3712	0,5867	0,9273
3	2	0,1847	0,3087	0,5158	11	12	0,5257	0,7863	1,1760
4	3	0,1758	0,3106	0,5487	12	13	0,2773	0,5183	0,9686
5	5	0,5370	0,7933	1,1720	13	14	0,1448	0,3491	0,8421
6	6	0,4414	0,6726	1,0251	14	15	0,5685	0,9831	1,7001
7	8	0,8643	1,2492	1,8054	15	16	0,2104	0,6453	1,9789
8	9	0,8180	1,1613	1,6486	16	17	1,0201	2,2065	4,7727

Uma provável explicação para estimativas e para intervalos de confiança maiores, principalmente, nos dias 7 e 8 (Tabela 17 e Figura 20(b)) é a redução da clorofila na região equatorial do fruto, gerando, por consequência, o aumento da síntese de carotenoides, aumentando, por consequência, a variabilidade da cor da casca nessa região.

Como o amadurecimento do fruto ocorre em faixas que partem da região estilar em direção a região peduncular do fruto, torna-se fundamental o emprego da técnica de imagens digitalizadas no estudo da cor. Novamente, para o método instrumental por meio de um colorímetro, deve-se ressaltar que a forma usual de amostragem é insuficiente para o cálculo da tonalidade média da casca nos estádios iniciais do fruto quando o objetivo é determinar a tonalidade média de toda a superfície da casca do fruto.

De forma geral, os resultados apresentados para o colorímetro na Tabela 17, mostram, novamente, a necessidade de amostrar um maior número de pontos por fruto, além da necessidade de desconsiderar a ideia de observá-los apenas na região equatorial mesmo. Esses resultados confluem com os resultados apresentados por Mendoza e Aguilera (2004), os quais verificaram que a utilização de imagens digitalizadas aplicadas a análise de maturidade fisiológica da casca de bananas é mais versátil, mais consistente, mais eficiente e mais econômica do que a identificação dos estádios de maturação fisiológica por meio de técnicas instrumentais, como por exemplo as que utilizam colorímetros manuais.

Também foram calculados os intervalos de 95% de confiança para o erro padrão dos modelos de ambas as metodologias de análise da cor. Para a metodologia do *scanner*, o intervalo de confiança para o erro padrão é $[0,9803; 1,9893]$, sendo sua estimativa igual a 1,3966. Para a metodologia do colorímetro, por sua vez, o intervalo de confiança para o erro padrão é $[0,6877; 1,4135]$, sendo sua estimativa igual a 0,9859.

Segundo Pinheiro e Bates (1995), quando algum dos intervalos de confiança para os componentes de variâncias apresenta amplitude anormal, como por exemplo uma amplitude de doze ordens de grandeza em relação ao valor estimado do parâmetro, pode ser um indicativo de que há problemas com a definição do modelo proposto. Como, tanto para a metodologia do *scanner* quanto para a do colorímetro, o erro padrão, bem como os demais componentes de variâncias tiveram intervalos de confiança com amplitude pequena, menor do que duas ordens de grandeza em relação ao valor estimado do parâmetro, têm-se que, aparentemente, não houve problemas com a definição dos modelos ajustados aos dados de tonalidade média da casca de mamão.

Quanto à normalidade dos resíduos condicionais padronizados para o *scanner*, pode-se verificar na Figura 21(a) que 26,73% dos pontos estão fora do envelope simulado, sugerindo que, aparentemente, os resíduos não tenham distribuição normal. Apesar disso, ao realizar o teste de Shapiro-Wilk, para verificar a pressuposição de normalidade dos resíduos condicionais padronizados, obtiveram-se um p-valor igual a 0,0840, ou seja, os resíduos condicionais padronizados, para o *scanner*, têm distribuição normal, considerando-se o nível de 5% de significância.

Segundo Hilden-Minton (1995) e Nobre e Singer (2007), problemas envolvendo a normalidade dos resíduos condicionais podem ser devido ao confundimento de $\hat{\epsilon}$ com $\mathbf{b}$, ou seja, $\hat{\epsilon}$ pode não apresentar normalidade quando na verdade ele é normal (ver item 2.7). Assim, em trabalho futuro pode-se considerar os resíduos condicionais padronizados com confundimento mínimo, propostos por Nobre e Singer (2007).

Por outro lado, a Figura 21(b) mostra uma aparente normalidade dos resíduos condicionais para o colorímetro, sendo que apenas 7,14% então fora do envelope simulado. Ao realizar o teste de Shapiro-Wilk, para verificar a pressuposição de normalidade dos resíduos condicionais padronizados, obteve-se um valor-p igual a 0,3481, ou seja, os resíduos condicionais padronizados, para o colorímetro, têm distribuição normal, considerando-se um nível de significância de 5%.

A partir da Figura 22(a), pode-se verificar que uma possível causa para o número de pontos fora do envelope simulado (Figura 21(a)) seja devido a curtose dos resíduos condicionais padronizados. Já para o colorímetro (Figura 22(b)), nota-se que não há problemas de assimetria e de curtose evidentes.

Para elucidar se o problema está relacionado a assimetria ou a curtose dos re-

(a) *Scanner* (b) *Colorímetro*

Figura 21 – Gráfico de normalidade dos resíduos condicionais padronizados para o *scanner* (a) e para o colorímetro (b)

(a) *Scanner* (b) *Colorímetro*

Figura 22 – Histograma de normalidade dos resíduos condicionais padronizados, com adição de uma curva em vermelho indicando a distribuição amostral e uma curva em azul indicando a distribuição normal padrão

síduos condicionais padronizados foram construídos os histogramas apresentados na Figura 23. Para a construção desses histogramas, fez-se uma simulação de 1000 amostras aleatórias da distribuição normal padrão de tamanho 304 para obter a distribuição empírica do coeficiente de assimetria e de curtose dessa distribuição. Foram consideradas amostras de tamanho 304 pois o número de observações totais para a tonalidade média é igual a 303. Em seguida, calculou-se o valor de assimetria e de curtose dos resíduos condicionais padronizados para realizar uma análise descritiva do problema.

Para a metodologia do *scanner* (Figura 23(a)) aparentemente não há problemas de assimetria nos resíduos condicionais padronizados, por outro lado, a Figura 23(b) mostra que, aparentemente, o problema esteja associado a curtose, ou seja, possivelmente os resíduos condicionais padronizados, apresentados na Figura 21(a), estejam concentrados ao redor do valor zero (leptocúrtica), explicando, aparentemente, a causa para o número de pontos fora do envelope simulado (Figura 21(a)).

Já para a metodologia do colorímetro, a Figura 23(c) mostra que, aparentemente, há uma leve assimetria dos resíduos condicionais padronizados, já em relação a curtose, a Figura 23(d) mostra que, aparentemente, não há problemas evidentes.

Figura 23 – Histograma da distribuição empírica do coeficiente de assimetria e de curtose obtidas a partir da simulação de 1000 amostras aleatórias da distribuição normal padrão de tamanho 303, com adição de uma linha vermelha indicando a distribuição empírica e de uma linha azul indicando o valor calculado (V.C.) da assimetria e da curtose dos resíduos condicionais padronizados do modelo da expressão (38) para o *scanner* (a) e (b) e para o colorímetro (c) e (d)

Em seguida, avaliou-se a normalidade dos efeitos aleatórios extraindo-se as

predições dos mesmos por meio da função *ranef* implementada por Pinheiro e Bates (1995). Assim construiu-se o gráfico dos quantis da distribuição normal em função das predições dos efeitos aleatórios para o *scanner* e para o colorímetro, apresentado na Figura 24.

Figura 24 – Gráficos normais de probabilidade com envelope simulado para $\mathbf{b}_{0i}$ ((a) e (b)), $\mathbf{b}_{1i}$ ((c) e (d)) e $\mathbf{b}_{2i}$ ((e) e (f)) para o *scanner* e colorímetro, respectivamente

Para o *scanner*, observa-se que os efeitos aleatórios $\mathbf{b}_{0i}$ (Figura 24(a)) e $\mathbf{b}_{2i}$

(Figura 24(e)) têm distribuição normal, enquanto que o efeito aleatório $\mathbf{b}_{1i}$ (Figura 24(c)) apresenta quatro pontos fora do envelope simulado. No entanto, os valores preditos do efeito aleatório $\mathbf{b}_{1i}$ referentes aos frutos 1, 3, 6 e 11 podem ser justificados, uma vez que esses frutos pertencem ao grupo de maturação fisiológica mais lenta. Isso significa que em análises futuras deve-se levar em consideração dois grupos de maturação fisiológica para mamões cultivar “Sunrise Solo”.

Uma alternativa para resolver o problema de normalidade associado ao efeito aleatório $\mathbf{b}_{1i}$ para o *scanner*, é separar os frutos em dois grupos de maturação: um de maturação “lenta” composto pelos frutos 1, 3, 6 e 11 e outro de maturação “normal” composto pelos demais frutos. Assim, mais ensaios a respeito do ponto ótimo de colheita do mamão “Sunrise Solo” devem ser conduzidos, sendo que a análise de imagens digitalizadas pode ser uma importante ferramenta para estimar o ponto ótimo de colheita.

Já para o colorímetro observa-se que todos os efeitos aleatórios, apresentados na Figura 24(b), 24(d) e 24(f), apresentam distribuição normal, indicando que quando utiliza-se o colorímetro para avaliar a tonalidade média é mais difícil distinguir os grupos de maturação fisiológica.

Em seguida, elaborou-se o gráfico dos resíduos condicionais padronizados em função dos valores preditos para o *scanner* e para o colorímetro, apresentado na Figura 25. As Figura 25(a) e 25(b) aparentemente não mostram nenhum padrão nos resíduos, sendo que os mesmos estão distribuídos de forma aleatória em torno da média zero, satisfazendo a condição de independência.

Também verificou-se o comportamento dos resíduos condicionais padronizados individuais em função do tempo após o início do experimento, uma vez que esse procedimento é importante para identificar padrões de comportamento nos indivíduos. A Figura 26(a) revela um mesmo padrão no comportamento dos resíduos condicionais dos frutos 7, 14, 16 e 19 medidos pelo *scanner*. Por outro lado, o modelo dado na expressão (38) consegue explicar bem a evolução da tonalidade média ao longo do tempo para os demais frutos. Além disso, a inclusão de uma estrutura de covariâncias autorregressiva de ordem um com heterocedasticidade de variâncias não foi significativo pelo teste razão de verossimilhanças (Tabela 7).

Já para o colorímetro, a Figura 26(b) revela um padrão de comportamento nos resíduos condicionais dos frutos 2, 5, 7, 11 e 18, no entanto, para os demais frutos o

Figura 25 – Gráfico dos resíduos condicionais padronizados em função dos valores preditos para o *scanner* (a) e para o colorímetro (b)

modelo consegue explicar bem a evolução da tonalidade média ao longo do tempo. Assim como para o *scanner*, a utilização de uma estrutura de correlação não é necessária (Figura 14(c)).

Figura 26 – Resíduos condicionais padronizados individuais em função do tempo e curvas não paramétricas de regressão LOESS para o modelo apresentado na expressão (38 para o *scanner* (a) e para o colorímetro (b)

Segundo Pinheiro e Bates (1995), quando se detectam padrões periódicos nos resíduos individuais e os resíduos estão concentrados ao redor da média zero, pode-se incorporar essa periodicidade ao modelo por meio de funções trigonométricas. No entanto, as Figuras 26(a) e 26(b) mostram que apenas alguns frutos apresentam certa periodicidade

ao longo do tempo, sendo esse o motivo principal de não incorporar essa periodicidade ao modelo, em nenhuma das metodologias.

Para avaliar a qualidade do ajuste do modelo foi utilizado o gráfico dos valores ajustados em função dos valores observados para ambas as metodologias de análise da cor (Figura 27). As Figuras 27(a) e 27(b) mostram que os valores estimados estão próximos aos valores observados, indicando que o modelo dado pela expressão (38) consegue explicar a variabilidade da tonalidade média de frutos de mamoeiro “Sunrise Solo” para ambas as metodologias.

Figura 27 – Gráfico dos valores ajustados (em graus) em função dos valores observados (em graus) para o *scanner* (a) e para o colorímetro (b)

As Figuras 28(a) e 28(b) mostram que as curvas ajustadas para o *scanner* aparentemente são diferentes das curvas ajustadas para o colorímetro, sendo a principal diferença a presença dos grupos de frutos com maturação fisiológica distintas para o *scanner*, sendo um grupo é composto, aparentemente, por 16 indivíduos de amadurecimento “normal” e o outro grupo composto, aparentemente, por 4 indivíduos de amadurecimento “lento”.

A Figura 29 mostra que as curvas ajustadas aproximam-se bem às observações dos frutos obtidas pelo *scanner*, captando a variabilidade existente dentro do mesmo indivíduo e, portanto, ajustando-se bem aos dois grupos de maturação mencionados anteriormente. Em relação ao colorímetro, observa-se que os modelos individuais ajustaram-se bem às observações dos frutos, captando a variabilidade existente dentro do mesmo indivíduo (Figura 30).

Portanto, o modelo dado na expressão (38) ajustou-se bem aos dados de

Figura 28 – Gráfico das curvas individuais ajustadas a partir do modelo (38) para a tonalidade média (em graus) em função do tempo (em dias) para o *scanner* (a) e para o colorímetro (b), bem como uma escala de cores baseada na tonalidade, fixando $L = 50$ e $C = 50$

tonalidade média da casca para ambas as metodologias, indicando que a utilização de LMEM para a análise de dados de tonalidade média medida ao longo do tempo para dados desbalanceados pode ser uma boa alternativa.

A análise de imagens digitalizadas também pode ser, ainda, estendida para a classificação do mamão “Golden”, que, segundo Fonseca et al. (2007), apresenta variação da cor da casca pequena devido a maior redução dos teores de clorofila comparada ao pequeno aumento no teor de carotenoides. De acordo com os autores, a diferença entre o teor de clorofila e de carotenoides é de 10 vezes para o “Golden” enquanto que para o “Sunrise Solo” é de apenas duas vezes.

Figura 29 – Gráfico dos ajustes individuais (curva rosa) para a tonalidade média (em graus) em função do tempo (em dias), com adição do modelo de efeitos fixos (curva azul) e das observações para a metodologia do *scanner*

Figura 30 – Gráfico dos ajustes individuais (curva rosa) para a tonalidade média (em graus) em função do tempo (em dias), com adição do modelo de efeitos fixos (curva azul) e das observações para a metodologia do colorímetro

6 CONSIDERAÇÕES FINAIS

A metodologia de análise da cor por meio de um *scanner*, bem como por meio de um colorímetro revelaram que os frutos de mamoeiro “Sunrise Solo” apresentam dois possíveis grupos de maturação fisiológica, sendo um de maturação “lenta” e outro de maturação “normal”. É importante notar que quando se utilizou a metodologia de análise da cor por meio de um colorímetro a presença desses grupos não ficou evidente na análise descritiva, no entanto, foi possível verificar a existências desses grupos a partir do cálculo da velocidade de maturação dos frutos.

Também é importante frisar que, independentemente da metodologia de análise da cor, existem frutos que apresentam, ao mesmo tempo, velocidade de maturação e tonalidade média inicial altas, como também existem frutos com velocidade de maturação baixa e tonalidade média inicial alta, portanto, aparentemente, os frutos que apresentam maturação fisiológica “lenta” (1, 3, 6 e 11) não estão relacionados ao ponto de colheita dos mesmos, indicando a provável existência de outros fatores fisiológicos que diretamente ou indiretamente interferem na velocidade de maturação dos frutos.

Quando o fruto foi avaliado em seu estadio inicial de maturação por meio de um colorímetro, a tonalidade média resultante a partir dessa metodologia para esse fruto foi diferente da tonalidade média resultante da metodologia utilizando-se um *scanner*. De forma geral, quando utilizou-se um *scanner* a tonalidade média inicial resultante foi mais amarelada do que a tonalidade média inicial resultante da metodologia utilizando-se um colorímetro.

Já a matriz de variâncias e covariâncias $\mathbf{D}_*$ não estruturada e a matriz de variâncias e covariâncias Σ_i com estrutura CV_H , indicam que a variabilidade entre os indivíduos e num mesmo indivíduo, respectivamente, são altas. Dessa forma, o emprego de técnicas de formulação de modelos lineares de efeitos mistos foram fundamentais para o estudo da tonalidade média de mamão “Sunrise Solo” e indicaram que, independentemente da metodologia de análise da cor, os indivíduos apresentaram comportamentos distintos.

Assim, para obter uma avaliação mais precisa da maturação da casca de toda a superfície do fruto deve-se utilizar o método de análise de imagens digitalizadas, obtidas por meio de um *scanner*, uma vez que foram mais consistentes e eficientes em relação ao método instrumental, por meio de um colorímetro, no cálculo da tonalidade média da casca

dos frutos de mamão “Sunrise Solo” ao longo do tempo. Isso está diretamente relacionado ao número de pontos observados, bem como a posição dos mesmos na casca dos frutos. Na prática, com o *scanner* é possível observar toda a região da casca do fruto, enquanto que com o colorímetro o mesmo não é possível.

REFERÊNCIAS

- AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on automatic control**, New York, v. 19, n. 6, p. 716–723, Dec. 1974.
- ALCARDE, R. **Modelos lineares mistos em dados longitudinais com o uso do pacote ASReml–R**. 2012. 156 p. Tese (Doutorado em Estatística e Experimentação Agrônômica) — Escola Superior de Agricultura “Luiz de Queiroz”, Piracicaba, 2012.
- ALMEIDA, R. F.; MARTINS, M. L. L.; RESENDE, E. D. de; VITORAZI, L.; CARLOS, L. A.; PINTO, L. K. A. Influência da temperatura de refrigeração sobre as características químicas do mamão cv. “golden”. **Ciência e Tecnologia de Alimentos**, Campinas, v. 26, n. 3, p. 577–581, 2006.
- ASMARE, M. H.; ASIRVADAM, V. S.; IZNITA, L. Color space selection for color image enhancement application. In: INTERNATIONAL CONFERENCE ON SIGNAL ACQUISITION AND PROCESSING, 2009, Kuala Lumpur. **Anais Eletrônicos...** Kuala Lumpur: Signal Acquisition and Processing, 2009. p. 208–212. Disponível em: <<http://www.informatik.uni-trier.de/ley/db/conf/icsap/icsap2009.html>>. Acesso em: 27 nov. 2013.
- ASSOCIAÇÃO BRASILEIRA DE HORTICULTURA. **Programa Brasileiro para a Modernização da Horticultura**. São Paulo, 2003. Disponível em: <<http://www.horti-brasil.org.br>>. Acesso em: 14 fev. 2013.
- BALBINO, J. M. S. **Efeitos de hidrotermia, refrigeração e ethephon na qualidade pós-colheita do mamão (*Carica papaya* L.)**. 1997. 104 p. Tese (Doutorado em Fisiologia Vegetal) — Universidade Federal de Viçosa, Viçosa, 1997.
- BALBINO, J. M. S.; COSTA, A. de F. S. Crescimento e desenvolvimento dos frutos do mamoeiro do “grupo solo” e padrão de qualidade. In: MARTINS, D. S.; COSTA, A. F. S. (Ed.). **A cultura do mamoeiro: tecnologias de produção**. Vitória: Incaper, 2003. p. 389–402.
- BATES, D.; CHAMBERS, J.; DALGAARD, P.; FALCON, S.; GENTLEMAN, R.; HORNIK, K.; IACUS, S.; IHAKA, R.; LEISCH, F.; LIGGES, U.; LUMLEY, T.; MAECHLER, M.; MURDOCH, D.; MURRELL, P.; PLUMMER, M.; RIPLEY, B.; SARKAR, D.; LANG, D. T.; TIERNEY, L.; URBANEK, S. **R**: Versão 2.15.2. 2013. Disponível em: <<http://www.r-project.org>>. Acesso em: 27 jul. 2013.
- BLEINROTH, E.W; SIGRIST, J.M.M. II - matéria-prima. In: INSTITUTO DE TECNOLOGIA DE ALIMENTOS, 1995. **Mamão: cultura, matéria-prima, processamento e aspectos econômicos**. Campinas, 1995. p. 179–254.
- BLEINROTH, E. W. Determinação do ponto de colheita. In: GAYET, J. P.; BLEINROTH, E. P.; MATALLO, M.; GARCIA, E. E. C.; GARCIA, A. E.; ARDITO, E.F.G.; BORDIN, M. R. (Ed.). **Mamão para Exportação: procedimentos de colheita e pós-colheita**. Brasília: EMBRAPA–SPI, 1995. p. 10–25.
- BOLKER, B. M.; BROOKS, M. E.; CLARK, C. J.; GEANGE, S. W.; POULSEN, J. R.; STEVENS, M. H. H.; WHITE, J. S. Generalized linear mixed models: a practical guide

for ecology and evolution. **Trends in Ecology and Evolution**, Miami, v. 24, n. 3, p. 127–135, 2008.

BREWER, M. T.; LANG, L.; FUJIMURA, K.; DUJMOVIC, N.; GRAY, S.; KNAAP, E. van der. Development of a controlled vocabulary and software application to analyze fruit shape variation in tomato and other plant species. **Plant Physiology**, Ohio, v. 141, n. 1, p. 15–25, May 2006.

BROSNAN, T.; SUN, D. W. Inspection and grading of agricultural and food products by computer vision – a review. **Computers and Electronics in Agriculture**, Irlanda, v. 36, n. 2, p. 193–213, Nov. 2002.

CASELLA, G.; BERGER, R. L. **Inferência estatística**. 2. ed. Tradução de S. A. Visconde. São Paulo: Cengage Learning, 2011. 588 p.

CHEN, Y. R.; CHAO, K.; KIM, M. S. Machine vision technology for agricultural applications. **Computers and Electronics in Agriculture**, Beltsville, v. 36, p. 173–191, 2002.

CHITARRA, M. I. F.; CHITARRA, A. D. **Pós-colheita de frutos e hortaliças – Fisiologia e Manuseio**. 2. ed. Lavras: Universidade Federal de Lavras, 2005. 785 p.

ColorMetrix. **Join us for Color Control Myths & Mysteries Part III**. 2013. Disponível em: <<http://colormetrix.com/blog/join-us-color-control-myths-mysteries-part-iii/>>. Acesso em: 13 nov. 2013.

COMMISSION INTERNATIONALE DE L'ÉCLAIRAGE. **Recommendations on uniform color spaces**: Color-difference equations, psychometric color terms. Paris, 1978. 1 v.

DARRIGUES, A.; HALL, J.; KNAAP, E. van der; FRANCIS, D. M. Tomato analyzer-color test: a new tool for efficient digital phenotyping. **Journal of the American Society for Horticultural Science**, Wooster, v. 133, n. 4, p. 579–586, Apr. 2008.

DEMIDENKO, E. **Mixed Models**. New Jersey: John Wiley & Sons, 2004. 704 p.

EMBRAPA. **Participação dos principais países na produção mundial de mamão em 2010**. 2010. Disponível em: <http://www.cnpmf.embrapa.br/planilhas/Mamao_Brasil_2010.pdf>. Acesso em: 15 fev. 2013.

FAGUNDES, G. R.; YAMANISHI, O. K. Características físicas e químicas de frutos de mamoeiro do grupo “solo” comercializados em 4 estabelecimentos de Brasília-DF. **Revista Brasileira de Fruticultura**, Jaboticabal, v. 23, n. 3, p. 541–545, dez. 2001.

FIORAVANÇO, J.C.; PAIVA, M.C.; CARVALHO, R.I.N. de.; MANICA, I. Características do mamão formosa comercializado em Porto Alegre de outubro/91 a junho/92. **Ciência Rural**, Santa Maria, v. 24, n. 3, p. 519–522, 1994.

FONSECA, M. J. O.; LEAL, N. R.; CENCI, S. A.; CECON, P. R.; BRESSAN-SMITH, R. E.; BALBINO, J. M. S. Evolução dos pigmentos durante o amadurecimento de mamão “sunrise solo” e “golden”. **Revista Brasileira de Fruticultura**, Jaboticabal, v. 29, n. 3, p. 451–455, jun. 2007.

FOX, J.; WEISBERG, S. **An R Companion to Applied Regression**. 2011.

GILMOUR, R.; THOMPSON, R.; CULLIS, B. R. An efficient algorithm for variance parameter estimation in linear mixed models. **Biometrics**, Arlington, v. 51, n. 4, p. 1440–1450, Dec. 1995.

GILMOUR, A. R.; THOMPSON, R.; CULLIS, B. R. Average information reml: an efficient algorithm for variance parameter estimation in linear mixed models. **Biometrics**, Arlington, v. 51, n. 4, p. 1140–1450, Dec 1995.

HARVILLE, D. Extension of the gauss-markov theorem to include the estimation of random effects. **The Annals of Statistics**, Philadelphia, v. 4, n. 2, p. 384–395, Mar. 1976.

HARVILLE, D. A. Maximum likelihood approaches to variance component estimation and to related problems. **Journal of the American Statistical Association**, Alexandria, v. 72, n. 358, p. 320–338, 1977.

HASLETT, S. J.; PUNTANEN, S. On the equality of the blups under two linear mixed models. **Metrika**, Nova Zelândia, v. 74, p. 381–395, 2011.

HENDERSON, C. R. Estimation of variance and covariance components. **Biometrics**, Arlington, v. 9, n. 2, p. 226–252, June 1953.

HILDEN-MINTON, J. A. **Multilevel diagnostics for mixed and hierarchical linear models**. 1995. 110 p. Tese (Philosophy in Statistics) — University of California, Los Angeles, 1995.

JACOMINO, A. P.; BRON, I. U.; KLUGE, R. A. Avanços em tecnologia pós-colheita de mamão. In: SIMPÓSIO DO PAPAYA BRASILEIRO, 2003, Vitória. **Papaya Brasil: qualidade do mamão para o mercado interno**. Vitória: INCAPER, 2003. p. 279–289.

JOHNSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis**. 6. ed. London: Pearson Prentice Hall, 2007. 773 p.

KAYS, S. J. **Postharvest physiology of perishable plant products**. Athens: Exon Press, 1991. 532 p.

KNIGHT, E. **Improved iterative schemes for REML estimation of variance parameters in linear mixed models**. 2008. 290 p. Tese (Philosophy in Statistics) — School of Agriculture , Food and Wine. The University of Adelaide, Adelaide, 2008.

LAIRD, N. M.; WARE, J. H. Random-effects models for longitudinal data. **Biometrics**, Nova York, v. 38, n. 4, p. 963–974, Dec. 1982.

LI, Q.; WANG, M.; GU, W. Computer vision based system for apple surface defect detection. **Computers and Electronics in Agriculture**, Oxon, v. 36, p. 215–223, Nov. 2002.

LIMA, K. S. C.; LIMA, A. L. S.; FREITAS, L. C.; DELLA-MODESTA, R. C.; GODOY, R. L. O. Efeito de baixas doses de irradiação nos carotenóides majoritários em cenouras prontas para o consumo. **Ciência e Tecnologia de Alimentos**, Campinas, v. 24, n. 2, p. 183–193, maio 2004.

LINDSTROM, M. J.; BATES, D. M. Nonlinear mixed effects models for repeated measures data. **Biometrics**, Nova York, v. 46, n. 3, p. 673–687, Sept. 1990.

LITTELL, L. C. Analysis of unbalanced mixed model data: A case study comparison of anova versus reml/gls. **Journal of Agricultural, Biological, and Environmental Statistics**, v. 7, n. 4, p. 472–490, Dec 2002.

MAGALHÃES, M. N. **Probabilidade e variáveis aleatórias**. 3. ed. São Paulo: Editora da Universidade de São Paulo, 2011. 424 p.

MENDOZA, F.; AGUILERA, J. M. Application of image analysis for classification of ripening bananas. **Journal of Food Science**, Santiago, v. 69, n. 9, p. 471–477, Nov. 2004.

MUNSELL COLOR COMPANY. Baltimore: Munsell Color Company, 1952. 2 v.

NOBRE, J. S.; SINGER, J. M. Residual analysis for linear mixed models. **Biometrical Journal**, Weinheim, v. 49, n. 6, p. 863–875, Jan. 2007.

OLIVEIRA, M. A. B. de; VIANNI, R.; SOUZA, G. de; ARAÚJO, T. M. R. Caracterização do estágio de maturação do papaia “golden” em função da cor. **Revista Brasileira de Fruticultura**, Jaboticabal, v. 24, n. 2, p. 559–561, ago. 2002.

OLIVEIRA, T. P.; ZOCCHI, S. S. Calibração de scanners de mesa para análise de cor de casca de mamão (*Carica papaya* L.) “golden”. In: SIMPÓSIO INTERNACIONAL DE INICIAÇÃO CIENTÍFICA DA UNIVERSIDADE DE SÃO PAULO, 18., 2010, Piracicaba. **Anais Eletrônicos...** Piracicaba: Universidade de São Paulo, 2010. Disponível em: <<https://uspdigital.usp.br/siicusp/>>. Acesso em: 13 jul. 2013.

OLIVEIRA, T. P.; ZOCCHI, S. S. Curvas de maturação de mamão cultivar “sunrise solo”. In: SIMPÓSIO INTERNACIONAL DE INICIAÇÃO CIENTÍFICA DA UNIVERSIDADE DE SÃO PAULO, 19., 2011, Piracicaba. **Anais Eletrônicos...** Piracicaba: Universidade de São Paulo, 2011. Disponível em: <<https://uspdigital.usp.br/siicusp/>>. Acesso em: 20 ago. 2013.

OLIVEIRA, T. P.; ZOCCHI, S. S. Análise de dados circulares com aplicação em tonalidade da cor de casca de mamão “sunrise solo”. In: REGIÃO BRASILEIRA DA SOCIEDADE INTERNACIONAL DE BIOMETRIA, 58.; SIMPÓSIO DE ESTATÍSTICA APLICADA À EXPERIMENTAÇÃO AGRONÔMICA, 15., 2013, Campina Grande. **Anais...** Campina Grande, 2013. p. 5.

O’SULLIVAN, M. G.; BYRNE, D. V.; MARTENS, H.; GIDSKEHAUG, L. H.; ANDERSEN, H. J.; MARTENS, M. Evaluation of pork color: prediction of visual sensory quality of meat from instrumental and computer vision methods of color analysis. **Meat Science**, Copenhagen, v. 65, p. 909–918, Nov. 2003.

PASCALE, D. **A review of RGB color space**: ...from xyY to RGB. Montreal: The BabelColor Company, 2003. 35 p.

PAULUS, I.; SCHREVEENS, E. Shape characterization of new apple cultivars by fourier expansion of digitalized images. **Journal of Agricultural Engineering Research**, Hverlee, v. 72, p. 113–118, Feb. 1999.

PICHA, D. Horticultural crop quality characteristics important in international trade. **Acta Horticulturae**, Bangkok, v. 712, n. 1, p. 423–426, June 2006. Disponível em: <http://www.actahort.org/books/712/712_49.htm>. Acesso em: 10 fev. 2013.

PINHEIRO, J. C.; BATES, D. M. Approximations to the log-likelihood function in the nonlinear mixed-effects model. **Journal of Computational and Graphical Statistics**, Alexandria, v. 4, n. 1, p. 12–35, 1995.

PINHEIRO, J. C.; BATES, D. M.; DEBROY, S.; SARKAR, D.; EISPACK; R-CORE. **nlme**: Versão 3.1-108. 2013.

PINTO, L. K. A.; MARTINS, M. L. L.; RESENDE, E. D. de; THIÈBAUT, J. T. L. Atividade da pectina metilesterase e da β -galactosidase durante o amadurecimento do mamão cv. “golden”. **Revista Brasileira de Fruticultura**, São Paulo, v. 33, n. 3, p. 713–722, 2011.

POLDERDIK, J. J.; TIJSKENS, L.M.M.; ROBBERS, J. E.; VALK, H. C.P.M. van der. Predictive model of keeping quality of tomatoes. **Postharvest Biology and Technology**, Amsterdam, v. 2, p. 179–185, 1993.

R Core Team and contributors worldwide. **The R stats package**: Version 3.0.1. 2010.

RENCHE, A. C.; SCHAALJE, G. B. **Linear Models in Statistics**. 2. ed. New Jersey: John Wiley & Sons, 2008. 672 p.

RODRÍGUEZ, G.; STRECKER, J.; BREWER M.; GONZALO, M. J.; ANDERSON, C.; LANG, L.; SULLIVAN, D.; WAGNER, E.; STRECKER, B.; DRUSHAL, R.; DUJMOVIC, N.; FUJIMURO, K.; JACK, A.; NJANJI, I.; THOMAS, J.; GRAY, S.; KNAAP, E. van der. **Tomato analyzer user manual version 3**. Wooster, 2010.

SCHWARZ, G. Estimating the dimension of a model. **The Annals of Statistics**, Jerusalem, v. 6, n. 2, p. 461–464, 1978.

SHAHIN, M. A.; SYMONS, S. J. A machine vision system for grading lentils. **Canadian Biosystems Engineering**, Winnipeg, v. 43, n. 7, p. 7–14, 2001.

SPIELKE, J.; PIEPHO, H. P.; HU, X. Analysis of unbalanced data by mixed linear models using the mixed procedure of the sas system. **Journal of Agronomy and Crop Science**, Berlin, n. 191, p. 47–54, 2005.

UNIVERSIDADE FEDERAL DE SANTA CATARINA (UFSC). 2013. Disponível em: <<http://www.arq.ufsc.br/labcon/arq5656/>>. Acesso em: 05 out. 2013.

VERBEKE, G.; MOLENBERGHS, G. **Linear mixed models for longitudinal data**. Nova York: Springer Verlag, 2009. 568 p.

VERBYLA, A.P.; CULLIS, B.R.; KENWARD, M.G.; WELHAN, S.J. The analysis of designed experiments and longitudinal data by using smoothing splines. **Journal of the Royal Statistical Society**, London, v. 48, n. 3, p. 269–311, 1999.

VIEIRA, G.; VIÉGAS, P. R. A.; NEVES, J. C. L.; AGNES, E. L.; OLIVEIRA, F. A. M. B. Influência do cultivar e do estágio de maturação em algumas características de frutos de mamão durante a pós-colheita. In: CONGRESSO BRASILEIRO DE FRUTICULTURA, 15., 1998, Poços de Caldas. **Anais...** Poços de Caldas: Universidade Federal de Lavras, 1998. p. 137–142.

WALSH, B. **Mixed Models**. 2010. Disponível em: <<http://nitro.biosci.arizona.edu/courses/EEB596/handouts/mixed.pdf>>. Acesso em: 06 jul. 2013.

YAM, K. L.; PAPADAKIS, S. P. A simple digital imaging method for measuring and analyzing color of food surfaces. **Journal of Food Engineering**, Oxon, v. 61, n. 1, p. 137–142, Jan. 2003.

APÊNDICES

APÊNDICE A - Prova de que $E(\boldsymbol{\eta}) = \mathbf{0}$ e de que $cov(\boldsymbol{\eta}) = \mathbf{V}_*(\boldsymbol{\gamma})$

i) Esperança

Proposição 1 (MAGALHÃES, 2011, p. 226). Sejam $X_1, X_2, \dots, X_M$ variáveis aleatórias independentes cujo valor esperado existe. Então, se a esperança da soma dessas variáveis existir, tem-se que

$$E(X_1 + X_2 + \dots + X_M) = E(X_1) + E(X_2) + \dots + E(X_M).$$

Proposição 2 (MAGALHÃES, 2011, p. 227). Sejam $X_1, X_2, \dots, X_M$ variáveis aleatórias cujo valor esperado é finito. Então, a esperança do produto dessas variáveis é finita e é igual ao produto das esperanças, ou seja,

$$E\left(\prod_{i=1}^M X_i\right) = E(X_1) E(X_2) \dots E(X_M).$$

Agora, seja $\boldsymbol{\eta}_i = \boldsymbol{\varepsilon}_i + \mathbf{Z}_i \mathbf{b}_i$, com $i = 1, 2, \dots, M$, tem-se que a $E[\boldsymbol{\eta}]$ é dada por:

$$\begin{aligned} E[\boldsymbol{\eta}] &= E\left[\begin{pmatrix} \boldsymbol{\eta}_1 \\ \boldsymbol{\eta}_2 \\ \vdots \\ \boldsymbol{\eta}_M \end{pmatrix}\right] = E\left[\begin{pmatrix} \boldsymbol{\varepsilon}_1 + \mathbf{Z}_1 \mathbf{b}_1 \\ \boldsymbol{\varepsilon}_2 + \mathbf{Z}_2 \mathbf{b}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_M + \mathbf{Z}_M \mathbf{b}_M \end{pmatrix}\right] = \begin{pmatrix} E[\boldsymbol{\varepsilon}_1 + \mathbf{Z}_1 \mathbf{b}_1] \\ E[\boldsymbol{\varepsilon}_2 + \mathbf{Z}_2 \mathbf{b}_2] \\ \vdots \\ E[\boldsymbol{\varepsilon}_M + \mathbf{Z}_M \mathbf{b}_M] \end{pmatrix} \\ &= \begin{pmatrix} E[\boldsymbol{\varepsilon}_1] + E[\mathbf{Z}_1 \mathbf{b}_1] \\ E[\boldsymbol{\varepsilon}_2] + E[\mathbf{Z}_2 \mathbf{b}_2] \\ \vdots \\ E[\boldsymbol{\varepsilon}_M] + E[\mathbf{Z}_M \mathbf{b}_M] \end{pmatrix} = \begin{pmatrix} E[\boldsymbol{\varepsilon}_1] + \mathbf{Z}_1 E[\mathbf{b}_1] \\ E[\boldsymbol{\varepsilon}_2] + \mathbf{Z}_2 E[\mathbf{b}_2] \\ \vdots \\ E[\boldsymbol{\varepsilon}_M] + \mathbf{Z}_M E[\mathbf{b}_M] \end{pmatrix} \end{aligned}$$

como assumiu-se que $\boldsymbol{\varepsilon}_i \stackrel{ind}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{I}\sigma^2)$, $\mathbf{b}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{D}_*)$ e que $\boldsymbol{\varepsilon}_i$ é independente de $\mathbf{b}_i$, tem-se

que:

$$E[\boldsymbol{\eta}] = \begin{pmatrix} \mathbf{0} + \mathbf{Z}_1 \mathbf{0} \\ \mathbf{0} + \mathbf{Z}_2 \mathbf{0} \\ \vdots \\ \mathbf{0} + \mathbf{Z}_M \mathbf{0} \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix};$$

ii) Covariância

$$\begin{aligned} cov[\boldsymbol{\eta}] &= E \left[\left(\begin{pmatrix} \boldsymbol{\eta}_1 \\ \boldsymbol{\eta}_2 \\ \vdots \\ \boldsymbol{\eta}_M \end{pmatrix} - E \begin{pmatrix} \boldsymbol{\eta}_1 \\ \boldsymbol{\eta}_2 \\ \vdots \\ \boldsymbol{\eta}_M \end{pmatrix} \right) \left(\begin{pmatrix} \boldsymbol{\eta}_1 \\ \boldsymbol{\eta}_2 \\ \vdots \\ \boldsymbol{\eta}_M \end{pmatrix} - E \begin{pmatrix} \boldsymbol{\eta}_1 \\ \boldsymbol{\eta}_2 \\ \vdots \\ \boldsymbol{\eta}_M \end{pmatrix} \right)^T \right] \\ &= E \left[\begin{pmatrix} \boldsymbol{\eta}_1 - E(\boldsymbol{\eta}_1) \\ \boldsymbol{\eta}_2 - E(\boldsymbol{\eta}_2) \\ \vdots \\ \boldsymbol{\eta}_M - E(\boldsymbol{\eta}_M) \end{pmatrix} \begin{pmatrix} \boldsymbol{\eta}_1 - E(\boldsymbol{\eta}_1) \\ \boldsymbol{\eta}_2 - E(\boldsymbol{\eta}_2) \\ \vdots \\ \boldsymbol{\eta}_M - E(\boldsymbol{\eta}_M) \end{pmatrix}^T \right] \end{aligned}$$

Teorema 1 (RENCER; SCHAALJE, 2008, p. 15). Seja $\mathbf{A}$ uma matriz $n \times p$ qualquer.

Então $\mathbf{A}^T \mathbf{A}$ e $\mathbf{A} \mathbf{A}^T$ têm as seguintes propriedades:

- 1) $\mathbf{A}^T \mathbf{A}$ é $p \times p$ e é obtida como produto das *colunas* de $\mathbf{A}$;
- 2) $\mathbf{A} \mathbf{A}^T$ é $n \times n$ e é obtida como produto das *linhas* de $\mathbf{A}$;
- 3) Ambas as matrizes $\mathbf{A}^T \mathbf{A}$ e $\mathbf{A} \mathbf{A}^T$ são simétricas;
- 4) Se $\mathbf{A}^T \mathbf{A} = \boldsymbol{\Phi}$ então $\mathbf{A} = \boldsymbol{\Phi}$.

Proposição 3 (RENCER; SCHAALJE, 2008, p. 11). Seja $\mathbf{b}$ um vetor $n \times 1$ então a multiplicação envolvendo $\mathbf{b} \mathbf{b}^T$ é uma matriz quadrada e simétrica de ordem $n \times n$.

Considere ainda que

$$var(\boldsymbol{\eta}_M) = E \left((\boldsymbol{\eta}_M - E(\boldsymbol{\eta}_M)) (\boldsymbol{\eta}_M - E(\boldsymbol{\eta}_M))^T \right)$$

e que

$$\text{cov}(\boldsymbol{\eta}_M, \boldsymbol{\eta}_{M^*}) = E \left((\boldsymbol{\eta}_M - E(\boldsymbol{\eta}_M)) (\boldsymbol{\eta}_{M^*} - E(\boldsymbol{\eta}_{M^*}))^T \right),$$

para qualquer $M \neq M^*$.

Dessa forma, a partir do **Teorema 1** e da **Proposição 3** tem-se que:

$$\text{cov}[\boldsymbol{\eta}] = \begin{bmatrix} \text{var}(\boldsymbol{\eta}_1) & \text{cov}(\boldsymbol{\eta}_1, \boldsymbol{\eta}_2) & \dots & \text{cov}(\boldsymbol{\eta}_1, \boldsymbol{\eta}_M) \\ \text{cov}(\boldsymbol{\eta}_1, \boldsymbol{\eta}_2) & \text{var}(\boldsymbol{\eta}_2) & \dots & \text{cov}(\boldsymbol{\eta}_2, \boldsymbol{\eta}_M) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(\boldsymbol{\eta}_1, \boldsymbol{\eta}_M) & \text{cov}(\boldsymbol{\eta}_2, \boldsymbol{\eta}_M) & \dots & \text{var}(\boldsymbol{\eta}_M) \end{bmatrix}$$

Proposição 4 (MAGALHÃES, 2011, p. 252). Sejam $X_1, X_2, \dots, X_M$ variáveis aleatórias num mesmo espaço de probabilidade. Então, sempre que as esperanças envolvidas estejam bem definidas, tem-se que:

$$\text{var} \left(\sum_{i=1}^n X_i \right) = \sum_{i=1}^n \text{var}(X_i) + 2 \sum_i \sum_j \text{cov}(X_i, X_j) \text{ com } 1 \leq i < j \leq n.$$

No caso das variáveis aleatórias serem independentes, além das condições mencionadas, então:

$$\text{var} \left(\sum_{i=1}^n X_i \right) = \sum_{i=1}^n \text{var}(X_i).$$

Teorema 2 (RENCER; SCHAALJE, 2008, p. 82). Seja $\mathbf{z} = \mathbf{A}\mathbf{y}$ e $\mathbf{w} = \mathbf{B}\mathbf{y}$, em que $\mathbf{A}$ é uma matriz de constantes de ordem $k \times p$, $\mathbf{B}$ é uma matriz de constantes de ordem $m \times p$ e $\mathbf{y}$ é um vetor aleatório de ordem p com matriz de variâncias e covariâncias $\boldsymbol{\Sigma}$. Então

$$1) \text{ cov}(\mathbf{z}) = \text{cov}(\mathbf{A}\mathbf{y}) = \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T;$$

$$2) \text{ cov}(\mathbf{z}, \mathbf{w}) = \text{cov}(\mathbf{A}\mathbf{y}, \mathbf{B}\mathbf{y}) = \mathbf{A}\boldsymbol{\Sigma}\mathbf{B}^T.$$

A partir da **Proposição 4**, do **Teorema 2** e assumindo independência entre

os $\boldsymbol{\eta}_i$, tem-se que:

$$\begin{aligned}
 cov[\boldsymbol{\eta}] &= \begin{bmatrix} var(\boldsymbol{\eta}_1) & \boldsymbol{\Phi} & \dots & \boldsymbol{\Phi} \\ \boldsymbol{\Phi} & var(\boldsymbol{\eta}_2) & \dots & \boldsymbol{\Phi} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Phi} & \boldsymbol{\Phi} & \dots & var(\boldsymbol{\eta}_M) \end{bmatrix} \\
 &= \begin{bmatrix} var(\mathbf{Z}_1 \mathbf{b}_1 + \boldsymbol{\varepsilon}_1) & \boldsymbol{\Phi} & \dots & \boldsymbol{\Phi} \\ \boldsymbol{\Phi} & var(\mathbf{Z}_2 \mathbf{b}_2 + \boldsymbol{\varepsilon}_2) & \dots & \boldsymbol{\Phi} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Phi} & \boldsymbol{\Phi} & \dots & var(\mathbf{Z}_M \mathbf{b}_M + \boldsymbol{\varepsilon}_M) \end{bmatrix} \\
 &= \begin{bmatrix} var(\mathbf{Z}_1 \mathbf{b}_1) + var(\boldsymbol{\varepsilon}_1) & \boldsymbol{\Phi} & \dots & \boldsymbol{\Phi} \\ \boldsymbol{\Phi} & var(\mathbf{Z}_2 \mathbf{b}_2) + var(\boldsymbol{\varepsilon}_2) & \dots & \boldsymbol{\Phi} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Phi} & \boldsymbol{\Phi} & \dots & var(\mathbf{Z}_M \mathbf{b}_M) + var(\boldsymbol{\varepsilon}_M) \end{bmatrix} \\
 &= \begin{bmatrix} \mathbf{Z}_1 var(\mathbf{b}_1) \mathbf{Z}_1^T + var(\boldsymbol{\varepsilon}_1) & \boldsymbol{\Phi} & \dots & \boldsymbol{\Phi} \\ \boldsymbol{\Phi} & \mathbf{Z}_2 var(\mathbf{b}_2) \mathbf{Z}_2^T + var(\boldsymbol{\varepsilon}_2) & \dots & \boldsymbol{\Phi} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Phi} & \boldsymbol{\Phi} & \dots & \mathbf{Z}_M var(\mathbf{b}_M) \mathbf{Z}_M^T + var(\boldsymbol{\varepsilon}_M) \end{bmatrix}
 \end{aligned}$$

Como se assumiu que $\boldsymbol{\varepsilon}_i \sim \mathcal{N}(\mathbf{0}, \sigma^2 \boldsymbol{\Lambda}_i)$, $\mathbf{b}_i \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{D})$ e que $\boldsymbol{\varepsilon}_i$ é independente de $\mathbf{b}_i$, tem-se que:

$$cov[\boldsymbol{\eta}] = \begin{bmatrix} \mathbf{Z}_1 (\sigma^2 \mathbf{D}) \mathbf{Z}_1^T + \sigma^2 \boldsymbol{\Lambda}_1 & \boldsymbol{\Phi} & \dots & \boldsymbol{\Phi} \\ \boldsymbol{\Phi} & \mathbf{Z}_2 (\sigma^2 \mathbf{D}) \mathbf{Z}_2^T + \sigma^2 \boldsymbol{\Lambda}_2 & \dots & \boldsymbol{\Phi} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Phi} & \boldsymbol{\Phi} & \dots & \mathbf{Z}_M (\sigma^2 \mathbf{D}) \mathbf{Z}_M^T + \sigma^2 \boldsymbol{\Lambda}_M \end{bmatrix}$$

Portanto, tem-se que:

$$cov[\boldsymbol{\eta}] = \sigma^2 \begin{bmatrix} \mathbf{Z}_1 \mathbf{D} \mathbf{Z}_1^T + \boldsymbol{\Lambda}_1 & \boldsymbol{\Phi} & \dots & \boldsymbol{\Phi} \\ \boldsymbol{\Phi} & \mathbf{Z}_2 \mathbf{D} \mathbf{Z}_2^T + \boldsymbol{\Lambda}_2 & \dots & \boldsymbol{\Phi} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Phi} & \boldsymbol{\Phi} & \dots & \mathbf{Z}_M \mathbf{D} \mathbf{Z}_M^T + \boldsymbol{\Lambda}_M \end{bmatrix} = \mathbf{V}_*(\boldsymbol{\gamma}).$$

APÊNDICE B - Logaritmo da Função da verossimilhança

A distribuição normal multivariada, segundo Johnson e Wichern (2007), é obtida substituindo a distância univariada $(x - \mu)(\sigma^2)^{-1}(x - \mu)$ pela distância generalizada multivariada $(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})$ na função densidade de probabilidade. Quando essa substituição ocorre, a constante $(2\pi)^{-\frac{1}{2}} (\sigma^2)^{-1}$ também deve ser substituída por uma constante mais geral, dada por $(2\pi)^{-\frac{M}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}}$, fazendo com que o volume abaixo da superfície da função densidade multivariada seja único para qualquer M. Consequentemente, uma densidade normal multivariada para um vetor aleatório $\mathbf{Y} = [\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_M]$ ($i = 1, 2, \dots, M$) tem a forma

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{-\frac{M}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\}$$

Como se assumiu que

$$\mathbf{Y}_i \stackrel{iid}{\sim} \mathcal{N}(\mathbf{X}_i \boldsymbol{\beta}, \mathbf{V}_{*i}(\gamma)), \quad i = 1, 2, \dots, M,$$

pode-se escrever a função de densidade conjunta de $\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_M$ como um produto das densidades marginais. Então, a função de $\boldsymbol{\theta}$ é definida por

$$\mathbf{L}(\boldsymbol{\theta}|\mathbf{y}, \mathbf{X}) = \prod_{i=1}^M \frac{1}{\sqrt{2\pi |\mathbf{V}_{*i}(\gamma)|}} \exp \left\{ -\frac{1}{2} [(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})^T \mathbf{V}_{*i}^{-1}(\gamma) (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})] \right\}.$$

Ao aplicarmos o logaritmo tem-se que:

$$\begin{aligned} \mathbf{l}(\boldsymbol{\theta}|\mathbf{y}, \mathbf{X}) &= -\frac{M}{2} \ln(2\pi) + \sum_{i=1}^M \ln \left[\frac{1}{\sqrt{|\mathbf{V}_{*i}(\gamma)|}} \right] - \frac{1}{2} \sum_{i=1}^M [(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})^T \mathbf{V}_{*i}^{-1}(\gamma) (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})] \\ &\propto -\frac{1}{2} \left\{ \sum_{i=1}^M [\ln |\mathbf{V}_{*i}(\gamma)| + (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})^T \mathbf{V}_{*i}^{-1}(\gamma) (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})] \right\}. \end{aligned}$$

Como $\mathbf{V}_{*i}(\gamma) = \sigma^2 (\mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T + \boldsymbol{\Lambda}_i)$ e $\mathbf{W}_i(\gamma) = (\mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T + \boldsymbol{\Lambda}_i)^{-1}$ tem-se que:

$$\begin{aligned} \mathbf{l}(\boldsymbol{\theta}|\mathbf{y}, \mathbf{X}) &= -\frac{1}{2} \left\{ \sum_{i=1}^M \left[\ln |\sigma^2 \mathbf{W}_i^{-1}(\gamma)| + (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})^T \frac{1}{\sigma^2} \mathbf{W}_i(\gamma) (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}) \right] \right\} \\ &= -\frac{1}{2} \left\{ M \ln \sigma^2 + \sum_{i=1}^M \left[\ln |\mathbf{W}_i^{-1}(\gamma)| + \frac{1}{\sigma^2} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})^T \mathbf{W}_i(\gamma) (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}) \right] \right\} \end{aligned}$$

Considerando $\mathbf{e}_i = \mathbf{e}_i(\boldsymbol{\beta}) = \mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}$, tem-se que:

$$l(\boldsymbol{\theta}|\mathbf{y}, \mathbf{X}) = -\frac{1}{2} \left\{ M \ln \sigma^2 + \sum_{i=1}^M \left[\ln |\mathbf{W}_i^{-1}(\gamma)| + \frac{1}{\sigma^2} \mathbf{e}_i^T \mathbf{W}_i(\gamma) \mathbf{e}_i \right] \right\}.$$

APÊNDICE C - Resíduo marginal

Seja $\boldsymbol{\eta} = \mathbf{Z}\mathbf{b} + \boldsymbol{\varepsilon}$, pode-se reescrever o modelo (3) como $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\eta}$. Assim, o erro marginal predito pode ser escrito como:

$$\hat{\boldsymbol{\eta}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} \quad (39)$$

Segundo Laird e Ware (1982), a estimativa para o vetor de parâmetros $\hat{\boldsymbol{\beta}}$ é dada por

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{V}_*^{-1}(\gamma) \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}_*^{-1}(\gamma) \mathbf{y}. \quad (40)$$

Assim, substituindo 40 em 39 tem-se que:

$$\hat{\boldsymbol{\eta}} = \mathbf{y} - \mathbf{X} (\mathbf{X}^T \mathbf{V}_*^{-1}(\gamma) \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}_*^{-1}(\gamma) \mathbf{y}$$

Como $\mathbf{V}_*(\gamma) = \sigma^2 \mathbf{V}(\gamma)$ então

$$\hat{\boldsymbol{\eta}} = \mathbf{y} - \mathbf{X} (\mathbf{X}^T \mathbf{V}^{-1}(\gamma) \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1}(\gamma) \mathbf{y} \quad (41)$$

Proposição 5 (RENCHE; SCHAALJE, 2008, p. 21). Seja $\mathbf{A}$ uma matriz não-singular, então $\mathbf{A}$ tem inversa única denotada por $\mathbf{A}^{-1}$, com a seguinte propriedade:

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$$

em que $\mathbf{I}$ é uma matriz identidade.

Como $\mathbf{V}(\gamma)$ é uma matriz não-singular, a partir da **Proposição 5**, têm-se que $\mathbf{V}(\gamma) \mathbf{V}^{-1}(\gamma) = \mathbf{I}$. Assim, o erro marginal predito (41) pode ser reescrito como:

$$\begin{aligned} \hat{\boldsymbol{\eta}} &= \mathbf{V}(\gamma) \mathbf{V}^{-1}(\gamma) \mathbf{y} - \mathbf{V}(\gamma) \mathbf{V}^{-1}(\gamma) \mathbf{X} (\mathbf{X}^T \mathbf{V}^{-1}(\gamma) \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1}(\gamma) \mathbf{y} \\ &= \mathbf{V}(\gamma) \left[\mathbf{V}^{-1}(\gamma) - \mathbf{V}^{-1}(\gamma) \mathbf{X} (\mathbf{X}^T \mathbf{V}^{-1}(\gamma) \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1}(\gamma) \right] \mathbf{y} \\ &= \mathbf{V}(\gamma) \left\{ \mathbf{V}^{-1}(\gamma) \left[\mathbf{I} - \mathbf{X} (\mathbf{X}^T \mathbf{V}^{-1}(\gamma) \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1}(\gamma) \right] \right\} \mathbf{y} \end{aligned}$$

Seja $\mathbf{T} = \mathbf{X} (\mathbf{X}^T \mathbf{V}^{-1}(\gamma) \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1}(\gamma)$ tem-se que

$$\hat{\boldsymbol{\eta}} = \mathbf{V}(\gamma) \mathbf{V}^{-1}(\gamma) [\mathbf{I} - \mathbf{T}] \mathbf{y},$$

Fazendo $\mathbf{Q} = \mathbf{V}^{-1}(\gamma) [\mathbf{I} - \mathbf{T}]$ tem-se que:

$$\hat{\boldsymbol{\eta}} = \mathbf{V}(\gamma) \mathbf{Q} \mathbf{y},$$

APÊNDICE D - Demonstração de que $E(\hat{\boldsymbol{\beta}}_i) = \boldsymbol{\beta}_i$ e que a $var(\hat{\boldsymbol{\beta}}_i) = \sigma^2 (\mathbf{X}_i^T \mathbf{X}_i)^{-1}$

Seja

$$y_i = \beta_{0i} + \beta_{1i}x_i + \beta_{2i}x_i^2 + \dots + \beta_{pi}x_i^p + \varepsilon_i,$$

um modelo de regressão linear para o i -ésimo indivíduo, com $i = 1, 2, \dots, M$ e $j = 1, 2, \dots, n_i$ o número de medidas repetidas observadas no indivíduo i . As pressuposições para ε_i e para y_i são:

- i) $E(\varepsilon_i) = 0$ para $i = 1, 2, \dots, M$, ou, equivalentemente, $E(y_i) = \beta_{0i} + \beta_{1i}x_i + \beta_{2i}x_i^2 + \dots + \beta_{pi}x_i^p$;
- ii) $var(\varepsilon_i) = \sigma_i^2$ para $i = 1, 2, \dots, M$, ou, equivalentemente, $var(y_i) = \sigma_i^2$.
- iii) $cov(\varepsilon_i, \varepsilon_{i^*}) = 0$ para todo $i \neq i^*$, com $i = 1, 2, \dots, M$, ou, equivalentemente, $cov(y_i, y_{i^*}) = 0$

Matricialmente, o modelo (42) pode ser escrito na forma

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i. \quad (42)$$

Assim, assumindo que $\boldsymbol{\varepsilon}_i \stackrel{iid}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{I}\sigma_i^2)$ tem-se que

$$E(\mathbf{y}_i) = \mathbf{X}_i \boldsymbol{\beta}_i + E(\boldsymbol{\varepsilon}_i) = \mathbf{X}_i \boldsymbol{\beta}_i$$

e que

$$var(\mathbf{y}_i) = var(\boldsymbol{\varepsilon}_i) = \mathbf{I}\sigma_i^2.$$

De forma equivalente, tem-se que $\mathbf{y}_i \stackrel{iid}{\sim} \mathcal{N}(\mathbf{X}_i \boldsymbol{\beta}_i, \mathbf{I}\sigma_i^2)$.

Para estimar os parâmetros de $\boldsymbol{\beta}_i$ pode-se utilizar o método de mínimos quadrados ordinários, cujas estimativas minimizam a soma de quadrados dos desvios das n_i observações do indivíduo i em relação aos valores preditos pelo modelo.

Derivando-se $\boldsymbol{\varepsilon}_i^T \boldsymbol{\varepsilon}_i$ em relação a $\boldsymbol{\beta}_i$ e igualando-o ao vetor de zeros $\mathbf{0}$, cuja solução, apresentada no item 4, é dada por

$$\hat{\boldsymbol{\beta}}_i = (\mathbf{X}_i^T \mathbf{X}_i)^{-1} \mathbf{X}_i^T \mathbf{y}_i. \quad (43)$$

Teorema 3 (RENCHE; SCHAALJE, 2008, p. 81). Suponha que $\mathbf{y}$ é um vetor aleatório, $\mathbf{X}$ é uma matriz de variáveis aleatórias, $\mathbf{a}$ e $\mathbf{b}$ são vetores de constantes e $\mathbf{A}$ e $\mathbf{B}$ são matrizes de constantes. Então, supondo que as matrizes e os vetores são conformes em cada produto, têm-se os seguintes valores esperados:

- i) $E(\mathbf{A}\mathbf{y}) = \mathbf{A}E(\mathbf{y})$;
- ii) $E(\mathbf{a}^T \mathbf{X} \mathbf{b}) = \mathbf{a}^T E(\mathbf{X}) \mathbf{b}$;
- iii) $E(\mathbf{A}^T \mathbf{X} \mathbf{B}) = \mathbf{A}^T E(\mathbf{X}) \mathbf{B}$.

A esperança do vetor de parâmetros $\boldsymbol{\beta}_i$, considerando o **Teorema 3**, é

$$\begin{aligned} E(\hat{\boldsymbol{\beta}}_i) &= E\left((\mathbf{X}_i^T \mathbf{X}_i)^{-1} \mathbf{X}_i^T \mathbf{y}_i\right) \\ &= (\mathbf{X}_i^T \mathbf{X}_i)^{-1} \mathbf{X}_i^T E(\mathbf{y}_i). \end{aligned}$$

Como $E(\mathbf{y}_i) = \mathbf{X}_i \boldsymbol{\beta}_i$, tem-se que

$$E(\hat{\boldsymbol{\beta}}_i) = (\mathbf{X}_i^T \mathbf{X}_i)^{-1} \mathbf{X}_i^T \mathbf{X}_i \boldsymbol{\beta}_i.$$

Proposição 6 (RENCHE; SCHAALJE, 2008, p. 21). Uma matriz quadrada de posto completo é dita não-singular. Uma matriz $\mathbf{A}$ não-singular tem inversa única, denominada $\mathbf{A}^{-1}$, com a seguinte propriedade

$$\mathbf{A} \mathbf{A}^{-1} = \mathbf{A}^{-1} \mathbf{A} = \mathbf{I}. \quad (44)$$

A partir da **Proposição 6**, tem-se que

$$E(\hat{\boldsymbol{\beta}}_i) = \mathbf{I} \boldsymbol{\beta}_i = \boldsymbol{\beta}_i.$$

Já a variância do vetor de parâmetros $\boldsymbol{\beta}_i$ pode ser escrita como

$$var(\hat{\boldsymbol{\beta}}_i) = var\left((\mathbf{X}_i^T \mathbf{X}_i)^{-1} \mathbf{X}_i^T \mathbf{y}_i\right).$$

A partir do **Teorema 2**, tem-se que

$$\text{var}(\hat{\beta}_i) = (\mathbf{X}_i^T \mathbf{X}_i)^{-1} \mathbf{X}_i^T \text{var}(\mathbf{y}_i) \mathbf{X}_i (\mathbf{X}_i^T \mathbf{X}_i)^{-1}.$$

Como $\text{var}(\mathbf{y}_i) = \mathbf{I}\sigma_i^2$ tem-se que

$$\begin{aligned} \text{var}(\hat{\beta}_i) &= (\mathbf{X}_i^T \mathbf{X}_i)^{-1} \mathbf{X}_i^T \mathbf{I} \sigma_i^2 \mathbf{X}_i (\mathbf{X}_i^T \mathbf{X}_i)^{-1} \\ &= (\mathbf{X}_i^T \mathbf{X}_i)^{-1} \mathbf{X}_i^T \mathbf{X}_i (\mathbf{X}_i^T \mathbf{X}_i)^{-1} \sigma_i^2 \\ &= (\mathbf{X}_i^T \mathbf{X}_i)^{-1} \sigma_i^2. \end{aligned}$$

APÊNDICE E - Sintaxe do software R para a análise descritiva para o *scanner* e para o colorímetro

```
dados=read.csv2("cores.csv",header=TRUE,sep = " ")
=====
## Pacote utilizado
=====
require(car)
=====
## Gráfico de perfis para o scanner e para o colorímetro.
=====
xyplot(Tonalidade_Scanner+Tonalidade_Colorimetro ~ Data |Fruto,dados,type=c("l"),
       xlab="Tempo após o início do ensaio (dias)",ylab="Tonalidade média do
       fruto", col=c("blue","red",2), auto.key = (list(columns=2,alpha=1,
       col=c("blue","red",2))), xlim=c(1:20,2),)
=====
## Gráfico de perfis para o scanner
=====
xyplot(Tonalidade_Scanner ~ dados$Data, groups=Fruto, dados, type=c("l"),
       xlab="Tempo após o início do ensaio (dias)",
       ylab="Tonalidade média do fruto", xlim=c(0,20))
=====
## Gráfico de perfis para o colorímetro
=====
```



```
xyplot(Tonalidade_Colorímetro ~ dados$Data, groups=Fruto, dados, type=c("l"),
       xlab="Tempo após o início do ensaio (dias)",
       ylab="Tonalidade média do fruto", xlim=c(0,20))
```

```
=====  
## Medidas de posição  
=====
```

```
#### Scanner  
-----
```

```
with(dados,tapply(S_H, Data, mean)) # Média da tonalidade média por dia  
with(dados,tapply(S_H, Data, median)) # Mediana da tonalidade média por dia  
-----
```

```
#### Colorímetro  
-----
```

```
with(dados,tapply(C_H, Data, mean)) # Média da tonalidade média por dia  
with(dados,tapply(C_H, Data, median)) # Mediana da tonalidade média por dia  
=====
```

```
## Medidas de dispersão  
=====
```

```
#### Scanner  
-----
```

```
with(dados,tapply(S_H, Data, var)) # Variância da tonalidade média por dia  
with(dados,tapply(S_H, Data, sd)) # Desvio padrão da tonalidade média por dia  
-----
```

```
#### Colorímetro  
-----
```

```
with(dados,tapply(C_H, Data, var)) # Variância da tonalidade média por dia  
with(dados,tapply(C_H, Data, sd)) # Desvio padrão da tonalidade média por dia
```

APÊNDICE F - Sintaxe do software R para o ajuste de modelos e análise inferencial para o *scanner*

```
=====  
## Pacotes utilizados  
=====
```

```

library(lattice)
library(nlme)
library(geoR)
library(car)
library(graphics)
library(stats)
library(agricolae)

=====

## Classe de objeto groupedData para compactar a fórmula com os dados.
=====

Dados=groupedData( S_H ~ Data | Fruto,
                  data = as.data.frame( dados ),
                  FUN = mean,
                  labels = list( Data = "Tempo após o início do ensaio",
                                S_H = "Tonalidade média pelo scanner" ),
                  units = list( Data = "(dias)", S_H = "graus") )

Dados

=====

## Construção de um modelo linear de efeitos fixos para cada indivíduo a partir
## do lmList.
=====

modelo3<-lmList(S_H~Data+I(Data^2),data=Dados);modelo3

=====

## Gráfico dos intervalos de confiança dos parâmetros de efeito fixo individuais.
=====

plot(intervals(modelo3))

=====

## Construção dos modelos lineares de efeitos mistos a partir da função lme do
## pacote de funções nlme utilizando o método de máxima verossimilhança para
## estimação dos parâmetros de efeito fixo e aleatório.
=====

modelo4m<-lme(S_H~1, data=Dados,random=~1,method="REML");modelo4m
modelo5m<-lme(S_H~Data, data=Dados,random=~1,method="REML");modelo5m
modelo6m<-update(modelo5,random=~Data-1,method="REML");modelo6m

```

```

modelo7m<-update(modelo5,random=~Data,method="REML");modelo7m
modelo8m<-lme(S_H~Data+I(Data^2),data=Dados,random=~1,method="REML");modelo8m
modelo9m<-update(modelo8,random=~Data-1,method="REML");modelo9m
modelo10m<-update(modelo8,random=~I(Data^2)-1,method="REML");modelo10m
modelo11m<-update(modelo8,random=~Data,method="REML");modelo11m
modelo12m<-update(modelo8,random=~I(Data^2),method="REML");modelo12m
modelo13m<-update(modelo8,random=~Data+I(Data^2)-1,method="REML");modelo13m
modelo14m<-update(modelo8,random=~Data+I(Data^2),method="REML");modelo14m
=====
## Comparação dos modelos M11 a M13 ao modelo de referência M14. Critérios de
## informação AIC e BIC para os modelos M4 a M10 ajustados aos dados de
## tonalidade média.
=====
AIC(modelo4m, modelo5m, modelo6m, modelo7m, modelo8m, modelo9m, modelo10m)
BIC(modelo4m, modelo5m, modelo6m, modelo7m, modelo8m, modelo9m, modelo10m)
anova(modelo11m,modelo14m)
anova(modelo12m,modelo14m)
anova(modelo13m,modelo14m)
=====
## Contrução do modelo selecionado utilizando o método de máxima verossimilhança
## restrita.
=====
modelo14.1<-lme(S_H~Data+I(Data^2),data=Dados,random=~Data+I(Data^2));modelo14.1
=====
## Mudando a estrutura da matriz de variâncias e covariâncias dos efeitos
## aleatórios (D)
=====
#### Matriz diagonal
-----
modelo14.diag=update(modelo14.1,random=list(Fruto=pdDiag(~Data+I(Data^2))))
-----
## Simetria composta
-----
modelo14.bd<-update(modelo14.1,random=pdCompSymm(~Data+I(Data^2)),

```

```

control=lmeControl(opt="optim",maxIter=100,msMaxIter=100,
niterEM=100))

-----

## Multiplo de uma identidade
-----

modelo14.id<-update(modelo14.1,random=list(Fruto=pdIdent(~Data+I(Data^2))))

-----

## Matriz positiva definida geral (Não estruturada)
-----

modelo14.pdg<-update(modelo14.1,random=list(Fruto=pdSymm(~Data+I(Data^2))))

=====

## Comparação entre os modelos pelo teste da razão de verossimilhanças
## para escolha da estrutura da matriz D.
=====

anova.lme(modelo14.id, modelo14.bd)
anova.lme(modelo14.id, modelo14.diag)
anova.lme(modelo14.id,modelo14.pdg)
anova(modelo14.bd,modelo14.diag)
anova(modelo14.bd,modelo14.pdg)
anova(modelo14.diag,modelo14.pdg)

=====

## Mudando a estrutura da matriz de variâncias e covariâncias do modelo
=====

## Componentes de variâncias com heterocedasticidade
-----

modelo14.pgdH<-lme(S_H~Data+I(Data^2),data=Dados,
random=~Data+I(Data^2)|Fruto,
weight=varIdent(form=~1|Data),
control=lmeControl(maxIter=150, msMaxIter=100, niterEM=100,
msMaxEval=300,
msVerbose = TRUE,
returnObject=TRUE));modelo14.pgdH

-----

## Auto regressiva de ordem 1 - AR(1)

```



```

modelo6m2<-update(modelo5,random=~Data-1,
  weight=varIdent(form=~1|Data),
  control=lmeControl(maxIter=150, msMaxIter=100, niterEM=100,
    msMaxEval=300,
    msVerbose = TRUE,
    returnObject=TRUE));modelo6m2
modelo7m2<-update(modelo5,random=~Data,
  weight=varIdent(form=~1|Data),
  control=lmeControl(maxIter=150, msMaxIter=100, niterEM=100,
    msMaxEval=300,
    msVerbose = TRUE,
    returnObject=TRUE));modelo7m2
modelo8m2<-lme(S_H~Data+I(Data^2),data=Dados,random=~1,
  weight=varIdent(form=~1|Data),
  control=lmeControl(maxIter=150, msMaxIter=100, niterEM=100,
    msMaxEval=300,
    msVerbose = TRUE,
    returnObject=TRUE));modelo8m2
modelo9m2<-update(modelo8,random=~Data-1,
  weight=varIdent(form=~1|Data),
  control=lmeControl(maxIter=150, msMaxIter=100, niterEM=100,
    msMaxEval=300,
    msVerbose = TRUE,
    returnObject=TRUE));modelo9m2
modelo10m2<-update(modelo8,random=~I(Data^2)-1,
  weight=varIdent(form=~1|Data),
  control=lmeControl(maxIter=150, msMaxIter=100, niterEM=100,
    msMaxEval=300,
    msVerbose = TRUE,
    returnObject=TRUE));modelo10m2
modelo11m2<-update(modelo8,random=~Data,
  weight=varIdent(form=~1|Data),
  control=lmeControl(maxIter=150, msMaxIter=100, niterEM=100,
    msMaxEval=300,

```

```

msVerbose = TRUE,
returnObject=TRUE));modelo11m2

modelo12m2<-update(modelo8,random=~I(Data^2),
weight=varIdent(form=~1|Data),
control=lmeControl(maxIter=150, msMaxIter=100, niterEM=100,
msMaxEval=300,
msVerbose = TRUE,
returnObject=TRUE));modelo12m2

modelo13m2<-update(modelo8,random=~Data+I(Data^2)-1,
weight=varIdent(form=~1|Data),
control=lmeControl(maxIter=150, msMaxIter=100, niterEM=100,
msMaxEval=300,
msVerbose = TRUE,
returnObject=TRUE));modelo13m2

modelo14m2<-update(modelo8,random=~Data+I(Data^2),
weight=varIdent(form=~1|Data),
control=lmeControl(maxIter=200, msMaxIter=150, niterEM=150,
msMaxEval=300,
msVerbose = TRUE,
returnObject=TRUE));modelo14m2

```

```

=====
## Comparação dos modelos M11 a M13 ao modelo de referência M14. Critérios de
## informação AIC e BIC para os modelos M4 a M10 ajustados aos dados de
## tonalidade média.
=====

```

```

AIC(modelo4m, modelo5m, modelo6m, modelo7m, modelo8m, modelo9m, modelo10m)
BIC(modelo4m, modelo5m, modelo6m, modelo7m, modelo8m, modelo9m, modelo10m)
anova(modelo11m,modelo14m)
anova(modelo12m,modelo14m)
anova(modelo13m,modelo14m)

```

```

=====
## Testando efeitos fixos por meio do teste F condicional.
anova(modelo14.pgDH, test=F)
=====

```

```
## Resumo das informações sobre o modelo
=====

summary(modelo14.pgdH)
=====

## Gráfico dos resíduos padronizados para um mesmo indivíduo em função do tempo.
## Este gráfico mostra claramente a existência de padrões periódicos nos
## resíduos.
=====

plot(modelo14.pgdH,resid(.,type="p")~Data|Fruto, xlab="Tempo após início do
ensaio (Dias)", ylab="Resíduos condicionais padronizados",
panel=function(x,y,...){
panel.grid()
panel.xyplot(x,y)
panel.loess(x,y,lty=2)
panel.abline(0,0)})
=====

## Teste de Shapiro-Wilk para normalidade dos resíduos padronizados.
=====

shapiro.test(residuals(modelo14.pgdH,type="p"))
=====

## Verificando a assimetria dos resíduos padronizados.
=====

skewness((residuals(modelo14.pgdH,type="p")))
=====

## Verificando a curtose dos resíduos padronizados.
=====

kurtosis((residuals(modelo14.pgdH,type="p")))
=====

## Histograma dos resíduos padronizados com adição da curva de uma distribuição
## normal padrão.
=====

hist((residuals(modelo14.pgdH,type="p")),prob=T,,main="", ylim=c(0,0.5),
      ylab="Densidade",xlab="Resíduo padronizado")
lines(density((residuals(modelo14.pgdH,type="p"))),col="red",lwd=3)
```



```

curve(dnorm(x),add=T,col="blue",lwd=3)

=====

## Estudo da curtose de uma distribuição normal padrão por meio de simulação.
=====

fn<-function(x) kurtosis(rnorm(303))
k<-sapply(1:1000,fn)
quantile(k,c(0,0.1, .025, 0.05,0.5,0.95,0.975,0.99,1))
hist(k,prob=T, ylab="Densidade", xlab="Amostra de uma normal padrão",
      ylim=c(0,1.55),main="");
lines(density(k),col="red",lwd=3)
abline(v=kurtosis((residuals(modelo14.pgdH,type="p"))),col="blue",lwd=3)

=====

## Estudo da assimetria de uma distribuição normal padrão por meio de simulação.
=====

fs<-function(x) skewness(rnorm(303))
k2<-sapply(1:1000,fn)
quantile(k2,c(0,0.1, .025, 0.05,0.5,0.95,0.975,0.99,1))
hist(k2,prob=T, ylab="Densidade", xlab="Amostra de uma normal padrão",main="");
lines(density(k2),lwd=3,col="red")
abline(v=skewness((residuals(modelo14.pgdH,type="p"))),col="blue",lwd=3)

=====

## Gráfico da normalidade dos resíduos condicionais padronizados com envelope
## simulado.
=====

hnp((residuals(modelo14.pgdH,type="p")),half=F,print=T, ylab="Quantil Amostral",
    xlab="Quantil teórico")

=====

## Gráficos de normalidade dos efeitos aleatórios
=====

fruto_ale<-c(18,2,15,12,10,13,8,20,16,14,17,9,7,19,1,5,3,4,6,11)

-----

#### Extraindo os valores preditos dos efeitos aleatórios
-----

b0AR<-random.effects(modelo14.pgdH)[[1]]

```

```

b1AR<-random.effects(modelo14.pgdH)[[2]]
b2AR<-random.effects(modelo14.pgdH)[[3]]

-----

#### Gráfico de normalidade para b0
-----

qqPlot(b0AR,ylab="Quantil teórico da distribuição normal",
        xlab=expression(b[0]),main="",grid=F)

-----

#### Gráfico de normalidade para b1
-----

qqPlot(b1AR,ylab="Quantil teórico da distribuição normal",
        xlab=expression(b[1]),main="",grid=F)

-----

#### Gráfico de normalidade para b2
-----

qqPlot(b2AR,ylab="Quantil teórico da distribuição normal",
        xlab=expression(b[2]),main="",grid=F)

=====

## Gráfico dos resíduos condicionais padronizados em função do tempo
=====

plot(Dados$Data,resid(modelo14.pgdH,type="p"),ylab="Tempo (Dias)",
     xlab="Resíduos condicionais padronizados")
abline(h=0)

=====

## Gráfico para verificar auto correlação entre os efeitos aleatórios
=====

pairs(modelo14.pgdH)

=====

## Gráfico dos resíduos condicionais padronizados em função dos valores preditos
## para verificar o ajuste do modelo.

=====

plot(Dados$S_H,fitted((modelo14.pgdH)),data=Dados,ylab="Valores Ajustados",
     xlab="Valores Observados")
abline(0,1)

```

```

=====
## Intervalos de confiança para os efeitos fixos, para os efeitos aleatórios e
## para os componentes de variância e para o erro padrão.
=====
intervals(modelo14.pgdH)
=====
## Matriz de variâncias e covariâncias dos efeitos aleatórios.
=====
getVarCov(modelo14.pgdH)
=====
## Matriz de variâncias e covariâncias do modelo marginal para o indivíduo 2.
=====
getVarCov(modelo14.pgdH, type="marginal",individual=2))
=====
## Matriz de variâncias e covariâncias do modelo condicional.
getVarCov(modelo14.pgdH, type="conditional",individual=4)
=====
## Função de autocorrelação empírica.
=====
plot(ACF(modelo14.pgdH),0.1)
=====
## Gráfico dos valores populacionais (obtido fazendo os efeitos aleatórios
## iguais a zero) e das predições dentro do grupo com adição dos valores
## observados.
=====
plot(augPred(modelo14.pgdH, level=c(0,1),order.groups=T),xlab="Tempo após início
do ensaio (Dias)", ylab="Tonalidade média",aspect=1)
=====
## Gráfico das predições dentro de um mesmo grupo.
=====
xyplot(predict(modelo14.pgdH)~ Dados$Data,groups=Fruto,Dados,type=c("l")
      ,xlab="Tempo após o início do ensaio (dias)",ylab="Tonalidade média do
      fruto", lwd=2,aspect = "fill", relation="sliced",ylim=c(70,120))
=====

```

APÊNDICE G - Sintaxe do software R para o ajuste de modelos e análise inferencial para o colorímetro

```

=====
## Pacotes utilizados
=====

library(lattice)
library(nlme)
library(geoR)
library(car)
library(graphics)
library(stats)
library(agricolae)
=====

## Classe de objeto groupedData para compactar a fórmula com os dados.
## Foram retirados os dias 5 e 8.
=====

Dados=groupedData( C_H ~ Data | No..fruto,
                   data = subset(dados, dados$Data!=5 & dados$Data!=8 &
                                dados$Data!=19),
                   FUN = mean,
                   labels = list( Data = "Tempo após o início do ensaio",
                                C_H = "Tonalidade média pelo colorímetro" ),
                   units = list( Data = "(dias)", C_H = "graus") )

Dados
=====

## Construção de um modelo linear de efeitos fixos para cada indivíduo a partir
## do lmList.
=====

modelo3<-lmList(C_H~Data+I(Data^2),data=Dados);modelo3
=====

## Gráfico dos intervalos de confiança dos parâmetros de efeito fixo individuais.
=====

plot(intervals(modelo3))
=====

```

```
## Construção dos modelos lineares de efeitos mistos a partir da função lme do
## pacote de funções nlme utilizando o método de máxima verossimilhança para
## estimação dos parâmetros de efeito fixo e aleatório.
```

```
=====
modelo4m<-lme(C_H~1, data=Dados,random=~1,method="REML");modelo4m
modelo5m<-lme(C_H~Data, data=Dados,random=~1,method="REML");modelo5m
modelo6m<-update(modelo5,random=~Data-1,method="REML");modelo6m
modelo7m<-update(modelo5,random=~Data,method="REML");modelo7m
modelo8m<-lme(C_H~Data+I(Data^2),data=Dados,random=~1,method="REML");modelo8m
modelo9m<-update(modelo8,random=~Data-1,method="REML");modelo9m
modelo10m<-update(modelo8,random=~I(Data^2)-1,method="REML");modelo10m
modelo11m<-update(modelo8,random=~Data,method="REML");modelo11m
modelo12m<-update(modelo8,random=~I(Data^2),method="REML");modelo12m
modelo13m<-update(modelo8,random=~Data+I(Data^2)-1,method="REML");modelo13m
modelo14m<-update(modelo8,random=~Data+I(Data^2),method="REML");modelo14m
=====
```

```
## Comparação dos modelos M11 a M13 ao modelo de referência M14. Critérios de
## informação AIC e BIC para os modelos M4 a M10 ajustados aos dados de
## tonalidade média.
```

```
=====
AIC(modelo4m, modelo5m, modelo6m, modelo7m, modelo8m, modelo9m, modelo10m)
BIC(modelo4m, modelo5m, modelo6m, modelo7m, modelo8m, modelo9m, modelo10m)
anova(modelo11m,modelo14m)
anova(modelo12m,modelo14m)
anova(modelo13m,modelo14m)
=====
```

```
## Contrução do modelo selecionado utilizando o método de máxima verossimilhança
## restrita.
```

```
=====
modelo14.1<-lme(C_H~Data+I(Data^2),data=Dados,random=~Data+I(Data^2));modelo14.1
=====
```

```
## Mudando a estrutura da matriz de variâncias e covariâncias dos efeitos
## aleatórios (D)
```

```
=====
```

```
## Matriz diagonal
```

```
-----
modelo14.diag=update(modelo14.1,random=list(Fruto=pdDiag(~Data+I(Data^2))))
-----
```

```
## Simetria composta
```

```
modelo14.bd<-update(modelo14.1,random=pdCompSymm(~Data+I(Data^2)),
                    control=lmeControl(maxIter=100,msMaxIter=100,
                                       niterEM=100))
-----
```

```
## Multiplo de uma identidade
```

```
-----
modelo14.id<-update(modelo14.1,random=list(Fruto=pdIdent(~Data+I(Data^2))))
-----
```

```
## Matriz positiva definida geral (Não estruturada)
```

```
-----
modelo14.pdg<-update(modelo14.1,random=list(Fruto=pdSymm(~Data+I(Data^2))))
=====
```

```
## Comparação entre os modelos pelo teste da razão de verossimilhanças
## para escolha da estrutura da matriz D.
=====
```

```
anova.lme(modelo14.id, modelo14.bd)
anova.lme(modelo14.id, modelo14.diag)
anova.lme(modelo14.id,modelo14.pdg)
anova(modelo14.bd,modelo14.diag)
anova(modelo14.bd,modelo14.pdg)
anova(modelo14.diag,modelo14.pdg)
=====
```

```
## Mudando a estrutura da matriz de variâncias e covariâncias do modelo
=====
```

```
## Componentes de variâncias com heterocedasticidade
-----
```

```
modelo14.pgdH<-lme(C_H~Data+I(Data^2),data=Dados,
                  random=~Data+I(Data^2)|Fruto,
                  weight=varIdent(form=~1|Data),
```

```

        control=lmeControl(maxIter=150, msMaxIter=100, niterEM=100,
                            msMaxEval=300,
                            msVerbose = TRUE,
                            returnObject=TRUE));modelo14.pgdH
=====
## Comparação entre os modelos pelo teste da razão de verossimilhanças
## para escolha da estrutura da matriz.
=====
anova(modelo14.pdg,modelo14.pgdH)
=====
## Nova seleção para o modelo considerando as matrizes selecionadas
## anteriormente
=====
modelo4m2<-lme(C_H~1, data=Dados,random=~1,
               weight=varIdent(form=~1|Data),
               control=lmeControl(maxIter=150, msMaxIter=100, niterEM=100,
                                   msMaxEval=300,
                                   msVerbose = TRUE,
                                   returnObject=TRUE));modelo4m2

modelo5m2<-lme(C_H~Data, data=Dados,random=~1,
               weight=varIdent(form=~1|Data),
               control=lmeControl(maxIter=150, msMaxIter=100, niterEM=100,
                                   msMaxEval=300,
                                   msVerbose = TRUE,
                                   returnObject=TRUE));modelo5m2

modelo6m2<-update(modelo5,random=~Data-1,
                  weight=varIdent(form=~1|Data),
                  control=lmeControl(maxIter=150, msMaxIter=100, niterEM=100,
                                      msMaxEval=300,
                                      msVerbose = TRUE,
                                      returnObject=TRUE));modelo6m2

modelo7m2<-update(modelo5,random=~Data,
                  weight=varIdent(form=~1|Data),

```

```
control=lmeControl(maxIter=150, msMaxIter=100, niterEM=100,  
msMaxEval=300,  
msVerbose = TRUE,  
returnObject=TRUE));modelo7m2  
modelo8m2<-lme(C_H~Data+I(Data^2),data=Dados,random=~1,  
weight=varIdent(form=~1|Data),  
control=lmeControl(maxIter=150, msMaxIter=100, niterEM=100,  
msMaxEval=300,  
msVerbose = TRUE,  
returnObject=TRUE));modelo8m2  
modelo9m2<-update(modelo8,random=~Data-1,  
weight=varIdent(form=~1|Data),  
control=lmeControl(maxIter=150, msMaxIter=100, niterEM=100,  
msMaxEval=300,  
msVerbose = TRUE,  
returnObject=TRUE));modelo9m2  
modelo10m2<-update(modelo8,random=~I(Data^2)-1,  
weight=varIdent(form=~1|Data),  
control=lmeControl(maxIter=150, msMaxIter=100, niterEM=100,  
msMaxEval=300,  
msVerbose = TRUE,  
returnObject=TRUE));modelo10m2  
modelo11m2<-update(modelo8,random=~Data,  
weight=varIdent(form=~1|Data),  
control=lmeControl(maxIter=150, msMaxIter=100, niterEM=100,  
msMaxEval=300,  
msVerbose = TRUE,  
returnObject=TRUE));modelo11m2  
modelo12m2<-update(modelo8,random=~I(Data^2),  
weight=varIdent(form=~1|Data),  
control=lmeControl(maxIter=150, msMaxIter=100, niterEM=100,  
msMaxEval=300,  
msVerbose = TRUE,  
returnObject=TRUE));modelo12m2
```



```

modelo13m2<-update(modelo8,random=~Data+I(Data^2)-1,
                    weight=varIdent(form=~1|Data),
                    control=lmeControl(maxIter=150, msMaxIter=100, niterEM=100,
                                       msMaxEval=300,
                                       msVerbose = TRUE,
                                       returnObject=TRUE));modelo13m2

modelo14m2<-update(modelo8,random=~Data+I(Data^2),
                    weight=varIdent(form=~1|Data),
                    control=lmeControl(maxIter=200, msMaxIter=150, niterEM=150,
                                       msMaxEval=300,
                                       msVerbose = TRUE,
                                       returnObject=TRUE));modelo14m2

=====
## Comparação dos modelos M11 a M13 ao modelo de referência M14. Critérios de
## informação AIC e BIC para os modelos M4 a M10 ajustados aos dados de
## tonalidade média.
=====
AIC(modelo4m, modelo5m, modelo6m, modelo7m, modelo8m, modelo9m, modelo10m)
BIC(modelo4m, modelo5m, modelo6m, modelo7m, modelo8m, modelo9m, modelo10m)
anova(modelo11m,modelo14m)
anova(modelo12m,modelo14m)
anova(modelo13m,modelo14m)
=====
## Testando efeitos fixos por meio do teste F condicional.
=====
anova(modelo14.pgdH, test=F)
=====
## Resumo das informações sobre o modelo
=====
summary(modelo14.pgdH)
=====
## Gráfico dos resíduos padronizados para um mesmo indivíduo em função do tempo.
## Este gráfico mostra claramente a existência de padrões periódicos nos resíduos.
=====

```

```

plot(modelo14.pgdH,resid(.,type="p")~Data|Fruto, xlab="Tempo após início do
ensaio (Dias)", ylab="Resíduos condicionais padronizados",
panel=function(x,y,...){
panel.grid()
panel.xyplot(x,y)
panel.loess(x,y,lty=2)
panel.abline(0,0)})

=====

## Teste de Shapiro-Wilk para normalidade dos resíduos padronizados.
=====

shapiro.test(residuals(modelo14.pgdH,type="p"))

=====

## Verificando a assimetria dos resíduos padronizados.
=====

skewness((residuals(modelo14.pgdH,type="p")))

=====

## Verificando a curtose dos resíduos padronizados.
=====

kurtosis((residuals(modelo14.pgdH,type="p")))

=====

## Histograma dos resíduos padronizados com adição da curva de uma distribuição
## normal padrão.
=====

hist((residuals(modelo14.pgdH,type="p")),prob=T,,main="", ylim=c(0,0.5),
      ylab="Densidade",xlab="Resíduo padronizado")
lines(density((residuals(modelo14.pgdH,type="p"))),col="red",lwd=3)
curve(dnorm(x),add=T,col="blue",lwd=3)

=====

## Estudo da curtose de uma distribuição normal padrão por meio de simulação.
=====

fn<-function(x) kurtosis(rnorm(303))
k<-sapply(1:1000,fn)
quantile(k,c(0,0.1, .025, 0.05,0.5,0.95,0.975,0.99,1))
hist(k,prob=T, ylab="Densidade", xlab="Amostra de uma normal padrão",

```

```

      ylim=c(0,1.55),main="");
lines(density(k),col="red",lwd=3)
abline(v=kurtosis((residuals(modelo14.pgdH,type="p"))),col="blue",lwd=3)
=====

## Estudo da assimetria de uma distribuição normal padrão por meio de simulação.
=====

fs<-function(x) skewness(rnorm(303))
k2<-sapply(1:1000,fn)
quantile(k2,c(0,0.1, .025, 0.05,0.5,0.95,0.975,0.99,1))
hist(k2,prob=T, ylab="Densidade", xlab="Amostra de uma normal padrão",main="");
lines(density(k2),lwd=3,col="red")
abline(v=skewness((residuals(modelo14.pgdH,type="p"))),col="blue",lwd=3)
=====

## Gráfico da normalidade dos resíduos condicionais padronizados com envelope
## simulado.
=====

hnp((residuals(modelo14.pgdH,type="p")),half=F,print=T, ylab="Quantil Amostral",
    xlab="Quantil teórico")
=====

## Gráficos de normalidade dos efeitos aleatórios
=====

fruto_ale<-c(18,2,15,12,10,13,8,20,16,14,17,9,7,19,1,5,3,4,6,11)
-----

## Extraíndo os valores preditos dos efeitos aleatórios
-----

b0AR<-random.effects(modelo14.pgdH)[[1]]
b1AR<-random.effects(modelo14.pgdH)[[2]]
b2AR<-random.effects(modelo14.pgdH)[[3]]
-----

## Gráfico de normalidade para b0
-----

qqPlot(b0AR,ylab="Quantil teórico da distribuição normal",
    xlab=expression(b[0]),main="",grid=F)
-----

```

```

## Gráfico de normalidade para b1
-----

qqPlot(b1AR,ylab="Quantil teórico da distribuição normal",
        xlab=expression(b[1]),main="",grid=F)
-----

## Gráfico de normalidade para b2
-----

qqPlot(b2AR,ylab="Quantil teórico da distribuição normal",
        xlab=expression(b[2]),main="",grid=F)
=====

## Gráfico dos resíduos condicionais padronizados em função do tempo
=====

plot(Dados$Data,resid(modelo14.pgdH,type="p"),ylab="Tempo (Dias)",
     xlab="Resíduos condicionais padronizados")
abline(h=0)
=====

## Gráfico para verificar auto correlação entre os efeitos aleatórios
=====

pairs(modelo14.pgdH)
=====

## Gráfico dos resíduos condicionais padronizados em função dos valores preditos
## para verificar o ajuste do modelo.
=====

plot(Dados$C_H,fitted((modelo14.pgdH)),data=Dados,ylab="Valores Ajustados",
     xlab="Valores Observados")
abline(0,1)
=====

## Intervalos de confiança para os efeitos fixos, para os efeitos aleatórios e
## para os componentes de variância e para o erro padrão.
=====

intervals(modelo14.pgdH)
=====

## Matriz de variâncias e covariâncias dos efeitos aleatórios.
=====

```

```

getVarCov(modelo14.pgdH)

=====

## Matriz de variâncias e covariâncias do modelo marginal para o indivíduo 2.

=====

getVarCov(modelo14.pgdH, type="marginal",individual=2))

=====

## Matriz de variâncias e covariâncias do modelo condicional.

=====

getVarCov(modelo14.pgdH, type="conditional",individual=4)

=====

## Gráfico representando o semivariograma amostral obtido com a função Variogram

=====

plot(Variogram(modelo14.pgdH,form=~Data,robust=T,resType="pearson",data=Dados2))

=====

## Gráfico dos valores populacionais (obtido fazendo os efeitos aleatórios
## iguais a zero) e das predições dentro do grupo com adição dos valores
## observados.

=====

plot(augPred(modelo14.pgdH, level=c(0,1),order.groups=T),xlab="Tempo após início
do ensaio (Dias)", ylab="Tonalidade média",aspect=1)

=====

## Gráfico das predições dentro de um mesmo grupo.

=====

xyplot(predict(modelo14.pgdH)~ Dados$Data,groups=Fruto,Dados,type=c("l")
      ,xlab="Tempo após o início do ensaio (dias)",ylab="Tonalidade média do
      fruto",lwd=2,aspect = "fill", relation="sliced",ylim=c(70,120))

```

ANEXOS

ANEXO A - Dados de tonalidade média ao longo do tempo do fruto 1 ao 20 medidos por meio de um *scanner* e de um colorímetro.

Dados obtidos a partir do experimento para análise da tonalidade média da casca de mamões “Sunrise Solo” avaliados por meio de um colorímetro e de um *scanner* ao longo do tempo. Nas Tabelas 18 a 20 são apresentados os dados de tonalidade média dos dias 0 a 2 e 9 a 17. O experimento foi financiado pela Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) por meio do programa de Iniciação Científica (processo 2010/16955-1) no ano agrícola de 2010/2011.

Tabela 18 – Dados de tonalidade média (em graus) dos frutos 1 ao 4 medidos ao longo do tempo (em dias) por meio de um *scanner* (Scan) e de um colorímetro (Col)

Dia	Fruto 1		Fruto 2		Fruto 3		Fruto 4	
	Scan	Col	Scan	Col	Scan	Col	Scan	Col
0	111,56	116,54	113,83	115,41	113,75	116,98	112,09	115,71
1	108,84	113,82	109,29	114,57	112,71	116,83	109,92	114,44
2	107,98	112,57	107,30	112,90	111,93	116,87	108,35	112,34
:	:	:	:	:	:	:	:	:
8	101,18	106,17	94,07	98,67	104,55	109,07	98,58	102,21
9	99,41	102,87	93,33	96,11	103,34	107,08	96,27	100,33
10	98,13	101,22	90,63	93,33	102,33	104,50	95,27	97,29
11	95,51	97,83	88,41	90,86	100,05	101,88	91,42	95,29
12	94,06	94,25	86,36	88,05	98,15	99,89	92,16	92,61
13	92,78	93,58	85,60	85,88	96,78	98,60		
14	89,71	91,78	82,86	83,96	93,05	95,51		
15	89,60	87,94	82,77	80,93	90,67	92,91		
16			80,36	77,89	88,83	88,90		
17			78,40	76,95	85,15	86,41		

Tabela 19 – Dados de tonalidade média (em graus) dos frutos 5 ao 8 medidos ao longo do tempo (em dias) por meio de um *scanner* (Scan) e de um colorímetro (Col)

Dia	Fruto 5		Fruto 6		Fruto 7		Fruto 8	
	Scan	Col	Scan	Col	Scan	Col	Scan	Col
0	112,77	115,88	113,43	116,57	111,55	116,14	112,16	114,44
1	111,86	115,18	108,25	115,83	112,34	116,68	108,69	114,22
2	109,47	113,72	107,41	115,29	111,89	115,81	107,54	112,71
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
8	94,69	99,26	98,23	107,33	105,99	110,55	97,91	101,33
9	93,16	96,30	95,90	105,12	103,95	109,09	94,66	98,50
10	92,15	93,01	94,82	101,94	103,48	106,26	94,17	93,87
11			92,77	99,41	101,47	103,24	91,28	90,63
12			91,15	96,29	99,27	101,35	87,76	87,73
13			89,90	93,98	98,12	99,70	87,57	86,38
14			87,12	92,03	95,44	97,81		
15			86,48	90,10	94,80	96,21		
16			84,78	86,32				
17			83,20	86,42				

Tabela 20 – Dados de tonalidade média (em graus) dos frutos 9 ao 12 medidos ao longo do tempo (em dias) por meio de um *scanner* (Scan) e de um colorímetro (Col)

Dia	Fruto 9		Fruto 10		Fruto 11		Fruto 12	
	Scan	Col	Scan	Col	Scan	Col	Scan	Col
0	112,40	116,10	111,51	115,18	109,26	114,38	112,06	115,56
1	110,03	114,50	109,11	113,83	108,51	113,79	112,60	117,08
2	107,37	113,17	107,45	112,23	107,61	112,41	112,78	116,84
:	:	:	:	:	:	:	:	:
8	98,77	103,80	96,26	99,75	95,15	100,50	104,75	111,54
9	96,59	101,46	94,35	97,89	92,55	97,15	103,04	110,16
10	95,33	98,67	92,33	94,65	91,48	95,25	101,89	106,72
11	93,08	97,17	91,05	92,54	89,02	91,96	100,16	104,75
12	91,11	93,47	89,31	89,25	86,89	89,21	98,95	102,51
13	89,66	91,64			85,12	85,62		
14	85,90	88,56			82,36	83,38		
15	86,17	87,41			81,20	81,31		

Tabela 21 – Dados de tonalidade média (em graus) dos frutos 13 ao 16 medidos ao longo do tempo (em dias) por meio de um *scanner* (Scan) e de um colorímetro (Col)

Dia	Fruto 13		Fruto 14		Fruto 15		Fruto 16	
	Scan	Col	Scan	Col	Scan	Col	Scan	Col
0	111,16	112,96	112,31	115,48	109,77	115,05	111,48	114,32
1	109,02	111,28	110,29	115,29	109,24	113,49	109,51	113,27
2	107,63	108,78	109,94	113,79	108,10	111,23	104,25	112,01
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
8	95,01	96,35	98,60	102,92	95,53	98,14	97,10	98,97
9	93,18	95,31	96,42	98,14	92,69	96,16	94,66	98,14
10	90,99	91,47	94,57	95,84	91,06	91,62	93,04	93,21
11	88,77	90,04	91,74	92,75	88,76	88,77	91,08	91,62
12	87,69	86,45	89,56	89,44	87,48	85,30	89,16	88,04
13	86,03	84,92	87,80	86,90			87,70	85,90
14	83,33	82,60	84,21	83,97				
15	82,65	81,00	84,09	81,90				

Tabela 22 – Dados de tonalidade média (em graus) dos frutos 17 ao 20 medidos ao longo do tempo (em dias) por meio de um *scanner* (Scan) e de um colorímetro (Col)

Dia	Fruto 17		Fruto 18		Fruto 19		Fruto 20	
	Scan	Col	Scan	Col	Scan	Col	Scan	Col
0	111,65	115,75	107,88	113,15	111,24	115,59	111,42	116,41
1	111,49	115,30	107,00	111,64	109,89	115,16	111,76	116,26
2	110,68	114,11	106,34	110,11	109,66	112,93	110,52	114,65
:	:	:	:	:	:	:	:	:
8	99,84	106,66	95,05	97,43	96,41	98,03	100,07	105,37
9	97,12	104,12	92,66	95,54	93,07	96,14	95,93	100,26
10	95,35	100,15	91,35	91,61	91,14	91,00	94,44	98,88
11	93,18	97,85	89,38	89,31	87,64	88,92	92,16	96,13
12	91,22	94,49	87,54	86,45	86,08	85,68	89,88	92,78
13	89,13	92,54	85,94	84,71			88,19	90,32
14	86,40	89,23	83,04	82,12			85,36	87,24
15	85,35	87,26	82,11	80,10			84,54	85,44
16			80,56	77,87				
17			78,96	76,15				