# Universidade de São Paulo Escola Superior de Agricultura "Luiz de Queiroz"

# Análise da cor da casca do mamão cv. Sunrise Solo por meio de modelo de regressão linear misto

# Caroline Oliveira do Nascimento

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

Piracicaba 2019

Caroline Oliveira do Nascimento Bacharela em Matemática

# Análise da cor da casca do mamão cv. Sunrise Solo por meio de modelo de regressão linear misto

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Orientador: Prof. Dr. IDEMAURO ANTONIO RODRIGUES DE LARA

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

Piracicaba 2019

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

Nascimento, Caroline Oliveira do

Análise da cor da casca do mamão cv. Sunrise Solo por meio de modelo de regressão linear misto / Caroline Oliveira do Nascimento. – – versão revisada de acordo com a resolução CoPGr 6018 de 2011. – – Piracicaba, 2019 . 64 p.

Dissertação (Mestrado) – – USP / Escola Superior de Agricultura "Luiz de Queiroz".

Regressão 2. Dados longitudinais 3. Seleção de modelos 4. Fruticultura
 I. Título.

Aos meus pais, Adriana Nascimento e Cláudio Nascimento e ao meu noivo, Marcell Silva, **Dedico**.

#### AGRADECIMENTOS

A Deus por seu amor incondicional, por ser meu amigo, meu refúgio nos dias difíceis, se cheguei até aqui é por que esteve comigo sempre. Grata sou pela vida que confiou à mim, tua seja a glória.

Não há palavras suficientes para expressar minha gratidão à minha família. Aos meus pais, Adriana e Cláudio, por não medirem esforços para me proporcionar sempre o melhor e me ensinar o valor do estudo. Ao meu noivo, Marcell, pelo incentivo e compreensão em todos os momentos. Com certeza minhas conquistas não me trariam tanta alegria se não tivesse vocês ao meu lado para compartilhar, amo vocês.

Ao Prof. Dr. Idemauro Antonio Rodrigues de Lara, sempre disponível para ajudar e desde que passou a ser meu orientador me trouxe motivação. Obrigada pela paciência, por todo apoio e tempo dedicado à minha orientação para que esse trabalho fosse concluído.

Ao Prof. Dr. Thiago de Paula Oliveira, meu agradecimento pelos dados fornecidos e por toda ajuda, que foi indispensável para desenvolver esse trabalho.

Aos Professores que compuseram a banca no exame de qualificação, Dr. César Gonçalves de Lima e Dr. Antônio Francisco Iemma, e a banca de defesa, Dr. Antônio Francisco Iemma, Dra. Sônia Maria De Stefano Piedade e Dr. Thiago de Paula Oliveira. Grata sou pelas correções e sugestões que contribuíram para o trabalho final.

A todos docentes e funcionários do Departamento de Ciências Exatas da Escola Superior de Agricultura "Luiz de Queiroz" (ESALQ), agradeço pelo conhecimento compartilhado e pela ajuda direta ou indireta nesse período.

Agradeço à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo auxílio financeiro nesse período.

Muito obrigada a todos que contribuíram de alguma maneira para a realização desse trabalho.

EPÍGRAFE

"O correr da vida embrulha tudo. A vida é assim: esquenta e esfria, Aperta e daí afrouxa, Sossega e depois desinquieta. O que ela quer da gente é coragem." *Guimarães Rosa* 

# SUMÁRIO

Resu	ımo											
Abstract												
Lista de Figuras												
Lista de Tabelas												
1 Introdução												
2 F	Revis	ão de li	teratura									
2.1 Análise de cor da casca por meio de imagens digitais												
2.2 Amadurecimento e características fisiológicas do mamão												
2.3 Modelo de regressão linear												
		2.3.1	Estimação									
		2.3.2	Testes de hipóteses e seleção de modelos									
2	2.4	Modelc	) de regressão linear misto									
		2.4.1	Estimação									
		2.4.2	Estrutura das matrizes de variâncias e covariâncias									
		2.4.3	Seleção de modelos									
			2.4.3.1 Critério de informação de Akaike (AIC)									
			2.4.3.2 Critério de informação Bayesiano (BIC)									
			2.4.3.3 Teste da razão de verossimilhança (TRV)									
		2.4.4	Resíduos									
3 N	Mate	erial										
4 N	Véto	odos										
5 F	Resu	ltados e	Discussão									
5	5.1	Tonalid	lade									
		5.1.1	Seleção das estruturas das matrizes D e $\Sigma_i$									
		5.1.2	Diagnósticos									
5	5.2	Intensio	Jade									
		5.2.1	Seleção das estruturas das matrizes D e $\Sigma_i$									
		5.2.2	Diagnósticos									
6 (	Cons	ideraçõe	es finais									
Referências												
Ane>	xos .		59									

#### RESUMO

# Análise da cor da casca do mamão cv. Sunrise Solo por meio de modelo de regressão linear misto

O mamão (Carica papaya L.) tem importância destacada na fruticultura e se encontra entre os seis principais produtos que somam mais de 50% da produção nacional desse setor. O mamão tem uma maturação relativamente rápida. Visando aumentar o potencial de comércio e possivelmente diminuir as perdas pós-colheita, a análise de imagens digitais é um recurso tecnológico para avaliar a tonalidade e intensidade da cor da casca dos frutos no período de maturação, que serve de base para estabelecer modelos funcionais para mensurações realizadas num período de tempo. Nesse contexto tem como motivação um estudo longitudinal envolvendo a avaliação da intensidade e tonalidade da cor da casca do mamão da espécie Carica papaya L. no período de maturação. Para a análise dos dados é utilizada a metodologia dos modelos lineares de efeitos mistos e para selecionar os modelos que melhor se ajustavam aos dados, utilizou-se teste da razão de verossimilhanças e teste F, em um método de seleção top-down. Verifica-se que modelo polinomial quadrático com efeito aleatório em todos os parâmetros descreve de maneira satisfatória a variável tonalidade. Para a variável intensidade obteve-se um modelo polinomial cúbico para os efeitos aleatórios e apenas o intercepto como parâmetro de efeito fixo. As análises de diagnóstico confirmaram o ajuste satisfatório dos modelos.

Palavras-chave: Regressão, Dados longitudinais, Seleção de modelos, Fruticultura

### ABSTRACT

# Analysis of color peel of the papaya cv. Sunrise Solo through of the mixed linear regression model

The papaya (*Carica papaya* L.) has important importance in fruticulture and is among the six main products that add up to more than 50% of the national production of this sector. Papaya has a relatively rapid maturation. In order to increase commercial potential and possibly reduce post-harvest losses, digital image analysis is a technological tool to evaluate the color tone and intensity of fruit peel during the maturation period, which serves as the basis for establishing functional models for measurements performed over a period of time. In this context it has as motivation a longitudinal study involving the evaluation of the intensity and color tone of the shell of the papaya of the species *Carica papaya* L. in the maturation period. For the analysis of the data the methodology of the linear models of mixed effects is used and to select the models that best fit the data, was used a test of the likelihood ratio and test F, in a method of selection *top-down*. It can be verified that the quadratic polynomial model with random effect in all the parameters describes in a satisfactory way the variable tonality. For the intensity variable we obtained a cubic polynomial model for the random effects and only the intercept as a fixed effect parameter. Diagnostic analyzes confirmed the satisfactory fit of the models.

Keywords: Regression, Longitudinal data, Selection of models, Fruticulture

# LISTA DE FIGURAS

2.1	Região das cores pertencentes ao sistema de cores CIELab.	17
$3.1 \\ 3.2$	Lado A e Lado B de um mamão "Sunrise Solo" Papaya	31
3.3	em 2010/2011	34 35
5.1	Intervalos de confiança individuais dos coeficientes do modelo polinomial de grau 2 (M2) para a variável tonalidade.	39
5.2	Gráfico de dispersão entre os efeitos aleatórios do modelo maximal (M1.t)	41
5.3	Gráfico das predições do modelo final (linha conínua) com as observações (pontos)	
	para cada fruto	43
5.4	Gráfico dos resíduos em função do tempo (a) e gráfico dos valores ajustados em	
	função dos valores observados (b) para a variável tonalidade	44
5.5	Gráfico dos resíduos condicionais padronizados em função dos valores ajustados	
	para a variável tonalidade	44
5.6	Gráfico normal de probabilidade dos resíduos com envelope simulado (a) e Histo- grama dos resíduos com adicional de uma curva da distribuição normal (b) para a	
		45
5.7	Graficos normais de probabilidade com envelope simulado para os efeitos aleatorios $L_{\rm c}(\lambda) = L_{\rm c}(\lambda) = L_{\rm c}(\lambda) = L_{\rm c}(\lambda) = L_{\rm c}(\lambda)$	45
٣٥	$b_0$ (a), $b_1$ (b) e $b_2$ (c) do modelo quadratico para a variavel tonalidade	45
0.8	intervalos de comança individuais dos coencientes do modelo pomomíal de grau $2 (M1^*)$ para a variával intensidada	46
5.0	Cráfico de dispersão entre os efeitos alextário de modelo M8 i	40
5.10	Gráfico das predições do modelo final (linha conínua) com as observações (pontos)	41
0.10	para cada fruto	49
5.11	Gráfico dos resíduos em função do tempo (a) e gráfico dos valores ajustados em	чJ
0.11	função dos valores observados (b) para a variável intensidade.	50
5.12	Gráfico dos resíduos condicionais padronizados em função dos valores ajustados	
	para a variável intensidade	51
5.13	Gráfico de normalidade dos resíduos com envelope simulado da distribuição normal	
	(a) e Histograma de normalidade dos resíduos com adicional de uma curva da	
	distribuição normal (b) para a variável intensidade	51
5.14	Gráficos normais de probabilidade com envelope simulado para os efeitos aleatórios	
	$b_0$ (a), $b_1$ (b), $b_2$ (c) e $b_3$ (d) do modelo cúbico para a variável intensidade	52

# LISTA DE TABELAS

2.1	Classificação dos estágios de maturação do mamão Carica papaya L.	18
3.1	Médias e desvios padrões da variável tonalidade da casca do mamão ao longo do tempo no experimento referente à maturação do <i>Carica Papaya</i> L	32
3.2	Médias e desvios padrões da variável intensidade da casca do mamão ao longo do tempo no experimento referente à maturação do <i>Carica Papaya</i> L	32
3.3	Variâncias e covariâncias (em negrito) e correlações amostrais da variável tonalidade.	32
3.4	Variâncias e covariâncias (em negrito) e correlações amostrais da variável intensidade.	33
3.5	Número de frutos analisados por dia no experimento referente à maturação do <i>Carica Papaya</i> L	33
3.6	Número total de observações por fruto para cada variável do experimento referente à maturação do <i>Carica Papaya</i> L	33
4.1	Modelos lineares de efeitos fixos considerados para as variáveis tonalidade (M1 e M2) e intensidade (M1 <sup>*</sup> e M2 <sup>*</sup> ).	37
5.1	Número de parâmetros (np), Soma de Quadrados dos Resíduos (SQRes) e nível descritivo referente ao Teste da Razão de Verossimilhanças (valor-p) na comparação entre o modelo linear simples (M1) e o modelo polinomial de grau 2 (M2)	39
5.2	Modelos encaixados referentes à variável tonalidade, com respectivos números de parâmetros (np), logarítmo da função de verossimilhança ( $-\log(L)$ ), estatística $\chi^2$ , números de graus de liberdade (gl) e valor descritivo referente ao teste da razão de verossimilhança ( $valor p$ )	40
5.3	Estimativas para o desvio padrão dos efeitos aleatórios e para o coeficiente de correlação entre tais efeitos, com respectivos limites inferiores e limites superiores	40
5.4	dos intervalos de 95% de confiança	40
	(valor-p)	41
5.5	Estruturas consideradas para a matriz de variâncias e covariâncias dos erros do modelo M1.t, em que CVH é componentes de variâncias heterogêneas e ARH(1) é matriz autorregressiva heterogêneae de ordem 1, com respectivos números de parâmetros do modelo (np), valores do critério Bayesiano (BIC), logarítmo da fun- ção de verossimilhança (-log(L)), Estatística do teste da razão de verossimilhanças	
	(TRV) e valor descritivo referente ao teste da razão de veros similhanças (valor-p).	42

5.6	Estimativas dos efeitos fixos do modelo quadrático para a tonalidade, com respec-	
	tivos desvios padrões (DP), limites inferiores e limites superiores dos intervalos de	
	95% de confiança, teste F e valor-p. $\ldots$	42
5.7	Número de parâmetros (np), Soma de Quadrados dos Resíduos (SQRes) e o nível	
	descritível referente ao teste da razão de veros similhanças (valor-p) na comparação $\ensuremath{\sim}$	
	entre o modelo polinomial de grau 3 (M1*) e o modelo polinomial de grau 4 (M2*).	45
5.8	Modelos encaixados referente à variável intensidade, com respectivos números de	
	parâmetros (np), logaritmo da veros similhança (-log(L)), estatística $\chi^2,$ números	
	de graus de liberdade (gl) e nível descritivo referente ao teste da razão de verossi-	
	milhanças (valor-p).	46
5.9	Estimativas para o desvio padrão dos efeitos aleatórios e para o coeficiente de	
	correlação entre tais efeitos, com respectivos limites inferiores e limites superiores	
	dos intervalos de 95% de confiança	47
5.10	Estruturas consideradas para a matriz de variâncias e covariâncias dos efeitos	
	aleatórios do modelo M8.i, em que NE é matriz não estruturada e NE.CV é a	
	bloco-diagonal, com respectivos números de parâmetros (np), valores do critério	
	Bayesiano (BIC), Estatística do teste da razão de verossimilhanças (TRV) e o nível	
	descritivo referente ao TRV (valor-p)	48
5.11	Estruturas consideradas para a matriz de variância e covariância dos erros do	
	modelo M8.i, em que CV é matriz componentes de variâncias, CVH é matriz	
	componentes de variâncias heterogêneas e $\mathrm{AR}(1)$ é autor regressiva de ordem 1, com	
	respectivos números de parâmetros do modelo (np), valores do critério Bayesiano	
	(BIC), logaritmo da verossimilhança ( $-\log(L)$ ), Estatística do teste da razão de	
	verossimilhanças (TRV) e nível descritivo referente ao TRV (valor-p)	48
5.12	Estimativa do efeito fixo do modelo cúbico para a intensidade, com respectivo des-	
	vio padrão (DP), limite inferior e limite superior do intervalo de 95% de confiança,	
	teste F e valor-p	49

# 1 INTRODUÇÃO

A fruticultura é um setor da agricultura relacionada ao cultivo de plantas frutíferas para atender à industrialização ou suprir o mercado de consumo *in natura*. Este é um setor de grande importância social e econômica, pois gera empregos no campo e nos processos pós-colheita (Lima *et al.*, 2018). De acordo com o Serviço Brasileiro de Apoio às Micro e Pequenas Empresas (SEBRAE), o Brasil se destaca na classificação mundial como o terceiro maior produtor de frutas, ficando atrás apenas de China e Índia, tendo uma área colhida de 631,7 mil hectares que geraram cerca de R\$ 38,9 bilhões para a economia nacional (IBGE, 2017), o que resultou em um crescimento de R\$ 5,6 bilhões de reais quando comparado com os números de 2016 (IBGE, 2016).

Dentre a produção nacional no setor da fruticultura, o mamão encontra-se entre os seis principais produtos que somam mais de 50% dessa produção (IBGE, 2016). A produção nacional de mamão em 2016 foi de 1.424,650 toneladas, sendo 37,939 toneladas destinada para comércio exterior, colocando o Brasil na segunda posição na classificação de produtores mundiais dessa cultura (Embrapa, 2018). Os Estados do Espirito Santo, Bahia e Minas Gerais são os estados responsáveis pela maior parte da produção brasileira de mamão, para comercio nacional e internacional, no entanto, tiveram uma redução em suas áreas de plantio em 2018 (Barbieri, 2018).

Visando aumentar e melhorar o comércio fruticultor no Brasil e também nas exportações, é preciso atentar-se aos requisitos de qualidade como, por exemplo, a firmeza e a cor da casca dos frutos, os quais podem influenciar diretamente na escolha do produto pelos consumidores ou exportadores. Existem vários fatores que influenciam na qualidade e durabilidade do fruto até chegar ao seu destino final, como doenças, forma de manusear, local de armazenamento e frutos com estágios de maturação diversificada em um mesmo lote.

Especificamente na questão de maturação das frutas, é importante obter uma classificação dos estágios de maneira a conseguir maior uniformidade na separação dos frutos que serão comercializados. Portanto, para auxiliar na preservação do fruto no período da colheita até a comercialização, se faz necessário o uso de tecnologias adequadas (Chitarra; Chitarra, 2005).

O mamão, em seu período de maturação, apresenta modificações na coloração de sua casca como, por exemplo, a tonalidade alaranjada da casca que ocorre, naturalmente, em frutos que estão no final do estágio de maturação (SEBRAE, 2018). Propõe-se a utilização de sistemas automatizados capazes de captar informações presentes nas cascas (Motta *et. al*, 2015; Oliveira *et. al*, 2017), buscando medir a coloração dos frutos de forma rápida e objetiva durante os períodos pós-colheita (Wu; Sun, 2013) e relacionando com o estágio de maturação. No caso da comercialização, ajudaria a selecionar os frutos de forma mais uniforme e no estágio de maturação adequado, para que não ocorram perdas antes de chegar ao destino ou passe do estágio exigido pelo comprador.

O uso de tecnologias voltadas para a fase de pós-colheita é cada vez maior, afim de manter a qualidade dos frutos quanto à aparência, textura, sabor, valor nutricional, segurança alimentar bem como reduzir perdas qualitativas e quantitativas entre a colheita e o consumo (Silva, 2016). A análise de imagens digitais é considerada um recurso tecnológico na avaliação do fruto quanto a sua parte externa (Carvalho et. al, 2014; Oliveira *et al.*, 2018).

Câmera digital e *scanner* são exemplos de equipamentos utilizados na obtenção de imagens digitais, de maneira objetiva, ajudando na avaliação e classificação dos estágios de maturação de frutos. Para fazer associação entre os *pixels* captados e as cores reais, os equipamentos utilizam um sistema de cores padrão, geralmente RGB (sigla referente a *Red Green Blue*) ou sRGB (standard RGB). Usualmente o sistema de cor é o RGB, em que cada *pixel* da imagem digitalizada é representado pelas intensidades de cada um dos três tons, sendo que o vermelho puro é representado pela coordenada (255,0,0), verde puro pela coordenada (0,255,0), azul puro pela coordenada (0,0,255) e as demais cores são dadas pelas combinações dos valores das variáveis R, G e B (Oliveira *et al.*, 2016).

Buscando maior aproximação possível das cores da imagem digital com as cores vistas pela visão humana, para que a representatividade seja a mais fidedigna da realidade, deve-se levar em conta pequenas variações no espaço de cores RGB os quais não são perceptíveis aos olho humano. Isto é, dependendo da localização do pixel nesse espaço, a cor que o representará pode não ser perceptível pelo olho humano.

Para eliminar esse problema, geralmente os aparelhos de medição de cor usam o sistema criado pela Comissão Internacional de *L'Eclairage* (CIE), considerada uma autoridade na ciência da luz e cor. Como exemplos pode-se citar o espaço de cor  $L^*a^*b^*$  (sigla referente a luminosidade, cores que variam de vermelhas a verdes e de amarelas a azuis, respectivamente) e  $L^*C^*h$  (refere-se a luminosidade, intensidade e tonalidade, respectivamente).

Assim, pode-se utilizar a imagem digital da superfície da casca do mamão para analisar características presentes na mesma. Como o mamão tem uma maturação relativamente rápida, o que possibilita maiores perdas em curto período de tempo (Balbino, 1997), utilizar o recurso de análise de imagens seria interessante para facilitar na classificação do estágio de maturação, além de poder descrever, por meio de modelos de regressão, o amadurecimento do fruto ao longo do tempo. Na literatura há trabalhos que já foram realizados para estudar esse recurso. Mendoza e Aguilera(2004) estudaram a análise de imagem para a classificação de maturação de bananas, Carvalho *et al.*(2014) verificaram que é possível o sistema de classificação da manga *Tommy* pela análise de cores da casca, Motta *et al.* (2015) mostraram ser possível estimar a maturação do mamão pela cor da casca e Pereira *et al.* (2018) estudaram a predição do amadurecimento do mamão por meio de imagens digitais, podendo concluir que é possível desenvolver um sistema automatizado baseado em imagens para classificar o mamão de acordo com os níveis de maturação.

Para se estudar os estágios de maturação são realizadas medições da característica de interesse no mesmo fruto por um período de tempo e dessa forma é esperado uma correlação entre estas medidas repetidas. Nesse contexto, a utilização de modelos lineares de efeitos mistos (parâmetros de efeitos fixos e de efeitos aleatórios) é uma alternativa metodológica apropriada, sendo possível estudar o comportamento médio de tais características por meio da parte fixa e a variação em cada fruto por meio da parte aleatória. Ademais, há flexibilidade de escolha de diferentes estruturas para as matrizes de variâncias e covariâncias dos efeitos aleatórios e dos erros, possibilitando a acomodação da variabilidade entre os frutos e assim obter um modelo que descreva os dados mais fielmente. Sendo assim, esse trabalho tem como motivação um estudo longitudinal buscando selecionar os melhores modelos de regressão que descrevam o comportamento da tonalidade e da intensidade da cor da casca do mamão da espécie *Carica papaya* L. O objetivo é descrever as curvas de amadurecimento individual dos frutos utilizando a abordagem de modelos mistos.

# 2 REVISÃO DE LITERATURA

#### 2.1 Análise de cor da casca por meio de imagens digitais

Algumas variáveis associadas a parâmetros físicos podem ser utilizadas para classificação dos estágios de maturação dos frutos, isto é, características da casca podem estar relacionadas com a polpa da fruta (Motta *et. al*, 2015). Manga, mamão e banana são exemplos de frutas com características em suas cascas que se modificam gradualmente no período de amadurecimento, como a tonalidade da casca e a intensidade da cor. Para identificar os estágios de amadurecimento do fruto de acordo com a cor da casca, um dos métodos é por meio de análise de imagem digital (Mendoza; Aguilera, 2004).

A aparência dos alimentos é de grande importância e influencia a escolha dos consumidores. A tonalidade e intensidade da cor são exemplos de características principais da casca que são observadas, pois estão associados ao frescor e qualidade de grande parte das frutas.

O scanner é um exemplo de equipamento utilizado para obter imagens da casca dos frutos. Por meio dele, consegue-se obter imagem de toda a superfície da fruta. O sistema de cor utilizado pelo hardware é, de forma geral, o RGB. No entanto, o espaço RGB possui cores que não são perceptíveis pelo cérebro humano. Assim, é necessário trabalhar com um espaço de cores que a visão humana consiga identificar. Visando minimizar esse problema, a CIE elaborou, dentre seus sistemas de cores, o L\*a\*b\* e L\*C\*h.

No sistema L\*a\*b\* tem-se 3 eixos (usam-se coordenadas retangulares), L\* refere-se à luminosidade da amostra, com valores que variam de 0 (representando o preto) a 100 (representando o branco), no eixo a as cores variam de vermelhas a verdes e no eixo b a variação é entre a cor amarela e azul, como mostra a Figura 2.1. Já o sistema  $L^*C^*h$  apesar de usar o mesmo diagrama que o espaço de cor  $L^*a^*b^*$ , não usa coordenadas retangulares e sim coordenadas cilíndricas.

Nesse contexto, a utilização de *softwares* como estratégia de análise baseada em imagens possibilita obter informações das cores presentes na casca, sendo possível estudar a relação dessas características com os estágios de maturação e descrever curvas de amadurecimento individual, bem como curva de amadurecimento médio dos frutos.



Figura 2.1: Região das cores pertencentes ao sistema de cores CIELab. Fonte: Konica Minolta

#### 2.2 Amadurecimento e características fisiológicas do mamão

Em algumas frutas é possível saber a qualidade de sua polpa quanto à firmeza por meio das características expressas na casca como acontece com o mamão (Motta *et al.*, 2015). O mamão tem maturação relativamente rápida e apresenta um nível significativo de perda de frutos na fase pós-colheita (Balbino,1997). Pode-se atribuir a perda pós-colheita ocasionada por um conjunto de fatores, como manuseio e transporte inadequado, desenvolvimento de doenças nas frutas e maturação acelerada causada pela temperatura elevada.

Visando à comercialização nacional e internacional, é importante saber determinar o estágio de maturação que se encontra os frutos para que se possa classificá-los e separá-los de forma mais homogênea. As possíveis classificações do estágio de maturação do mamão estão apresentadas na Tabela 2.1.

Estágio	Casca					
0	100%da superfície verde					
1	até 15% da superfície amarela					
2	até 25% da superfície amarela					
3	até 50% da superfície amarela					
4	50%à $75%$ da superfície amarela					
5	75%à 100% da superfície amarela					
Fonte: SEBRAE (2018).						

Tabela 2.1: Classificação dos estágios de maturação do mamão Carica papaya L.

O momento mais adequado para a colheita varia de acordo com a destinação que se dará aos frutos, por exemplo, para comercialização e consumo locais, os frutos podem ser colhidos quando apresentarem 50% de coloração amarela (estágio 3). Já os frutos destinados à exportação ou à armazenagem por períodos longos, são preferíveis no estágio 2 de maturação (Embrapa Mandioca e Fruticultura Tropical, 2009).

De acordo com Oliveira, Vianni e Araújo (2002) a utilização de parâmetros físicos para classificação da cor dos estágios de maturação do mamão abre a possibilidade da automatização do procedimento pela utilização de equipamentos com sensores óticos, que poderia aumentar a eficiência da seleção.

#### 2.3 Modelo de regressão linear

Modelos de regressão são usados na descrição do comportamento de uma variável aleatória Y como função de variáveis explanatórias  $\mathbf{X} = (X_1, X_2, \cdots, X_k)$ , com a seguinte estrutura:

$$Y = f(X_1, X_2, \dots, X_k, \boldsymbol{\beta}) + \epsilon \tag{2.1}$$

sendo que  $f(X_1, X_2, \ldots, X_k, \beta)$  representa a informação que  $X_1, X_2, \ldots, X_k$  fornece sobre Y, o qual  $\beta$  é o vetor de parâmetros desconhecidos e de interesse e  $\epsilon$  é o erro aleatório.

Os modelos de regressão linear são caracterizados por sua linearidade nos parâmetros  $\beta$ , isto é, uma função linear nos parâmetros.

Para uma amostra de n observações, o modelo de regressão linear múltipla pode ser expresso por:

$$Y_{i} = \beta_{0} + \beta_{1} X_{i1} + \dots + \beta_{k} X_{ik} + \epsilon_{i}, \quad i = 1, 2, \cdots, n$$
(2.2)

em que  $\beta_0, \beta_1, \dots, \beta_k$  são os coeficientes de regresssão (p=k+1 parâmetros),  $X_{ij}$   $(j = 1, \dots, k)$ são as variáveis explanatórias e  $\epsilon_i$  o erro aleatório para o qual supõe  $\epsilon_i \sim N(0, \sigma^2)$ . Matricialmente, tem-se

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1k} \\ 1 & X_{21} & X_{22} & \cdots & X_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$
(2.3)

ou seja, na forma reduzida:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{2.4}$$

em que **Y** é vetor de observações de dimensão  $n \times 1$ , **X** é matriz do delineamento de dimensão  $n \times p$ ,  $\beta$  é vetor de parâmetros de dimensão  $p \times 1$  e  $\epsilon$  é vetor de erros aleatórios de dimensão  $n \times 1$ .

Em particular, se a variável aleatória Y é função de uma variável explanatória X, tem-se o modelo de regressão linear simples:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i, \quad i = 1, \cdots, n.$$

$$(2.5)$$

em que  $\beta_0$  é o intercepto,  $\beta_1$  é o coeficiente de regressão e  $\epsilon_i$  é a variação aleatória em  $Y_i$ .

Assumindo  $E(\boldsymbol{\epsilon}) = \mathbf{0}$  e Var $(\boldsymbol{\epsilon}) = \sigma^2 \boldsymbol{I}$ , sendo que  $\boldsymbol{I}$  é a matriz identidade de dimensão  $n \times n$ , tem-se o valor esperado e a variância do modelo (2.4) dado, respectivamente, por:

$$E(\mathbf{Y}) = E(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon})$$
$$= E(\mathbf{X}\boldsymbol{\beta}) + E(\boldsymbol{\epsilon})$$
$$= \mathbf{X}\boldsymbol{\beta}$$

е

$$Var(\boldsymbol{Y}) = Var(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon})$$
$$= Var(\boldsymbol{X}\boldsymbol{\beta}) + Var(\boldsymbol{\epsilon})$$
$$= \sigma^{2}\boldsymbol{I}.$$

Portanto,

$$\boldsymbol{Y} \sim \mathbf{N}(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I}). \tag{2.6}$$

#### 2.3.1 Estimação

De acordo com Searle (1997), o método dos mínimos quadrados ordinários é umas das técnicas usuais para estimação dos parâmetros de modelos lineares. Em particular para o modelo

de regressão (2.4), o objetivo é minimizar a soma de quadrados:

$$S(\beta) = \sum_{i=1}^{n} \epsilon_i^2 \tag{2.7}$$

$$= \epsilon^T \epsilon \tag{2.8}$$

$$= (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$$
(2.9)

em que  $\epsilon^T$  é o vetor de erros aleatórios transposto. Em seguida, fazendo a derivada parcial de (2.9) com relação aos parâmetros,  $\beta$ , e igualando a zero, obtém-se o sistema de equações normais:

$$\mathbf{X}^T \mathbf{X} \hat{\boldsymbol{\beta}} = \mathbf{X}^T \mathbf{Y} \tag{2.10}$$

Então, a menos de multicolinearidade, o estimador para modelo de regressão linear é a solução única de (2.10):

$$\widehat{\boldsymbol{\beta}} = \left( \mathbf{X}^T \mathbf{X} \right)^{-1} \mathbf{X}^T \mathbf{Y}$$
(2.11)

coincidindo com o estimador de máxima verossimilhança. No caso de  $(\mathbf{X}^T \mathbf{X})^{-1}$  não existir, pode-se utilizar uma inversa generalizada  $(\mathbf{X}^T \mathbf{X})^G$  em (2.11), sendo a inversa de Moore-Penrose a de melhores propriedades. De acordo com Iemma (1987), adicionalmente tem-se

$$\widehat{\boldsymbol{Y}} = \mathbf{X}\widehat{\boldsymbol{eta}} = \mathbf{X} ig(\mathbf{X}^T\mathbf{X}ig)^{-1}\mathbf{X}^T\mathbf{Y} = \mathbf{H}\mathbf{Y}$$

que é a projeção ortogonal de  $\mathbf{Y}$  no espaço coluna da matrix  $\mathbf{X}$ , sendo  $\mathbf{H}$  a matriz de projeção, simétrica e idempotente.

Do ponto de vista da Estatística Matemática, os estimadores obtidos são normalmente distribuídos, não viesados e de variância mínima (dentre todos os estimadores possíveis) (Mood; Graybill; Boes, 1963). As expressões do valor esperado e da variância desses estimadores é descrita por Searle (1997) como:

$$E(\widehat{\boldsymbol{\beta}}) = E((\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y})$$
  
=  $(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T E(\mathbf{Y})$   
=  $(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{X} \boldsymbol{\beta}$   
=  $\boldsymbol{\beta}$ 

 $\mathbf{e}$ 

$$\operatorname{Var}(\widehat{\boldsymbol{\beta}}) = \operatorname{Var}((\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y})$$
  
=  $(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \operatorname{Var}(\mathbf{Y}) \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1}$   
=  $(\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^T \sigma^2 \mathbf{I} \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1}$   
=  $(\mathbf{X}^T \mathbf{X})^{-1} \sigma^2 \mathbf{I}.$ 

Dado que  $\mathbf{Y} \sim \mathbf{N}(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I})$  e que  $\widehat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$ , segue que  $\widehat{\boldsymbol{\beta}}$  é combinação linear de distribuições normais e, portanto:

$$\widehat{\boldsymbol{\beta}} \sim \mathbf{N}(\boldsymbol{\beta}, (\mathbf{X}^T \mathbf{X})^{-1} \sigma^2 \mathbf{I}).$$

Consequentemente:  $\hat{\boldsymbol{\beta}}_j \sim N(\beta_j, \sigma^2 C_{jj})$  em que  $C_{jj}$  é o j-ésimo elemento da diagonal da matriz  $(\mathbf{X}^T \mathbf{X})^{-1}$ .

#### 2.3.2 Testes de hipóteses e seleção de modelos

Para testar a significância da regressão, primeiramente considere a partição da soma de quadrados totais:

$$\underbrace{\sum_{i=1}^{n} (Y_i - \overline{Y})^2}_{\text{SQT}} = \underbrace{\sum_{i=1}^{n} (\widehat{Y}_i - \overline{Y})^2}_{\text{SQReg}} + \underbrace{\sum_{i=1}^{n} (Y_i - \widehat{Y}_i)^2}_{\text{SQRes}},$$
(2.12)

em que  $\sum_{i=1}^{n} (Y_i - \overline{Y})^2$  é a soma de quadrados total (SQT),  $\sum_{i=1}^{n} (\widehat{Y}_i - \overline{Y})^2$  é a soma de quadrados da regressão (SQReg) e  $\sum_{i=1}^{n} (Y_i - \widehat{Y}_i)^2$  é a soma de quadrados dos resíduos (SQRes). A estas somas de quadrados associam-se (n-1), (p-1) e (n-p) graus de liberdade, respectivamente.

A proporção da variação de Y que é explicada pelo modelo de regressão linear pode ser mensurada por meio do Coeficiente de Determinação,  $R^2$ :

$$R^2 = \frac{SQReg}{SQT},$$
(2.13)

sendo uma medida tal que  $0 < \mathbb{R}^2 < 1$ . O Coeficiente de Determinação ajustado é dado por

$$R^{2}{}_{aj} = 1 - \frac{n-1}{n-p}(1-R^{2}).$$
(2.14)

De acordo com Montgomery (2012) é preferível utilizar o coeficiente de determinação ajustado  $(R_{aj}^2)$ , pois incorpora os graus de liberdade referente ao número de fatores do modelo e é uma medida descritiva que pode ser cautelosamente usada na seleção de modelos. O teste global a respeito do modelo de regressão múltipla, refere-se às hipóteses:

$$H_0: \beta_1 = \cdots = \beta_k = 0$$
  
 $H_a: \beta_j \neq 0$ , para pelo menos algum j  $(1 \le j \le k)$ .

Essas hipóteses são avaliadas por meio da estatística:

$$F_0 = \frac{\text{SQReg}/k}{\text{SQRes}/(n-p)} \sim F_{k,n-p}$$
(2.15)

em que rejeita-se a hipótese  $H_0$ , a um nível  $\alpha$  de significância (usualmente  $\alpha = 0,05$ ), se  $F_0 > F_{k,n-p}$ . Nesse caso, conclui-se que pelo menos uma variável explanatória  $X_j$ ,  $j = 1, \dots, k$ , contribui significativamente para o modelo (Montgomery *et al.*, 2012).

Uma vez que foi verificado a significância do modelo pelo teste F, podem ser construídos testes para parâmetros individuais ou para subconjuntos deles, bem como intervalos de confiança. Considerando parâmetros individuais, pode-se testar às hipóteses:

$$H_0: \beta_j = 0$$
$$H_a: \beta_j \neq 0 \quad (j = 1, \cdots, k)$$

para as quais, pode-se usar a estatística

$$t_0 = \frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{\sigma}^2 C_{jj}}} = \frac{\hat{\beta}_j}{\sqrt{\hat{\sigma}^2 C_{jj}}} \sim t_{n-p-1}$$
(2.16)

em que  $\hat{\sigma}^2$  é o estimador não viesado da variância dos erros,  $\sigma^2$ , dado por

$$\widehat{\sigma^2} = \sqrt{\text{QMRes}}.$$

Na estatística (2.16) rejeita-se  $H_0$  se  $|t_0| > t_{(1-\alpha/2,n-p-1)}$ , indicando que a variável explanatória  $X_j$  contribui significativamente para o modelo.

Logo, pode-se obter o intervalo de confiança para  $\beta_j$ , a um nível  $\alpha = 0,05$  de significância:

$$\widehat{\beta}_j \pm t_{(\alpha/2,n-p-1)} \sqrt{\widehat{\sigma^2}} C_{jj}.$$

#### 2.4 Modelo de regressão linear misto

Modelos lineares mistos constituem-se em uma classe de modelos que podem ser empregados às situações em que se tem observações feitas no mesmo indivíduo, como é clássico dos estudos longitudinais no tempo ou espaço. Nestes estudos as  $n_i$  observações repetidas no i-ésimo indivíduo constituem um cluster (West; Welcki; Galecki, 2007) e espera-se que exista uma correlação entre essas observações (intra-unidade).

Os modelos mistos, nestas situações, são uma alternativa metodológica pois permitem levar em conta a correlação intra-unidade ao considerar na parte estocástica do modelo efeitos aleatórios pertinentes (Lara *et al.*, 2009). Além disso, permite considerar variadas estruturas para a matriz de variâncias e covariâncias, buscando representar ao máximo a variabilidade dos dados. Tais modelos são compostos por duas partes a saber: parte fixa que descreve o comportamento médio dos indivíduos e a parte aleatória que acomoda padrões individuais.

Seja  $\mathbf{Y}_i$  um vetor de  $n_i$  observações associado ao i-ésimo indivíduo,  $i = 1, \ldots, n$ , o modelo linear misto proposto por Lair e Ware (1982) satisfaz:

$$\mathbf{Y}_{i} = \overbrace{\mathbf{X}_{i}\beta}^{\text{fixo}} + \overbrace{\mathbf{Z}_{i}\mathbf{b}_{i} + \boldsymbol{\epsilon}_{i}}^{\text{aleatório}}$$
(2.17)

em que  $\mathbf{Y}_i$  é vetor  $(n_i \times 1)$  de respostas do i-ésimo indivíduo;  $\mathbf{X}_i$  é matriz  $(n_i \times p)$  de delineamento correspondente ao vetor de parâmetros de efeitos fixos;  $\boldsymbol{\beta}$  é vetor  $(p \times 1)$  de parâmetros desconhecidos de efeitos fixos;  $\mathbf{Z}_i$  é matriz  $(n_i \times q)$  de delineamento correspondente ao vetor de parâmetros de efeitos aleatórios;  $\mathbf{b}_i$  é vetor  $(q \times 1)$  de efeitos aleatórios que segue uma distribuição  $\mathbf{b}_i \sim \mathbf{N}(\mathbf{0}, \mathbf{D})$ ;  $\boldsymbol{\epsilon}_i$  é vetor  $(n_i \times 1)$  de erros aleatórios que segue uma distribuição  $\boldsymbol{\epsilon}_i \sim$  $\mathbf{N}(\mathbf{0}, \boldsymbol{\Sigma}_i)$ .

Logo, os vetores e matrizes em (2.17) para o i-ésimo indivíduo são:

$$\mathbf{Y}_{i} = \begin{bmatrix} \mathbf{Y}_{i1} \\ \mathbf{Y}_{i2} \\ \vdots \\ \mathbf{Y}_{in_{i}} \end{bmatrix}, \quad \mathbf{X}_{i} = \begin{bmatrix} 1 & X_{i11} & \dots & X_{ik1} \\ 1 & X_{i12} & \dots & X_{ik2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{i1n_{i}} & \dots & X_{ikn_{i}} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_{0} \\ \beta_{1} \\ \vdots \\ \beta_{k} \end{bmatrix}$$
$$\mathbf{Z}_{i} = \begin{bmatrix} Z_{i11} & Z_{i21} & \dots & Z_{iq1} \\ Z_{i12} & Z_{i22} & \dots & Z_{iq2} \\ \vdots & \vdots & \ddots & \vdots \\ Z_{i1n_{i}} & Z_{i2n_{i}} & \dots & Z_{iqn_{i}} \end{bmatrix}, \quad \mathbf{b}_{i} = \begin{bmatrix} b_{i1} \\ b_{i2} \\ \vdots \\ b_{iq} \end{bmatrix}, \quad \boldsymbol{\epsilon}_{i} = \begin{bmatrix} \epsilon_{i1} \\ \epsilon_{i2} \\ \vdots \\ \epsilon_{in_{i}} \end{bmatrix}$$

em que tem-se p=k+1 parâmetros de efeitos fixos, q parâmetros de efeitos aleatórios e  $b_1, \ldots, b_n$ e  $\epsilon_1, \ldots, \epsilon_n$  são independentes.

Considerando n indivíduos, o modelo (2.17) pode ser escrito compactamente como (West; Welcki; Galecki, 2007):

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\epsilon} \tag{2.18}$$

em que

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_n \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_n \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}, \quad \mathbf{Z} = \begin{bmatrix} \mathbf{Z}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{Z}_n \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_n \end{bmatrix}, \quad \boldsymbol{\epsilon} = \begin{bmatrix} \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon}_2 \\ \vdots \\ \boldsymbol{\epsilon}_n \end{bmatrix}.$$

com N=  $\sum_{i=1}^{n} n_i$  o total de observações, **Y** é vetor (N × 1) contendo as respostas dos n indivíduos obtido por meio do empilhamento dos vetores **Y**<sub>i</sub>; **X** é matriz (N × p) obtida por meio do empilhamento de **X**<sub>i</sub>;  $\beta$  é vetor (p × 1) de parâmetros de efeito fixos; **Z** é matriz (N × nq) bloco-diagonal composta com blocos das matrizes **Z**<sub>i</sub> na diagonal; **b** é vetor (nq × 1) que empilha todos os **b**<sub>i</sub>;  $\epsilon$  é vetor (N × 1) de erros aleatórios.

As matrizes de variâncias e covariâncias dos efeitos aleatórios do modelo (2.17), **D** e  $\Sigma_i$ , são matrizes simétricas e positivas definidas. **D** é matriz de dimensão  $q \times q$  e  $\Sigma_i$  é matriz de dimensão  $n_i \times n_i$ , assim são especificadas por:

$$\mathbf{D} = \operatorname{Var}(\mathbf{b}_i) = \begin{bmatrix} \operatorname{Var}(b_{i1}) & \operatorname{Cov}(b_{i1}, b_{i2}) & \cdots & \operatorname{Cov}(b_{i1}, b_{iq}) \\ & \operatorname{Var}(b_{i2}) & \cdots & \operatorname{Cov}(b_{i2}, b_{iq}) \\ & & \ddots & \vdots \\ & & \operatorname{Var}(b_{iq}) \end{bmatrix},$$
$$\mathbf{\Sigma}_i = \operatorname{Var}(\boldsymbol{\epsilon}_i) = \begin{bmatrix} \operatorname{Var}(\boldsymbol{\epsilon}_{i1}) & \operatorname{Cov}(\boldsymbol{\epsilon}_{i1}, \boldsymbol{\epsilon}_{i2}) & \cdots & \operatorname{Cov}(\boldsymbol{\epsilon}_{i1}, \boldsymbol{\epsilon}_{in_i}) \\ & \operatorname{Var}(\boldsymbol{\epsilon}_{i2}) & \cdots & \operatorname{Cov}(\boldsymbol{\epsilon}_{i2}, \boldsymbol{\epsilon}_{in_i}) \\ & & \ddots & \vdots \\ & & & \operatorname{Var}(\boldsymbol{\epsilon}_{in_i}) \end{bmatrix}$$

Supondo que **b** e  $\epsilon$  são normalmente distribuídos e não correlacionados, com vetor de médias **0** e as respectivas matrizes de variâncias e covariâncias D e  $\Sigma_i$ , obtém-se que

$$Var(\boldsymbol{Y}_i) = Var(\boldsymbol{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i)$$
(2.19)

$$= \operatorname{Var}(\boldsymbol{X}_{i}\boldsymbol{\beta}) + \operatorname{Var}(\boldsymbol{Z}_{i}\mathbf{b}_{i}) + \operatorname{Var}(\boldsymbol{\epsilon}_{i})$$
(2.20)

$$= \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T + \mathbf{\Sigma}_i \tag{2.21}$$

$$= \mathbf{V}_i \tag{2.22}$$

e como

$$E(\boldsymbol{Y}_i) = E(\mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i) = \boldsymbol{X}_i\boldsymbol{\beta}$$

segue que o modelo marginal é dado por:

$$\boldsymbol{Y}_i \sim N\left(\boldsymbol{X}_i\boldsymbol{\beta}, \mathbf{V}_i\right).$$

#### 2.4.1 Estimação

Na questão de estimação em modelos mistos, tem-se três aspectos fundamentais: a estimação e os testes de hipóteses sobre os efeitos fixos, estimação dos componentes de variância e a predição dos efeitos aleatórios (Marcelino; Iemma, 2000).

No que se refere à estimação dos parâmetros de efeitos fixos e dos componentes de variâncias e covariâncias, os métodos usuais são os de máxima verossimilhança (MV) e máxima verossimilhança restrita (MVR), ambos baseados na soma de quadrados dos resíduos. Davidian e Giltinan (1995) sugerem estimar os componentes de variância por MVR, pois corrige a perda dos graus de liberdade associados às estimativas de efeitos fixos. No caso da MV, por não corrigir essa perda, podem-se obter estimativas tendenciosas.

De acordo com Davidian e Giltinan (1995), a função de veros<br/>similhança por MV é dada por

$$\mathcal{L}(\boldsymbol{\theta}) = \prod_{i=1}^{n} (2\pi)^{-n_i/2} |\mathbf{V}_i|^{-1/2} \exp\left\{\frac{-(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta})^T \mathbf{V}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta})}{2}\right\},\tag{2.23}$$

em que  $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\alpha}^T)^T$  sendo que  $\boldsymbol{\beta}$  é o vetor de efeitos fixos do modelo (2.18) e  $\boldsymbol{\alpha}$ é o conjunto de todos os parâmetros dos componentes de variâncias e covariâncias. E para o método da MVR, a função de verossimilhança é expressa por:

$$L_{R}(\boldsymbol{\theta}) = \left\| \sum_{i=1}^{n} \mathbf{X}_{i}^{T} \mathbf{V}_{i}^{-1} \mathbf{X}_{i} \right\|_{i=1}^{-1/2} \prod_{i=1}^{n} (2\pi)^{-n_{i}/2} \left\| \mathbf{V}_{i} \right\|_{i=1}^{-1/2} \exp\left\{ \frac{-(\mathbf{Y}_{i} - \mathbf{X}_{i}\boldsymbol{\beta})^{T} \mathbf{V}_{i}^{-1} (\mathbf{Y}_{i} - \mathbf{X}_{i}\boldsymbol{\beta})}{2} \right\}$$
(2.24)

$$= \left|\sum_{i=1}^{n} \mathbf{X}_{i}^{T} \mathbf{V}_{i}^{-1} \mathbf{X}_{i}\right|^{-1/2} \mathbf{L}(\boldsymbol{\theta})$$

$$(2.25)$$

Maximizando as funções (2.23) e (2.25), com relação aos parâmetros desconhecidos, obtém-se um sistema de equações escore que tem por solução as estimativas de MV e MVR para os componentes de variâncias e covariâncias. Porém, estas equações não têm forma analítica fechada e são necessários processos iterativos para resolução. Dentre os algoritmos apresentados na literatura, pode-se citar o EM (*Expectation Maximum*), apresentado em Laird e Ware (1982) e também o algoritmo Newton-Rapson (Littell *et al.*, 2006).

Se os componentes de variâncias e covariâncias são conhecidos, tem-se que o vetor de parâmetros fixos é estimado por:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{Y}$$
(2.26)

em que  $\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1})$ . Para um caso geral, em que não se conhecem os componentes de variâncias e covariâncias, tem-se:

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^T \hat{\boldsymbol{V}}^{-1} \mathbf{X}\right)^{-1} \left(\mathbf{X}^T \hat{\boldsymbol{V}}^{-1} \mathbf{Y}\right)$$
(2.27)

em que  $\hat{V} = (\mathbf{Z}\hat{\mathbf{D}}\mathbf{Z}^T + \hat{\boldsymbol{\Sigma}}).$ 

Para os efeitos aleatórios, a predição é dada por:

$$\tilde{\mathbf{b}} = \mathbf{D}\mathbf{Z}^T \mathbf{V}^{-1} (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}).$$
(2.28)

Henderson *et. al* (1959) mostraram uma maneira de se obter a estimação de  $\hat{\beta}$  e a predição de  $\hat{b}$  simultaneamente, resolvendo o sistema de equações:

$$\begin{bmatrix} \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X} & \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{Z} \\ \mathbf{Z}^T \boldsymbol{\Sigma}^{-1} \mathbf{X} & \mathbf{Z}^T \boldsymbol{\Sigma}^{-1} \mathbf{Z} + \mathbf{D}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \tilde{\mathbf{b}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{Y} \\ \mathbf{Z}^T \boldsymbol{\Sigma}^{-1} \mathbf{Y} \end{bmatrix}$$

que tem como solução

$$\begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \tilde{\mathbf{b}} \end{bmatrix} = \begin{bmatrix} (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{Y} \\ \mathbf{D} \mathbf{Z}^T \mathbf{V}^{-1} \left( \mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}} \right) \end{bmatrix}$$

Como propriedade dos estimadores, tem-se que  $\hat{\boldsymbol{\beta}}$  é o melhor estimador linear não viesado (*BLUE* - *Best Linear Unbiased Estimator*) de  $\boldsymbol{\beta}$  e  $\hat{\mathbf{b}}$  é a melhor predição linear não viesada (*BLUP* - *Best Linear Unbiased Predictor*) de  $\mathbf{b}$  (Perri; Iemma, 1999).

As estimativas por MV para os componentes de variâncias são sempre não negativas, mas são viesadas, uma vez que o método não considera a perda dos graus de liberdade, resultante da estimação dos efeitos fixos do modelo. Já as estimativas obtidas pelo método MVR são não viesadas para dados balanceados, porém para dados desbalanceados as estimativas dos componentes de variância são viesadas, embora o método considere a perda dos graus de liberdade devida ao ajuste dos efeitos aleatórios (Marcelino; Iemma, 2000).

#### 2.4.2 Estrutura das matrizes de variâncias e covariâncias

Como já descrito anteriormente, a estimação dos componentes de variâncias depende do método utilizado (MV ou MVR). Porém o processo de estimação também dependente da escolha da estrutura das matrizes  $\mathbf{D} \in \mathbf{\Sigma}_i$ , uma vez que diferentes números de variâncias e covariâncias devem ser estimados (Perri; Iemma, 1999). Essas matrizes podem assumir diversas estruturas, pois nos modelos mistos os resíduos e os efeitos aleatórios podem apresentar variâncias que não sejam constantes ou observações que não sejam independentes.

Verbeke e Molenberghs (2009) ressaltam que o uso de uma estrutura de variâncias e covariâncias adequada é importante na interpretação da variação aleatória nos dados e é essencial para fazer inferências para os parâmetros. Assim, ao ajustar modelos mistos aos dados, procuram-se matrizes de variâncias e covariâncias que melhor descrevam as relações entre os indivíduos e entre as observações.

Na literatura, como por exemplo em Pinheiro e Bates (2000), Littel *et al.* (2006) e Alcarde (2012), são encontradas várias especificações para as estruturas de variâncias e covariâncias. A título de ilustração, considerando 4 observações por indivíduo, pode-se citar:

1) Componentes de Variância: variâncias são iguais e covariâncias nulas.

$$\begin{bmatrix} \sigma^2 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 \\ 0 & 0 & \sigma^2 & 0 \\ 0 & 0 & 0 & \sigma^2 \end{bmatrix} = \sigma^2 \mathbf{I}_4$$

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

2) Componentes de Variância com Heterogeneidade (CVH): admite-se variâncias diferentes e covariâncias nulas.

$$\begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & \sigma_3^2 & 0 \\ 0 & 0 & 0 & \sigma_4^2 \end{bmatrix}$$

 Simetria Composta(CS): variâncias homogêneas e covariâncias constantes entre quaisquer observações.

$$\begin{bmatrix} \sigma^{2} + \sigma_{1}^{2} & \sigma_{1}^{2} & \sigma_{1}^{2} & \sigma_{1}^{2} \\ \sigma_{1}^{2} & \sigma^{2} + \sigma_{1}^{2} & \sigma_{1}^{2} & \sigma_{1}^{2} \\ \sigma_{1}^{2} & \sigma_{1}^{2} & \sigma^{2} + \sigma_{1}^{2} & \sigma_{1}^{2} \\ \sigma_{1}^{2} & \sigma_{1}^{2} & \sigma_{1}^{2} & \sigma^{2} + \sigma_{1}^{2} \end{bmatrix} = \sigma^{2} \mathbf{I}_{4} + \sigma_{1}^{2} \mathbf{J}_{4}$$

em que  $\mathbf{J}$  é uma matriz de uns.

4) Não estruturada(UN): todas as variâncias e covariâncias podem ser desiguais, isto é, especifica uma matriz geral.

$$\begin{bmatrix} \sigma_{11} & \sigma_{21} & \sigma_{31} & \sigma_{41} \\ \sigma_{21} & \sigma_{22} & \sigma_{32} & \sigma_{42} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} & \sigma_{43} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_{44} \end{bmatrix}$$

5) Autorregressiva de primeira ordem (AR(1)): variâncias homogêneas e correlações que diminuem exponencialmente à medida em que se aumenta o intervalo de tempo entre as medidas repetidas. Nessa estrutura, as medidas repetidas precisam ser igualmente espaçadas.

$$\begin{bmatrix} \sigma^2 & \sigma^2 \rho & \sigma^2 \rho^2 & \sigma^2 \rho^3 \\ \sigma^2 \rho & \sigma^2 & \sigma^2 \rho & \sigma^2 \rho^2 \\ \sigma^2 \rho^2 & \sigma^2 \rho & \sigma^2 & \sigma^2 \rho \\ \sigma^2 \rho^3 & \sigma^2 \rho^2 & \sigma^2 \rho & \sigma^2 \end{bmatrix}$$

em que  $\rho$  é o parâmetro autorregressivo, com  $|\rho| < 1$ . Essa estrutura admite assumir variâncias heterogêneas, nesse caso tem-se uma estrutura autorregressiva heterogênea de primeira ordem (ARH(1)).

6) Autorregressiva heterogênea de primeira ordem(ARH(1)): Variâncias heterogêneas e correlações que diminuem exponencialmente à medida em que aumenta o intervalo de tempo entre as observações.

26

7) Bloco diagonal: faz-se combinações de estruturas, anteriormente descritas por exemplo, de modo que obtenha-se relações em bloco na diagonal.

$$\begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & \sigma_2 \sigma_3 & \sigma_2 \sigma_4 \\ 0 & \sigma_2 \sigma_3 & \sigma_3^2 & \sigma_3 \sigma_4 \\ 0 & \sigma_2 \sigma_4 & \sigma_3 \sigma_4 & \sigma_4^2 \end{bmatrix}$$

#### 2.4.3 Seleção de modelos

Na seleção de modelos o objetivo é encontrar, quando possível, um modelo que se ajusta bem aos dados e que possui o menor número de parâmetros, ou seja, um modelo parcimonioso.

Primeiramente se especifica o conjunto de modelos que podem ser considerados candidatos a descrever melhor os dados, e então do maximal ao minimal é feita a seleção de modelo intermediário com o auxílio de avaliadores de qualidade de ajuste e testes de hipóteses (Burnham; Anderson, 1998).

Dentre as opções de avaliadores de qualidade de ajuste que auxiliam na seleção de modelos podem-se citar o Critério de Informação de Akaike (AIC), Critério de Informação Bayesiano (BIC) e Teste da razão de verossimilhanças (TRV). Segundo Pinheiro e Bates (2000), quando os modelos são ajustados por máxima verossimilhança restrita pode-se comparar apenas modelos com a mesma estrutura dos efeitos fixos.

#### 2.4.3.1 Critério de informação de Akaike (AIC)

O AIC, proposto por Akaike (1974), é uma medida baseada no logaritmo da função de verossimilhança do modelo ajustado que leva em consideração o número de parâmetros envolvidos. Então o AIC é dado por:

$$AIC = -2 \log L + 2p \tag{2.29}$$

onde L é a função de máxima verossimilhança e p refere-se ao número de parâmetros do modelo. Considera-se o melhor modelo aquele que apresenta o menor valor de AIC.

#### 2.4.3.2 Critério de informação Bayesiano (BIC)

Schwarz (1978) apresentou uma abordagem Bayesiana ao AIC, o qual simplificadamente consiste em multiplicar o número de parâmetros por  $1/2 \log n$ . Então o Critério Bayesiano, BIC, é dado por:

BIC = 
$$-2 \log L + 2p \frac{1}{2} \log n = -2 \log L + p \log n$$
 (2.30)

onde L é a função de verossimilhança, p refere-se ao número de parâmetros e n é o número de observações. Como no Critério de Akaike, é considerado o melhor modelo quando apresenta menor valor de BIC.

Yamanouchi (2017) estudou o desempenho dos critérios AIC e BIC na seleção de modelos mistos aninhados, verificou que para seleção de efeitos fixos eles apresentaram melhores desempenhos quando o tamanho amostral foi maior. Já no estudo de seleção dos efeitos aleatórios e matriz de variâncias e covariâncias, o critério BIC apresentou melhor desempenho do que o critério AIC, sendo considerado o mais indicado nesse contexto.

#### 2.4.3.3 Teste da razão de verossimilhança (TRV)

Este teste baseia-se na comparação dos valores das funções de verossimilhança para dois modelos encaixados,  $M_1 \in M_2$  com  $p_1 \in p_2$  ( $p_1 < p_2$ ) parâmetros respectivamente, sendo  $M_1$  a versão restrita de  $M_2$  (West; Welch; Galecki, 2007). A estatística desse teste pode ser calculada para testar hipóteses sobre parâmetros de variâncias e covariâncias ou parâmetros de efeitos fixos, de forma que as hipóteses consideradas são:

H<sub>0</sub>: Modelo restrito  $(M_1)$  é preferível H<sub>a</sub>: Modelo completo  $(M_2)$  é preferível.

Considere  $L_1 \in L_2$  as funções de verossimilhança dos modelos  $M_1 \in M_2$ , respectivamente, então a estatística do teste da razão de verossimilhanças é dada por:

$$\text{TRV} = -2\log\left(\frac{L_1}{L_2}\right) = -2\left(\log L_1 - \log L_2\right) = 2\left(\log L_2 - \log L_1\right)$$
(2.31)

a qual segue, assintoticamente, uma distribuição qui-quadrado,  $\chi^2$ , com  $\gamma = p_2 - p_1$  graus de liberdade, isto é, os graus de liberdade correspondem à diferença entre os números de parâmetros dos modelos completo e restrito (Wilks, 1938). De acordo com Self e Liang (1987), para os casos em que os parâmetros estão na fronteira do espaço paramétrico tem-se que a estatística (2.31) segue uma mistura de  $\chi^2$ .

#### 2.4.4 Resíduos

Técnicas informais, como construção de gráficos, podem ser utilizadas para diagnósticos de resíduos, e assim verificar as pressuposições do modelo (West; Welch; Galecki, 2007). De acordo com Nobre e Singer (2007), devido a presença de efeitos aleatórios e pela flexibilidade de assumir diferentes estruturas de variâncias e covariâncias, em modelos mistos os seguintes resíduos podem ser usados para diagnósticos desses modelos:

1. Resíduos marginais: considerando o modelo marginal  $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ , obtém-se

$$\boldsymbol{\varepsilon} = \mathbf{Y} - \mathbf{X}\boldsymbol{\beta} \tag{2.32}$$

e no contexto de modelos misto (2.18) pode-se escrever  $\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{Z}\mathbf{b} + \boldsymbol{\epsilon}$ , ou seja

$$\hat{\boldsymbol{\varepsilon}} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{Z}\mathbf{b} + \boldsymbol{\epsilon}, \qquad (2.33)$$

que são os erros marginais preditos pelos resíduos marginais.

2. Resíduos condicionais: nesse caso, têm-se

$$\hat{\boldsymbol{\epsilon}} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{Z}\hat{\boldsymbol{b}},\tag{2.34}$$

que estima os erros condicionais

$$\boldsymbol{\epsilon} = \mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{b},\tag{2.35}$$

obtidos do modelo (2.18).

Por meio dos resíduos marginais pode-se verificar a linearidade dos efeitos, em que graficamente faz-se tais resíduos *versus* variáveis explanatórias e espera-se um comportamento aleatório em torno de zero no caso de ser linear (Nobre; Singer, 2007). Já os resíduos condicionais podem ser utilizados para verificar a homocedasticidade e normalidade dos erros condicionais e também para detectar valores discrepantes. West, Welch e Galecki (2007) afirmam que tais resíduos não são muito adequados pois os resíduos condicionais tendem a ser correlacionados e ter variâncias desiguais para diferentes subgrupos de indivíduos. Nesse caso, uma alternativa é considerar resíduos studentizados ou padronizados.

#### 3 MATERIAL

Os dados a serem utilizados neste trabalho referem-se à maturação do mamão (*Carica Papaya* L.) do cultivar "Sunrise Solo". O experimento foi conduzido no Departamento de Produção Vegetal na Escola Superior de Agricultura "Luiz de Queiroz", USP, em 2010/2011.

O estudo observacional envolveu 20 frutos, colhidos no mesmo dia e apresentando até 10% da casca na coloração amarela. Foram armazenados em câmara de refrigeração a 18 °C, com  $80\% \pm 5\%$  de umidade relativa.

Para a medição da coloração da casca o equipamento utilizado foi um *scanner* flat-bed (HP Scanjet G2410) com resolução de 200 *pixels* por polegada. Para ser possível analisar toda a superfície dos frutos, digitalizou-se 2 lados de cada mamão, Lado A e Lado B conforme descrito na Figura 3.1, que são faces opostas e que juntas formam a surperfície inteira. Para digitalizar a superfície do mamão, o *scanner* foi coberto com uma caixa de papelão revestida com tecido preto para que ficasse em evidência apenas a casca e fosse selecioná-la mais facilmente para a análise da coloração. Maiores detalhes desse experimento podem ser encontrados em Oliveira *et al.* (2016).



Figura 3.1: Lado A e Lado B de um mamão "Sunrise Solo" Papaya. Fonte: (Oliveira *et al.*,2016)

Neste trabalho consideram-se duas variáveis, a intensidade e a tonalidade da casca de 20 mamões. Em cada um dos frutos, foram obtidos os valores da tonalidade e intensidade do lado A e do lado B, calculando-se a média circular para a tonalidade e média aritmética para a intensidade, e considerou-se essa medida para a análise.

Frutos que apresentavam desenvolvimento de doenças foram descartados, acarretando em um número de avaliações diferente em cada fruto. O período de análise foi de 14 dias, totalizando em 277 observações para cada uma das variáveis.

A fim de identificar os padrões e características dos dados referentes às variáveis tonalidade e intensidade, apresenta-se uma análise exploratória que é base para definição de procedimentos metodológicos, definidos na seção 4. Sendo assim, estão apresentadas na Tabela 3.1 e na Tabela 3.2 as médias, por dia, referente à variável tonalidade e variável intensidade, respectivamente. Verifica-se que no período do experimento a tonalidade dos frutos, no geral, mudou de verde claro (tonalidade média de 112,05 do dia inicial) para amarelo (tonalidade média de 86,72 graus no último dia). No caso da intensidade observam-se que os valores são bem homogêneas ao longo dos dias, iniciando com média de 34,27 e atingindo 35 no último dia, indicando que as cascas dos frutos tiveram coloração mais escura nesse período de análise. Logo, isso mostra que há possivelmente um efeito do tempo sobre à tonalidade mas não sobre a intensidade, ou seja, a tonalidade da casca dos frutos sofre alterações no período de tempo enquanto que a intensidade não se tem alterações conforme o tempo.

Tabela 3.1: Médias e desvios padrões da variável tonalidade da casca do mamão ao longo do tempo no experimento referente à maturação do *Carica Papaya* L.

Dia	Média	Desvio padrão	Dia	Média	Desvio padrão
0	$112,\!05$	1,42	0	$34,\!27$	1,31
1	110,70	$1,\!85$	1	$34,\!56$	$1,\!42$
2	$109,\!38$	$2,\!37$	2	34,75	$1,\!43$
3	$107,\!13$	$2,\!43$	3	$35,\!03$	$1,\!47$
4	$106,\!08$	$2,\!58$	4	34,70	$1,\!59$
5	$104,\!53$	2,78	5	34,79	$1,\!97$
6	$102,\!88$	$3,\!31$	6	$34,\!24$	$1,\!62$
7	100,76	$3,\!32$	7	$34,\!61$	$1,\!51$
8	$98,\!42$	3,75	8	$34,\!90$	$1,\!92$
9	$96,\!12$	$3,\!78$	9	34,71	$1,\!50$
10	94,73	4,05	10	$34,\!59$	$1,\!46$
11	$92,\!40$	$4,\!22$	11	$34,\!66$	$1,\!23$
12	$91,\!01$	$4,\!51$	12	$35,\!03$	$1,\!22$
13	89,33	$4,\!58$	13	$35,\!16$	$0,\!99$
14	86,72	$4,\!68$	14	$35,\!00$	$0,\!83$

Fonte: (Oliveira et al., 2016).

Fonte: (Oliveira et al., 2016).

Tabela 3.2: Médias e desvios padrões da variá-

vel intensidade da casca do mamão ao longo do

tempo no experimento referente à maturação do

Carica Papaya L.

Na Tabela 3.3 tem-se da diagonal principal para cima (valores em negrito) as variâncias e covariâncias dos dados da tonalidade e abaixo dela estão as correlações entre as medidas tomadas no período de 0 à 14 dias. Observam-se que as medidas mais próximas uma das outras ao longo dos dias são mais correlacionadas. As variâncias aumentam com o passar dos dias, atingindo o maior valor no último dia de análise.

Tabela 3.3: Variâncias e covariâncias (em negrito) e correlações amostrais da variável tonalidade.

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
0	2,01	1,51	1,54	1,53	1,76	1,93	1,84	1,75	1,87	1,95	2,03	1,79	2,10	-1,01	0,08
1	0,57	3,46	3,78	3,82	4,16	4,32	$^{5,03}$	4,90	4,87	4,94	5,16	$0,\!62$	0,97	0,86	$2,\!48$
2	0,46	0,86	$^{5,65}$	4,84	$^{5,05}$	5,36	6,42	6,21	$6,\!65$	6, 49	$6,\!80$	0,60	1,82	-1,42	$2,\!68$
3	0,44	0,84	0,84	5,93	5,74	6, 17	7,32	7,34	7,83	$^{8,17}$	8,88	1,75	0,55	1,49	$4,\!42$
4	0,48	$^{0,87}$	0,82	0,91	6,70	7,07	$^{8,10}$	$^{8,19}$	$^{8,68}$	$^{8,78}$	9,42	2,41	2,19	2,51	$3,\!66$
5	0,49	0,84	0,81	0,91	0,98	7,73	8,69	8,78	9,33	9,46	10, 17	1,62	1,28	1,89	$3,\!63$
6	0,39	0,82	0,81	0,91	0,94	0,94	11,00	10,79	11,88	11,78	$12,\!67$	3,22	2,71	2,04	$^{3,05}$
7	0,37	0,79	0,79	0,91	0,95	0,95	0,98	11,06	12,21	12,09	13,06	$^{3,49}$	2,50	1,67	$^{3,13}$
8	0,35	0,70	0,75	0,86	0,89	0,89	0,95	0,98	14, 12	13,88	14,89	4,88	$^{3,27}$	$0,\!64$	1,93
9	0,36	0,70	0,72	0,89	0,90	0,90	0,94	0,96	0,98	14,29	15, 19	4,21	2,19	1,99	2,80
10	0,35	0,69	0,71	0,90	0,90	0,90	0,94	0,97	0,98	0,99	16,45	5,24	2,85	2,05	3,08
11	0,30	0,08	0,06	$^{0,17}$	$^{0,22}$	0,14	$^{0,23}$	$^{0,25}$	$^{0,31}$	$^{0,27}$	$^{0,31}$	$17,\!39$	15,42	$^{3,52}$	0,21
12	0,33	0,12	0,17	$^{0,05}$	0,19	0,10	0,18	$^{0,17}$	0,19	0,13	$^{0,15}$	0,81	20,72	0,31	$0,\!69$
13	-0,15	0,10	-0,13	$^{0,13}$	$^{0,21}$	0,15	$_{0,13}$	0,11	0,04	$^{0,11}$	$_{0,11}$	0,18	$^{0,01}$	$21,\!62$	$^{8,81}$
14	0,01	0,29	0,25	0,40	0,31	0,29	$^{0,20}$	$^{0,21}$	0,11	0,16	$^{0,17}$	0,01	$^{0,03}$	0,42	20,79

Analogamente, tem-se na Tabela 3.4 variâncias, covariâncias e correlações da intensidade. As variâncias tem valores variados, sem apresentar um padrão, atingindo seu maior valor (3,90) no quinto dia de experimento e menor valor (0,65) no último dia. Então, pode-se dizer que ao final do experimento a intensidade da coloração da casca estava com pouca variação entre um fruto e outro.

Tabela 3.4: Variâncias e covariâncias (em negrito) e correlações amostrais da variável intensidade.

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
0	1,72	1,49	1,24	1,01	0,90	1,06	1,04	1,02	0,24	0,18	0,23	-0,25	-0,45	0,23	-0,24
1	0,79	2,04	$1,\!39$	$0,\!60$	0,81	$0,\!82$	0,90	0,91	-0,75	-0,64	-0,47	-0,48	-0,61	0,28	-0,28
2	$0,\!66$	$0,\!68$	2,04	$0,\!98$	$1,\!37$	1,72	1,76	$1,\!61$	$0,\!48$	0,31	$0,\!43$	$0,\!10$	-0,25	$0,\!69$	-0,23
3	0,52	$0,\!29$	$0,\!46$	$2,\!18$	1,28	1,70	1,20	$1,\!37$	$0,\!61$	$0,\!22$	0,38	-0,02	0,08	0,31	$0,\!11$
4	$0,\!43$	0,36	$0,\!60$	$0,\!54$	$2,\!54$	$2,\!90$	$1,\!87$	1,85	0,91	$0,\!23$	0,82	-0,41	-0,54	$0,\!99$	-0,11
5	0,41	$0,\!29$	$0,\!61$	$0,\!58$	0,92	$3,\!90$	2,33	2,36	$1,\!19$	$0,\!25$	0,96	-0,30	-0,46	$1,\!15$	-0,15
6	$0,\!49$	$0,\!39$	0,76	0,50	0,72	0,73	$2,\!62$	2,26	1,23	0,60	0,88	$0,\!41$	0,21	1,10	$0,\!11$
7	0,51	$0,\!42$	0,75	$0,\!61$	0,77	0,79	0,92	2,29	0,93	$0,\!43$	$0,\!63$	0,11	0,07	1,06	0, 19
8	0,09	-0,27	$0,\!18$	$_{0,21}$	$0,\!30$	0,31	$^{0,40}$	0,32	$3,\!68$	2,38	$2,\!42$	0, 19	$0,\!20$	$0,\!40$	-0,45
9	0,09	-0,30	$0,\!15$	$_{0,10}$	$0,\!10$	$0,\!08$	0,24	$0,\!19$	0,83	$^{2,27}$	2,00	$0,\!47$	-0,08	0,09	-0,17
10	$0,\!12$	-0,22	$^{0,20}$	$0,\!18$	$0,\!35$	0,33	0,37	0,28	0,86	0,91	$2,\!14$	0,23	-0,38	0,21	-0,27
11 -	-0,16	-0,28	0,06	-0,01	-0,21	-0,13	$_{0,21}$	0,06	0,08	0,26	$_{0,13}$	$1,\!45$	$0,\!92$	$0,\!10$	0,03
12 -	-0,28	-0,35	-0,14	$0,\!05$	-0,28	-0,19	$0,\!11$	0,04	0,09	-0,04	-0,22	$0,\!63$	$1,\!47$	$0,\!17$	0,08
13	$0,\!18$	$0,\!20$	$0,\!50$	0,22	$0,\!64$	$0,\!60$	0,70	0,72	0,22	0,06	$0,\!15$	$0,\!08$	$_{0,15}$	0,94	$0,\!15$
14 -	-0,23	-0,24	-0,20	0,09	-0,09	-0,09	0,09	0,16	-0,29	-0,14	-0,23	0,03	0,08	0,20	$0,\!65$

Na Tabela 3.5 e na Tabela 3.6 tem-se o número de frutos por dia e o número de observações de cada fruto para cada variável analisada, respectivamente.

Tabela 3.5: Número de frutos analisados por dia no experimento referente à maturação do *Carica Papaya* L.

Tabela 3.6: Número total de observações por fruto para cada variável do experimento referente à maturação do *Carica Papaya* L.

	N.C	Fruto	Número de observações
Dia	Numero de frutos		15
0	20	$\overline{2}$	15
1	20	3	15
2	20	4	11
2	20	5	13
5	20	6	15
4	20	7	13
5	20	8	14
6	20	9	13
7	20	10	15
1	20	11	13
8	20	12	14
9	20	13	15
10	20	14	12
11	10	15	15
11	19	16	12
12	16	17	15
13	12	18	15
1/	10	19	12
14	10	20	15
Fonte	: (Oliveira $et al., 2016$ ).	Fonte	e: (Oliveira <i>et al.</i> , 2016).

Para verificar o comportamento de cada fruto, tem-se na Figura 3.2 os perfis individuais referente à variável resposta tonalidade média. Apesar da aparência inicial dos frutos ser igual, com poucas faixas na coloração amarela (máximo 20%), existe uma variabilidade do valor médio inicial da tonalidade que pode ser observada na Figura 3.2. Como exemplo, tem-se o fruto 4 com tonalidade média inicial acima de 110 (tons amarelados predominantes), enquanto o fruto 18 tem sua tonalidade média inicial abaixo de 110 (tons esverdeados predominantes).

Esses perfis apresentam comportamento decrescente conforme aumenta o tempo de maturação, com algumas curvas que decrescem mais rápido do que outras, demonstrando que existem diferentes velocidade de maturação entre esses frutos. Apesar de estarem em um ambiente controlado e com temperatura igual para todos, há frutos que apresentam maturação mais lenta do que outros. Isso se deve ao fato do processo de maturação envolver fontes não controláveis, como a composição química que pode afetar diretamente.

A maioria dos frutos apresenta observações com um comportamento linear, porém há observações que definem comportamento levemente curvilíneo. As curvas relacionadas aos frutos 16, 17, 19, 6 e 11, destacam-se por apresentar observações com comportamento diferente dos demais frutos.



Figura 3.2: Perfis individuais das tonalidades médias dos 20 frutos (*Carica Papaya* L.) medidas no período de 14 dias, no experimento referente à maturação dos mesmos, realizado em 2010/2011.

Para a variável intensidade, tem-se o gráfico de perfis individuais na Figura 3.3. Verificamse curvas com comportamento diferente em cada fruto, causado pela variabilidade dentro dos frutos, o que contribui para a variabilidade entre os frutos. Pode-se considerar que não houve grande variação dentro dos frutos, já que se manteve na faixa de 31 à 39.



Figura 3.3: Perfis individuais das intensidades médias dos 20 mamões (Carica Papaya L.) medidas no período de 14 dias, no experimento referente à maturação dos mesmos, realizado em 2010/2011.

## 4 MÉTODOS

Com base na análise exploratória dos dados apresentada na seção 3, observa-se para a variável tonalidade uma tendência linear nos perfis individuais (Figura 3.2), com alguns frutos apresentando curvatura, sinalizando uma tendência quadrática. Nesse caso tem-se como candidatos iniciais para modelos de efeitos fixos, o modelo linear e polinomial de grau 2.

Com relação à variável intensidade, observa-se pelo gráfico de perfis individuais (Figura 3.3) uma tendência polinomial de grau maior e mais heterogênea em relação à tonalidade. Nesse caso, são considerados como modelos de efeitos fixos o polinomial de grau 3 e polinomial de grau 4. Outros modelos de grau maior ou menor ou não lineares podem ser considerados, caso sejam necessários.

Dessa forma, a fim de estabelecer os modelos funcionais associados às variáveis respostas, são considerados modelos lineares polinomiais tendo como variável regressora o tempo:

$$Y = \beta_0 + \beta_1 t + \dots + \beta_k t^k \tag{4.1}$$

em que  $\beta_0, \beta_1, \dots, \beta_k$  são parâmetros de efeitos fixos e trepresenta o tempo. Como são analisadas duas variáveis resposta, usa-se Y como sendo a variável resposta referente à tonalidade e  $Y^*$  a variável resposta referente à intensidade. Na Tabela 4.1 tem-se o resumo dos modelos de efeitos fixos considerados.

Tabela 4.1: Modelos lineares de efeitos fixos considerados para as variáveis tonalidade (M1 e M2) e intensidade (M1 $^*$  e M2 $^*$ ).

Modelo	Expressão
M1	$Y = \beta_0 + \beta_1 t$
M2	$Y = \beta_0 + \beta_1 t + \beta_2 t^2$
$M1^*$	$Y^* = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3$
$M2^*$	$Y^* = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \beta_4 t^4.$

Afim de avaliar a necessidade de incluir efeitos aleatórios, usa-se o procedimento descrito por Pinheiro e Bates (2000), que sugerem construir um intervalo de confiança para cada parâmetro de efeito fixo, em um nível de confiança  $1 - \alpha = \gamma$ ,  $0 < \alpha < 1$ , pré especificado para cada indivíduo.

Na prática tem-se que se ao menos dois intervalos de confiança para  $\beta_{ji}$  não se sobrepuserem, pode-se concluir que há indícios da necessidade de efeito aleatório ao respectivo parâmetro. Uma vez identificado em quais parâmetros há necessidade de efeitos aleatórios, parte-se de um modelo completo (maximal) para cada variável resposta. A seguir apresentam-se os modelos maximais (M1.t e M1.i) para a variável tonalidade ( $Y_i$ ) e variável intensidade ( $Y_i^*$ ), respectivamente:

$$Y_{i} = (\beta_{0} + b_{0i}) + (\beta_{1} + b_{1i})t_{i} + (\beta_{2} + b_{2i})t_{i}^{2} + \epsilon_{i},$$

$$(4.2)$$

$$Y_i^* = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})t_i + (\beta_2 + b_{2i})t_i^2 + (\beta_3 + b_{2i})t_i^3 + \epsilon_i.$$
(4.3)

Partindo dos modelos (4.2) e (4.3), mantém-se os parâmetros de efeitos fixos e retirase gradativamente os parâmetros de efeitos aleatórios, obtendo submodelos. Assim, utiliza-se do teste da razão de verossimilhança na comparação do modelo maximal com os submodelos. Em seguida, faz-se analogamente o processo descrito anteriormente, porém mantendo os efeitos aleatórios que foram relevantes e retiram-se gradativamente os parâmetros de efeitos fixos.

Assume-se, inicialmente, estruturas de variâncias e covariâncias  $\Sigma_i = \sigma^2 \mathbf{I}$  e não estruturada para a matriz **D**. Após selecionado os efeitos significativos do modelo, encerra-se o processo de seleção de modelos e parte-se para a seleção das estruturas das matrizes de variâncias e covariâncias dos efeitos aleatórios e dos resíduos, **D** e  $\Sigma_i$  respectivamente. Utiliza-se teste da razão de verossimilhança na seleção da estrutura que melhor descreve a relação dos dados.

Para a matriz **D**, as estruturas avaliadas para a variável tonalidade foram a não estruturada (NE), componentes de variâncias (CV), componentes de variância heterogênea (CVH) e a matriz em bloco-diagonal (NE.CV) a qual foi obtida baseada em análise de intervalos de confiança para os componentes de variâncias e covariâncias. Analogamente para a variável intensidade, porém as matrizes CV e CVH não foram consideradas.

Para a matriz  $\Sigma_i$ , para a variável tonalidade foram consideradas as estruturas componentes de variâncias com heterogeneidade (CVH) e autorregressiva heterogênea de ordem 1 (ARH1). No caso da variável intensidade, consideraram-se componentes de variâncias (CV), componentes de variâncias com heterogeneidade (CVH), autorregressiva heterogênea de ordem 1 (ARH1).

Faz-se o uso de gráficos de dispersão para os resíduos condicionais (conforme descrito na seção 2) e também gráficos de quantil-quantil com envelope simulado para verificar a normalidade dos parâmetros de efeitos aleatórios e dos resíduos. As análises são feitas no *software* R (R Core Team, 2018) por meio dos pacotes **lme4** (Bates *et al.*, 2019) e **nlme** (Pinheiro *et al.*, 2018).

#### 5 RESULTADOS E DISCUSSÃO

#### 5.1 Tonalidade

Conforme descrito na seção 4, inicialmente comparam-se dois modelos somente com efeitos fixos  $(M_1 \in M_2)$ . As somas de Quadrados dos Resíduos e o resultado do Teste da Razão de Verossimilhanças são apresentados na Tabela 5.1.

Tabela 5.1: Número de parâmetros (np), Soma de Quadrados dos Resíduos (SQRes) e nível descritivo referente ao Teste da Razão de Verossimilhanças (valor-p) na comparação entre o modelo linear simples (M1) e o modelo polinomial de grau 2 (M2).

Modelo	np	SQRes	valor-p
M1	3	$2995,\! 6$	
M2	4	2960,3	$0,\!070$

Verifica-se um menor valor da soma de quadrados dos resíduos para M2. Por meio do teste da razão de verossimilhanças o efeito quadrático não foi significativo para tal modelo, porém seleciona-se o modelo polinomial quadrático (M2) afim de verificar a significância do parâmetro de efeito aleatório no termo quadrático. Os intervalos de confiança individuais, para cada coeficiente do modelo, são apresentados na Figura 5.1. Observa-se que há intervalos disjuntos em todos os parâmetros, isto é, intervalos que não se sobrepõem quando comparados no mesmo parâmetro, indicando a necessidade de incluir efeitos aleatórios em todos os parâmetros do modelo, destacando-se também a necessidade para o parâmetro do termo quadrático.



Figura 5.1: Intervalos de confiança individuais dos coeficientes do modelo polinomial de grau 2 (M2) para a variável tonalidade.

Escolhido a parte fixa do modelo e propondo-se a parte aleatória por meio das análises exploratórias, obtém-se o modelo maximal M1.t descrito na equação (4.2) e a partir dele temse modelos encaixados, cujo resumo está apresentado na Tabela 5.2. As três primeiras linhas (modelos M1.t ao M3.t) são referentes à seleção dos efeitos aleatórios e as três últimas (modelos M4.t ao M6.t) são referentes à seleção dos efeitos fixos.

Modelo	Efeito fixo	Efeito aleatório	np	$-\log(L)$	Comparação	$\chi^2$	gl	valor-p
M1.t	$\beta_0, \beta_1, \beta_2$	$\beta_0, \beta_1, \beta_2$	10	400,93				
M2.t	$\beta_0, \beta_1, \beta_2$	$\beta_0, \beta_1$	$\overline{7}$	$425,\!51$	M1.t $\times$ M2.t	$49,\!15$	3	< 0,001
M3.t	$\beta_0, \beta_1, \beta_2$	$\beta_0$	5	456, 16	M2.t $\times$ M3.t	$237,\!95$	2	< 0,001
M4.t	$\beta_0, \beta_1, \beta_2$	$\beta_0, \beta_1, \beta_2$	10	400,93				
M5.t	$\beta_0, \beta_1$	$\beta_0, \beta_1, \beta_2$	9	$410,\!45$	M4.t $\times$ M5.t	$19,\!04$	1	< 0,001
M6.t	$\beta_0$	$\beta_0, \beta_1, \beta_2$	8	$437,\!55$	M5.t $\times$ M6.t	$54,\!19$	1	$<\!0,\!001$

Tabela 5.2: Modelos encaixados referentes à variável tonalidade, com respectivos números de parâmetros (np), logarítmo da função de verossimilhança ( $-\log(L)$ ), estatística  $\chi^2$ , números de graus de liberdade (gl) e valor descritivo referente ao teste da razão de verossimilhanças (valor-p).

Por meio do teste da razão de verossimilhanças, selecionou-se o modelo maximal (M1.t). As estimativas dos componentes de variâncias e covariâncias dos efeitos aleatórios do modelo são descrita na matriz (5.1), em que nota-se variâncias diferentes e relação positiva entre os efeitos  $b_0 e b_1 e$  relação negativa entre os efeitos  $b_0 e b_2 e$  também entre  $b_1 e b_2$ .

$$\widehat{\mathbf{D}} = \begin{bmatrix} 1,725 & 0,250 & -0,015\\ 0,250 & 0,287 & -0,012\\ -0,015 & -0,012 & 0,0007 \end{bmatrix}$$
(5.1)

Na Tabela 5.3 apresentam-se as estimativas dos desvios padrões dos efeitos aleatórios, as correlações e respectivos intervalos de 95% de confiança. Adicionalmente, visualiza-se pela Figura 5.2 o gráfico de dispersão das relações entre os efeitos aleatórios, permitindo estudar o tipo de estrutura de variâncias e covariâncias mais adequado para tais efeitos.

Tabela 5.3: Estimativas para o desvio padrão dos efeitos aleatórios e para o coeficiente de correlação entre tais efeitos, com respectivos limites inferiores e limites superiores dos intervalos de 95% de confiança.

Termo	Estimativas	Limites inferiores	Limites superiores
$\sigma_{b_0}$	1,313	0,911	1,894
$\sigma_{b_1}$	0,536	0,377	0,762
$\sigma_{b_2}$	0,027	0,018	0,041
$\operatorname{cor}(b_0, b_1)$	0,354	-0,159	0,717
$\operatorname{cor}(b_0, b_2)$	-0,419	-0,775	0,140
$\operatorname{cor}(b_1, b_2)$	-0,807	-0,928	-0,531

Observa-se que os efeitos aleatórios  $b_0 e b_1$  apresentam correlação linear positiva (matriz (5.1)), porém o intervalo de confiança para  $cor(b_0, b_1)$  inclui o zero (Tabela 5.3), logo pode-se dizer que a correlação entre eles não é significativa. Analogamente ocorre com os efeitos aleatórios  $b_0 e b_2$ , que possuem uma correlação linear negativa, com o zero incluído no respectivo intervalo de 95% de confiança.



Figura 5.2: Gráfico de dispersão entre os efeitos aleatórios do modelo maximal (M1.t).

Portanto, sugere-se uma matriz de variâncias e covariâncias bloco-diagonal, com o efeito aleatório de intercepto sendo um bloco e efeitos de tempo e tempo quadrático sendo outro bloco.

#### 5.1.1 Seleção das estruturas das matrizes D e $\Sigma_i$

Foram testadas as estruturas para a matriz  $\mathbf{D}$  de variâncias e covariâncias dos efeitos aleatórios do modelo selecionado. A Tabela 5.4 apresenta o resumo dos valores obtidos para cada uma das comparações feitas.

Tabela 5.4: Estruturas consideradas para a matriz de variâncias e covariâncias dos efeitos aleatórios do modelo M1.t, em que NE é matriz não estruturada, CVH é componentes de variâncias com heterogeneidade e NE.CV é a bloco-diagonal, com respectivos números de parâmetros (np), valores do critério Bayesiano (BIC), logarítmo da função de verossimilhança (-log(L)), Estatística do teste da razão de verossimilhanças (TRV) e valor descritivo referente ao teste da razão de verossimilhanças (valor-p).

Estrutura	np	BIC	$-\log(L)$	Comparação	TRV	valor-p
NE	10	870,10	406,98			
CVH	7	$872,\!99$	$416,\!85$	$\rm NE \times \rm CVH$	19,73	0,0002
NE.CV	8	$861,\!18$	408, 14	$\rm NE \times NE.CV$	$2,\!30$	0,3158
				$\mathrm{CVH} \times \mathrm{NE.CV}$	$17,\!42$	$<\!0,\!001$

Pelo valor-p do teste da razão de verossimilhança, pode-se concluir que a estrutura NE.CV é a mais adequada para acomodar a variância e covariância intra-frutos. Observase também que o valor BIC é o menor considerando a estrutura NE.CV. Então, a matriz de variâncias e covariâncias ajustada para os efeitos aleatórios é dada por:

$$\hat{\boldsymbol{D}} = \begin{bmatrix} 1,8707 & 0 & 0\\ 0 & 0,3046 & -0,0129\\ 0 & -0,0129 & 0,0008 \end{bmatrix}$$

A seguir, testam-se estruturas para a matriz de variâncias e covariâncias dos erros aleatórios,  $\Sigma_i$ , que estão resumidas na Tabela 5.5.

Tabela 5.5: Estruturas consideradas para a matriz de variâncias e covariâncias dos erros do modelo M1.t, em que CVH é componentes de variâncias heterogêneas e ARH(1) é matriz autorregressiva heterogêneae de ordem 1, com respectivos números de parâmetros do modelo (np), valores do critério Bayesiano (BIC), logarítmo da função de verossimilhança (-log(L)), Estatística do teste da razão de verossimilhanças (TRV) e valor descritivo referente ao teste da razão de verossimilhanças (valor-p).

Estrutura	np	BIC	$-\log(L)$	Comparação	$\mathrm{TRV}$	valor-p
CVH	22	890,64	$383,\!57$			
ARH(1)	23	$890,\!10$	$380,\!50$	$ARH(1) \times CVH$	$6,\!15$	0,0131

Por meio do teste da razão de veros similhanças, a estrutura ARH(1) foi selecionada ao nível de 5% de significância com parâmetro autor regressivo dado por  $\rho = 0,23$ . Portanto, o modelo final é dado por:

$$\hat{Y}_i = (\hat{\beta}_0 + \hat{b}_{0i}) + (\hat{\beta}_1 + \hat{b}_{1i})t + (\hat{\beta}_2 + \hat{b}_{0i})t^2.$$
(5.2)

cujas matrizes de variâncias e covariâncias dos efeitos aleatórios e dos erros são, respectivamente:

$$\widehat{\boldsymbol{D}} = \begin{bmatrix} \sigma_0^2 & 0 & 0 \\ 0 & \sigma_1^2 & \sigma_{12} \\ 0 & \sigma_{12} & \sigma_2^2 \end{bmatrix} \quad \mathbf{e} \quad \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{i}} = \begin{bmatrix} \sigma_1^2 & \sigma_1 \sigma_2 \rho & \cdots & \sigma_1 \sigma_{n_i} \rho^{n_i - 1} \\ \sigma_1 \sigma_2 \rho & \sigma_2^2 & \cdots & \sigma_2 \sigma_{n_i} \rho^{n_i - 1} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_1 \sigma_{n_i} \rho^{n_i - 1} & \sigma_2 \sigma_{n_i} \rho^{n_i - 2} & \cdots & \sigma_{n_i}^2 \end{bmatrix}$$

Na Tabela 5.6 apresentam-se as estimativas dos parâmetros de efeitos fixos com respectivos intervalos de 95% de confiança, os desvios padrões, estatística F e valor-p, que indicam a importância de todos esses efeitos.

Tabela 5.6: Estimativas dos efeitos fixos do modelo quadrático para a tonalidade, com respectivos desvios padrões (DP), limites inferiores e limites superiores dos intervalos de 95% de confiança, teste F e valor-p.

Parâmetros	Estimativas	DP	Limites inferiores	Limites superiores	F	valor-p
$\beta_0$	112,11	0,325	111,48	112,75	121777,40	< 0,001
$\beta_1$	-1,39	$0,\!136$	-1,66	-1,12	$796,\!30$	< 0,001
$\beta_2$	-0,04	0,007	-0,05	-0,02	$24,\!09$	$<\!0,\!001$

Na Figura 5.3 estão as observações e as curvas ajustadas com valores preditos pelo modelo final, para cada fruto. Observam-se que as curvas ajustadas (linha contínua) aproximamse bem das observações dos frutos, isto é, aparentemente o modelo está captando a variabilidade existente dentro do mesmo indivíduo e assim ajustando-se bem aos dados. Espera-se que ao longo do período de maturação os valores da tonalidade média diminuam, pois as cores que as cascas do mamão podem ter do início ao fim desse período percorrem do maior ao menor grau de tonalidade (Figura 2.1), respectivamente.



Figura 5.3: Gráfico das predições do modelo final (linha conínua) com as observações (pontos) para cada fruto.

#### 5.1.2 Diagnósticos

Na Figura 5.4 são apresentados os gráficos de diagnósticos referente ao modelo final para a variável tonalidade. Nota-se pela Figura 5.4 (a) que os resíduos estão nos limites consideráveis, que seria o intervalo [-3,3]. A Figura 5.4(b) mostra que os valores estimados estão próximos aos valores observados, indicando que o modelo final está conseguindo explicar a variabilidade da tonalidade média.

Já na Figura 5.5 tem-se os resíduos condicionais em relação aos valores ajustados. Verifica-se que os pontos estão distribuídos em torno de zero e aparentemente não há nenhum padrão nos resíduos. Assim, aparentemente, estão satisfazendo a condição de independência.

Pode-se verificar, por meio do gráfico quantil-quantil com envelope simulado da distribuição normal apresentado na Figura 5.6 (a), à normalidade dos resíduos condicionais padronizados, em que poucos pontos estão fora do envelope simulado (menos de 8%). Tem-se também o histograma, Figura 5.6 (b), que indica aderência à pressuposição de normalidade. Já no gráfico apresentado na Figura 5.7 tem-se os quantis da distribuição normal em função das predições dos efeitos aleatórios, indicando que a suposição de normalidade é satisfeita. Portanto, conclui-se que o modelo final está adequado para descrever a variável tonalidade.



Figura 5.4: Gráfico dos resíduos em função do tempo (a) e gráfico dos valores ajustados em função dos valores observados (b) para a variável tonalidade.



Figura 5.5: Gráfico dos resíduos condicionais padronizados em função dos valores ajustados para a variável tonalidade.



Figura 5.6: Gráfico normal de probabilidade dos resíduos com envelope simulado (a) e Histograma dos resíduos com adicional de uma curva da distribuição normal (b) para a variável tonalidade.



Figura 5.7: Gráficos normais de probabilidade com envelope simulado para os efeitos aleatórios  $b_0$  (a),  $b_1$  (b) e  $b_2$  (c) do modelo quadrático para a variável tonalidade.

# 5.2 Intensidade

Analogamente ao processo feito para a variável tonalidade, inicialmente comparam-se dois modelos somente com efeitos fixos (M1<sup>\*</sup> e M2<sup>\*</sup>). As Somas de Quadrados dos Resíduos e o valor-p do Teste da Razão de Verossimilhanças são apresentados na Tabela 5.7.

Tabela 5.7: Número de parâmetros (np), Soma de Quadrados dos Resíduos (SQRes) e o nível descritível referente ao teste da razão de verossimilhanças (valor-p) na comparação entre o modelo polinomial de grau 3 (M1<sup>\*</sup>) e o modelo polinomial de grau 4 (M2<sup>\*</sup>).

Modelo	np	SQRes	valor-p
$M1^*$	5	$598,\!01$	
$M2^*$	6	$595,\!69$	$0,\!30$

Foi preferível o modelo M1<sup>\*</sup> pelo teste da razão de verossimilhanças. Nota-se que a diferença das somas de quadrados dos resíduos dos modelos foi mínima, ou seja, os dois modelos estão descrevendo os dados de maneira similar. O modelo selecionado foi ajustado para cada fruto, sendo possível obter os intervalos de confiança individuais para cada coeficiente, apresentados na Figura 5.8. A análise desse gráfico, sugere a inclusão de efeitos aleatórios em todos os parâmetros do modelo em questão.



Figura 5.8: Intervalos de confiança individuais dos coeficientes do modelo polinomial de grau 3  $(M1^*)$  para a variável intensidade.

Assim, obtém-se o modelo maximal M1.i descrito na equação (4.3) e a partir dele temse modelos encaixados, no qual o resumo está apresentado na Tabela 5.8. As quatro primeiras linhas (modelos M1.i ao M4.i) são referentes à seleção dos efeitos aleatórios e as quatro últimas (modelos M5.i ao M8.i) são referentes à seleção dos efeitos fixos.

Tabela 5.8: Modelos encaixados referente à variável intensidade, com respectivos números de parâmetros (np), logaritmo da verossimilhança ( $-\log(L)$ ), estatística  $\chi^2$ , números de graus de liberdade (gl) e nível descritivo referente ao teste da razão de verossimilhanças (valor-p).

Modelo	Efeito fixo	Efeito aleatório	$^{\mathrm{np}}$	$-\log(L)$	Comparação	$\chi^2$	$\operatorname{gl}$	valor-p
M1.i	$\beta_0, \beta_1, \beta_2, \beta_3$	$\beta_0, \beta_1, \beta_2, \beta_3$	15	$415,\!60$				
M2.i	$\beta_0, \beta_1, \beta_2, \beta_3$	$\beta_0, \beta_1, \beta_2$	11	$421,\!97$	M1.i $\times$ M2.i	12,74	4	0,013
M3.i	$\beta_0, \beta_1, \beta_2, \beta_3$	$\beta_0, \beta_1$	8	$436,\!25$	M2.i $\times$ M3.i	$28,\!56$	3	< 0,001
M4.i	$\beta_0, \beta_1, \beta_2, \beta_3$	$\beta_0$	6	$457,\!33$	M3.i $\times$ M4.i	42,16	2	< 0,001
M5.i	$\beta_0, \beta_1, \beta_2, \beta_3$	$\beta_0, \beta_1, \beta_2, \beta_3$	15	$421,\!10$				
M6.i	$\beta_0, \beta_1, \beta_2$	$\beta_0, \beta_1, \beta_2, \beta_3$	14	$427,\!40$	M5.i $\times$ M6.i	$12,\!61$	1	< 0,001
M7.i	$\beta_0, \beta_1$	$\beta_0, \beta_1, \beta_2, \beta_3$	13	$424,\!59$	M6.i $\times$ M7.i	$5,\!62$	1	< 0,001
M8.i	$\beta_0$	$\beta_0, \beta_1, \beta_2, \beta_3$	12	$426,\!33$	M7. i $\times$ M8. i	$3,\!48$	1	0,062

Com base no teste da razão de verossimilhanças, seleciona-se o modelo M8.i. As estimativas dos componentes de variâncias e covariâncias dos efeitos aleatórios do modelo são descritas na matriz (5.3). Observam-se variâncias de diferentes valores e relação positiva apenas entre os efeitos  $b_0$  e  $b_3$ .

$$\widehat{\mathbf{D}} = \begin{vmatrix} 1,31070 & -0,13686 & -0,00348 & 0,00055 \\ -0,13686 & 0,40508 & -0,06527 & 0,00253 \\ -0,00348 & -0,06527 & 0,01272 & -0,00055 \\ 0,00055 & 0,00253 & -0,00055 & 0,00003 \end{vmatrix}$$
(5.3)

Tem-se na Tabela 5.9 tem-se as estimativas pontuais dos desvios padrões, as correlações e respectivos intervalos de 95% de confiança para os efeitos aleatórios. Adicionalmente, visualiza-se pela Figura 5.9 o gráfico de dispersão das relações entre os efeitos aleatórios.

Tabela 5.9: Estimativas para o desvio padrão dos efeitos aleatórios e para o coeficiente de correlação entre tais efeitos, com respectivos limites inferiores e limites superiores dos intervalos de 95% de confiança.

Termo	Estimativa	Limites inferiores	Limites superiores
$\sigma_{b_0}$	1,145	0,700	1,880
$\sigma_{b_1}$	$0,\!637$	0,460	0,884
$\sigma_{b_2}$	0,112	0,082	$0,\!154$
$\sigma_{b_3}$	0,005	0,003	0,007
$\operatorname{cor}(b_0, b_1)$	-0,188	-0,684	$0,\!427$
$\operatorname{cor}(b_0, b_2)$	-0,027	-0,613	0,578
$\operatorname{cor}(b_0, b_3)$	0,094	-0,598	0,706
$\operatorname{cor}(b_1, b_2)$	-0,909	-0,926	-0,888
$\operatorname{cor}(b_1, b_3)$	0,785	0,744	0,820
$\operatorname{cor}(b_2, b_3)$	-0,970	-0,980	-0,957



Figura 5.9: Gráfico de dispersão entre os efeitos aleatório do modelo M8.i.

Observam-se que os efeitos aleatórios  $b_0$  e  $b_1$ ,  $b_0$  e  $b_2$ ,  $b_0$  e  $b_3$  apresentam correlações lineares negativas e tem-se o zero incluso nos intervalos de confiança. Portanto pode-se dizer que a correlação entre eles não é significativa e sugere-se uma matriz de variâncias e covariâncias bloco-diagonal. Supõe-se que a correlação do intercepto com os outros efeitos é nulo.

#### 5.2.1 Seleção das estruturas das matrizes D e $\Sigma_i$

Para a seleção de estruturas de matrizes de variâncias e covariâncias foram consideradas as apresentadas na seção 4, cujo resumo está na Tabela 5.10.

Tabela 5.10: Estruturas consideradas para a matriz de variâncias e covariâncias dos efeitos aleatórios do modelo M8.i, em que NE é matriz não estruturada e NE.CV é a bloco-diagonal, com respectivos números de parâmetros (np), valores do critério Bayesiano (BIC), Estatística do teste da razão de verossimilhanças (TRV) e o nível descritivo referente ao TRV (valor-p).

Estrutura	np	BIC	-logLik	Comparação	TRV	valor-p
NE	12	894,021	413,288			
NE.CV	9	$885,\!336$	$417,\!376$	$NE \ge NE.CV$	$^{8,176}$	$0,\!043$

Pelo critério BIC, a estrutura NE.CV apresentou o menor valor, porém o teste da razão de verossimilhanças foi favorável à não estruturada (NE). Selecionou-se a estrutura NE.CV, uma vez que apresenta menor número de parâmetros e o nível descritivo referente ao teste da razão de verossimilhança foi muito próximo à 5%. Então, tem-se a matriz de variâncias e covariâncias ajustada para os efeitos aleatórios é dada por:

$$\widehat{\boldsymbol{D}} = \begin{bmatrix} 0,67785 & 0 & 0 & 0\\ 0 & 0,40184 & -0,07178 & 0,00294\\ 0 & -0,07178 & 0,01485 & -0,00067\\ 0 & 0,00295 & -0,00067 & 0,00003 \end{bmatrix}$$
(5.4)

Em seguida, testam-se estruturas para a matriz de variâncias e covariâncias dos erros aleatórios,  $\Sigma_i$ , resumidas na Tabela 5.11.

Tabela 5.11: Estruturas consideradas para a matriz de variância e covariância dos erros do modelo M8.i, em que CV é matriz componentes de variâncias, CVH é matriz componentes de variâncias heterogêneas e AR(1) é autorregressiva de ordem 1, com respectivos números de parâmetros do modelo (np), valores do critério Bayesiano (BIC), logaritmo da verossimilhança (-log(L)), Estatística do teste da razão de verossimilhanças (TRV) e nível descritivo referente ao TRV (valor-p).

Estrutura	np	BIC	$-\log(L)$	Comparação	TRV	valor-p
CV	9	$885,\!33$	$417,\!37$			
$\operatorname{CVH}$	23	$916,\!28$	$393,\!50$	$\mathrm{CV} \times \mathrm{CVH}$	47,73	< 0,001
AR(1)	10	837,90	$408,\!95$	$\text{CVH} \times \text{AR}(1)$	$30,\!89$	0,003
				$\mathrm{CV} \times \mathrm{AR}(1)$	$16,\!84$	$<\!0,\!001$

Observa-se que a estrutura AR(1) teve o menor valor do BIC, porém, pelo teste da razão de verossimilhanças, selecionou-se a estrutura CVH. Portanto, o modelo final é dado por:

$$\widehat{\mathbf{Y}}_{i} = (\widehat{\beta}_{0} + \widehat{b}_{0i}) + \widehat{b}_{1i}t + \widehat{b}_{2i}t^{2} + \widehat{b}_{3i}t^{3}$$
(5.5)

cujas matrizes de variâncias e covariâncias dos efeitos aleatórios e dos erros são, respectivamente:

$$\widehat{\boldsymbol{D}} = \begin{bmatrix} \sigma_0^2 & 0 & 0 & 0\\ 0 & \sigma_1^2 & \sigma_{12} & \sigma_{13}\\ 0 & \sigma_{12} & \sigma_2^2 & \sigma_{23}\\ 0 & \sigma_{13} & \sigma_{23} & \sigma_3^2 \end{bmatrix} \quad \mathbf{e} \quad \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{i}} = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0\\ 0 & \sigma_2^2 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \sigma_{n_i}^2. \end{bmatrix}$$

Na Tabela 5.12 apresentam-se a estimativa do único parâmetro de efeitos fixo, o intercepto, com respectivo intervalo de 95% de confiança, o desvio padrão, estatística F e valor-p.

Tabela 5.12: Estimativa do efeito fixo do modelo cúbico para a intensidade, com respectivo desvio padrão (DP), limite inferior e limite superior do intervalo de 95% de confiança, teste F e valor-p.

Parâmetros	Estimativa	DP	Limite inferior	Limite superior	teste F	valor-p
$\beta_0$	34,77	0,20	34,36	35,18	$27803,\!8$	< 0,001

Na Figura 5.10 tem-se as observações de cada fruto juntamente com as curvas ajustadas por meio das predições obtidas com o modelo final, possibilitando visualizar a qualidade dos ajustes com relação aos perfis individuais. Nota-se que as observações dentro de cada fruto estão flutuando em torno das curvas individuais, porém há frutos em que o comportamento da curva não é similar ao comportamento das observações ao longo do tempo de maturação, como acontece nos frutos 1 e 12 por exemplo. Nesse contexto, pode-se dizer que aparentemente o modelo está satisfatório, porém pode ser que outros tipos de modelos possam descrever melhor os dados de intensidade, ao captar melhor a variabilidade dentro dos frutos.



Figura 5.10: Gráfico das predições do modelo final (linha conínua) com as observações (pontos) para cada fruto.

#### 5.2.2 Diagnósticos

Na Figura 5.11 são apresentados os gráficos de diagnósticos referente ao modelo final para a variável intensidade. Nota-se que os resíduos (Figura 5.11 (a)), estão dentro dos limites consideráveis, que seria o intervalo [-3,3]. Aparentemente a estrutura CVH incorporou ao modelo a heterocedasticidade. A Figura 5.11 (b) mostra que a maior parte dos valores estimados estão próximos aos valores observados, com indícios de que o modelo final está conseguindo explicar a variabilidade da intensidade parcialmente.

Já na Figura 5.12 tem-se os resíduos condicionais em relação aos valores ajustados, o qual apresentam pontos distribuídos em torno de zero e aparentemente não indicam nenhum padrão nos resíduos. Assim, estão satisfazendo a condição de independência.

Pode-se verificar, por meio do gráfico quantil-quantil com envelope simulado da distribuição normal apresentado na Figura 5.13 (a), à normalidade dos resíduos condicionais padronizados, em que poucos pontos estão fora do envelope simulado (menos de 5%). Tem-se também o histograma, Figura 5.13 (b), apresentando uma boa simetria dos resíduos. Já no gráfico apresentado na Figura 5.14 tem-se os quantis da distribuição normal em função das predições dos efeitos aleatórios, indicando que a suposição de normalidade para tais efeitos é satisfeita. Portanto, conclui-se que o modelo final está descrevendo a variável intensidade porém há indícios que não está captando toda a variabilidade.



Figura 5.11: Gráfico dos resíduos em função do tempo (a) e gráfico dos valores ajustados em função dos valores observados (b) para a variável intensidade.



Figura 5.12: Gráfico dos resíduos condicionais padronizados em função dos valores ajustados para a variável intensidade.



Figura 5.13: Gráfico de normalidade dos resíduos com envelope simulado da distribuição normal (a) e Histograma de normalidade dos resíduos com adicional de uma curva da distribuição normal (b) para a variável intensidade.



Figura 5.14: Gráficos normais de probabilidade com envelope simulado para os efeitos aleatórios  $b_0$  (a),  $b_1$  (b),  $b_2$  (c) e  $b_3$  (d) do modelo cúbico para a variável intensidade.

# 6 CONSIDERAÇÕES FINAIS

O estudo da maturação de frutos é relevante para a fruticultura, sendo de grande importância a classificação correta dos estágios de maturação. O uso de equipamentos para automatizar o processo de separação de frutos de acordo com seu estágio de maturação é cada vez mais frequente. Diante disso, a análise de imagens digitais é utilizada como método de avaliar características da parte externa dos frutos.

A tonalidade e intensidade são características expressas na casca de alguns frutos, que podem ser utilizadas para estudar a maturação dos mesmos. No entanto, adicionalmente se faz necessário a busca de modelos estatísticos adequados para a análise dessas variáveis, sobretudo em estudos longitudinais.

A proposta desse trabalho apresenta uma introdução à classes dos modelos lineares mistos aplicados ao estudo da tonalidade e intensidade da cor da casca de frutos da espécie *Carica papaya L.*, no seu período de maturação, obtidas por meio de um *scanner*. Para a variável tonalidade, o modelo polinomial quadrático, com efeito aleatório adicionado a todos os termos, é o que melhor explica a variabilidade dos dados de tonalidade média da casca do fruto.

Inicialmente, na investigação do modelo fixo (página 39) o modelo linear foi o selecionado pelo TRV. Fazendo as etapas de seleção de modelo, análogo ao que se descreveu no item 5.1, tal modelo se mostrou simples demais para ajustar os dados da tonalidade, sendo visível na análise de resíduos. Por tal motivo, selecionou-se o modelo quadrático inicialmente. Na prática, o modelo de segundo grau pode ajudar no agrupamento de frutos de acordo com a sua velocidades de maturação, por meio da derivação dessa equação de segundo grau.

Para a variável intensidade, o modelo polinomial cúbico com efeito fixo apenas no intercepto e efeito aleatório em todos os termos, apresentou-se como o que melhor explica a variabilidade dos dados de intensidade. Porém na análise de diagnósticos, os valores estimados não estão tão próximos aos valores observados, indicando que tal modelo não está conseguindo explicar toda a variabilidade da intensidade.

Há outras abordagens metodológicas que podem ser utilizadas, uma delas é o uso de Modelos não lineares ou Modelos não paramétricos. Como perspectiva de trabalhos futuros tem-se o estudo de modelos não lineares mistos, desenvolvimento de técnica de predição e consequentemente pode-se fazer um comparativo com o estudo aqui apresentado.

# REFERÊNCIAS

- AKAIKE, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, v. 19, n. 6, p. 716–723, 1974.
- [2] ALCARDE, R. Modelos lineares mistos em dados longitudinais com o uso do pacote ASReml-R. 2012. p. 156. Tese - Universidade de São Paulo/Escola Superior de Agricultura "Luiz de Queiroz", 2012.
- [3] AMADOR, J. P. Modelos mistos no ajuste de curvas de crescimento de Ricinus communis L. 2010. p. 141. Tese - Universidade de Santa Maria, 2010.
- [4] BALBINO, J. M. S. Efeitos de hidrotermia, refrigeração e ethephon na qualidade póscolheita do mamão (Carica papaya L.). 1997. p. 104. Tese - Universidade Federal de Viçosa, 1997.
- [5] BARBIERI, M. G. Retrospectiva 2017 & perspectiva 2018. Hortifruti Brasil, p. 40–41, 2018.
- [6] BATES, D. M.; WATTS, D. G. Nonlinear Regression Analysis and Its Applications. 2nd. ed. Wiley, 1988.
- [7] BATES, D.; MÄCHLER, M.; BOLKER, B. M.; WALKER, S. C. Fitting Linear Mixed-Effects Models using lme4. *Journal of statistical software*, v. 67, n. 1, p. 48, 2015.
- [8] BATES, D. M.; MAECHLER, M.; BOLKER, B.; WALKER, S.; CHRISTENSEN, R. H. B.; SINGMANN, H.; DAI, B.; SCHEIPL, F.; GABOR. Package 'lme4'. Technical report, 2019.
- [9] BURNHAM, K. P.; ANDERSON, D. R. Model Selection; Multimodel inference. 1998.
- [10] CARVALHO, J. N.; GURJÃO, E. C.; MATA, M. E. R. M. C.; DUARTE, M. E. M. Classificação automática de frutas por análise de imagem – o caso da manga Tommy Atkins. *Memórias do WTA 2014*, p. 77–91, 2014.
- [11] CHITARRA, M. I. F.; CHITARRA, A. B. Pós-colheita de frutas e hortaliças. 2. ed.. ed. Lavras: UFLA, 2005.
- [12] DAVIDIAN, M.; GILTINAN, D. M. Nonlinear models for repeated measurement data. Chapman Hall, 1995.
- [13] Embrapa Mandioca e Fruticultura Tropical. A cultura do mamão. 3. ed. rev. ampl.. ed. Brasília: coleção plantar, 2009.
- [14] Empresa Brasileira de Pesquisa Agropecuária-EMBRAPA. Base de dados dos produtos. Disponível em: <a href="http://www.cnpmf.embrapa.br/Base\_de\_Dados/index\_pdf/index\_pdf.htm">http://www.cnpmf.embrapa.br/Base\_de\_Dados/index\_pdf/index\_pdf.htm</a> , Acessadoem : 26/11/2018, 2018.
- [15] FRAZÉN, R.; HOPIA, A.; PAAKKI, M.; HANNA, M. Measuring the green color of vegetables from digital images using image analysis. *Food Science and Technology*, v. 63, p. 1184–1190, 2015.

- [16] HENDERSON, C. R.; KEMPTHORNE, O.; SEARLE, S. R. The estimation of environmental and genetic trends from records subject to culling<sup>\*</sup>. *Biometrics*, n. 2, p. 192–218, 1959.
- [17] IBGE. Produção Agrícola Municipal: culturas temporarias e permanentes. Rio de Janeiro: , 2016. v. 43.
- [18] INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA-IBGE. Produção agrícola municipal 2017. Disponível em: <www.ibge.gov.br/cidade@>, Rio de Janeiro, 2017.
- [19] IEMMA, A. F. Modelos lineares: Uma introdução para profissionais na pesquisa agropecuária. Londrina: Universidade Estadual de Londrina, Departamento de Matemática Aplicada, 1987.
- [20] Konica Minolta. Compreendendo o Espaço de Cor CIE L\*C\*h. Disponível em: <a href="http://sensing.konicaminolta.com.br/2015/08/compreendendo-o-espaco-de-cor-cie-lch/">http://sensing.konicaminolta.com.br/2015/08/compreendendo-o-espaco-de-cor-cie-lch/</a>, Acessado em: 12/11/2018, 2018.
- [21] LAIRD, N. M.; WARE, J. H. Random-Effects Models for Longitudinal Data. *Biometrics*, v. 38, n. 4, p. 963–974, 1982.
- [22] LARA, I. A. R.; DEMÉTRIO, C. G. B.; SHIMAKURA, S. E. Considerações sobre a inclusão de efeitos aleatórios em modelos de transição de Markov. *Rev. Bras. Biom.*, v. 27, n. 2, p. 269–277, 2009.
- [23] LIMA, U. D. A.; FERREIRA, A.; ARNALDI, D.; SONODA, D.; FANTINI, R. Agroindustrialização de frutas. 2. ed., Biblioteca de Ciências Agrárias Luiz de Queiroz v. 5. Piracicaba: FEALQ, 2018.
- [24] LITTELL, R. C.; MILLIKEN, G. A.; STROUP, W. W.; WOLFINGER, R. D.; SCHABEN-BERGER, O. SAS for Mixed Models. 2nd. ed. 2006.
- [25] LÓPEZ, L.; IEMMA, A. On the Estimation and Prediction in Mixed Linear Models. Scientia Agricola, v. 55, n. 2, 1998.
- [26] MARCELINO, S. D. R.; IEMMA, A. F. Métodos de estimação de componentes de variância em modelos mistos desbalanceados. *Scientia Agricola*, v. 57, n. 4, p. 643–652, 2000.
- [27] MARTINS, D. S.; COSTA, A. F. S. A cultura do mamoeiro: tecnologias de produção. Vitória: ES:Incaper, 2003.
- [28] MENDOZA, F.; DEJMEK, P.; AGUILERA, J. M. Calibrated color measurements of agricultural foods using image analysis. *Postharvest Biology and Technology*, v. 41, n. 3, p. 285–295, 2006.
- [29] MENDOZA, F.; AGUILERA, J. M. Application of Image Analysis for Classification of Ripening Bananas. of food science, v. 69, n. 9, p. 478–487, 2004.
- [30] MONTGOMERY, D. C.; PECK, E. A.; VINING, G. G. Introduction to Linear Regression Analysis. 2012.

- [31] MOTTA, J. D.; QUEIROZ, A. J. M.; FIGUEIRÊDO, R. M. F.; SOUSA, K. D. S. M. Índice de cor e sua correlação com parâmetros físicos e físico-químicos de goiaba, manga e mamão. *Comunicata Scientiae*, v. 6, n. 1, p. 74–82, 2015.
- [32] MOOD, A. M. Introduction to the theory of statistics. 2d ed [by] alexander m. mood [and] franklin a. graybill.. ed., McGraw-Hill series in probability and statistics. New York: McGraw-Hill, 1963.
- [33] NASCIMENTO, A. R. B. Análise da produtividade de cultivares de café utilizando modelos lineares mistos. Brasília: UnB, 2016.
- [34] OLIVEIRA, T. P.; ZOCCHI, S. S.; JACOMINO, A. P. Measuring color hue in 'sunrise solo' papaya using a flatbed scanner. *Rev. Bras. Frutic*, v. 39, n. 2, 2016.
- [35] OLIVEIRA, M. A. B.; VIANNI, R.; SOUZA, G.; ARAÚJO, T. M. R. Caracterização do estágio de maturação do papaya "golden"em função da cor. *rev. Bras. Frutic.*, v. 24, n. 2, p. 559–561, 2002.
- [36] PEREIRA, N. N. Modelos não lineares mistos na análise de curvas de crescimento de bovinos de raça Tabapuã. 2014. p. 51. Tese - 2014.
- [37] PEREIRA, L. F. S.; Barbon Jr, S. B.; VALOUS, N. A.; BASBIN, D. F. Predicting the ripening of papaya fruit with digital imaging and random forests. *Computers and Electronics in Agriculture*, v. 145, p. 76–82, 2018.
- [38] PERRI, S. H. V.; IEMMA, A. F. Procedimento "MIXED" do SAS® para análise de modelos mistos. *Scientia Agricola*, v. 56, n. 4, p. 959–967, oct 1999.
- [39] PINHEIRO, J. C.; BATES, D. M. Mixed-Effects Models in S and S-PLUS. Springer, 2000.
- [40] PINHEIRO, J. C.; BATES, D. M.; DEBROY, S.; SARKAR, D.; EISPACK; HEISTERKAMP, S.; WILLIGEN, B. V.; R-CORE. Package 'nlme'. 2018.
- [41] R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2016.
- [42] REGAZZI, A. J.; SILVA, C. H. O. Teste para verificar a igualdade de parâmetros e a identidade de modelos de regresão não-linear. I. Dados no delineamento inteiramente casualizado. *Revista* de Matemática e Estatística, v. 22, n. 2004, p. 33–45, 2004.
- [43] RENCHER, A. C.; SCHAALJE, G. B. Linear Models in Statistics. 2nd. ed. Wiley, 2008.
- [44] RODRIGUES, M. I.; IEMMA, A. F. Planejamento de experimentos e otimização de processos.
   1. ed.. ed. Campinas: Casa do Pão Editora, 2005.
- [45] SCHWARZ, G. Estimating the dimension of a model. The Annals of Statistics, v. 6, n. 2, p. 461–464, 1978.
- [46] SEARLE, S. R. Linear models. New York: New York Wiley, 1997.

- [47] Serviço brasileiro de apoio às micro e pequenas empresas-SEBRAE. O cultivo e o mercado do mamão. Disponível em: <a href="http://www.sebrae.com.br/sites/PortalSebrae/artigos/o-cultivoe-o-mercado-do-mamao,937a9e665b182410VgnVCM100000b272010aRCRD#>">http://www.sebrae.com.br/sites/PortalSebrae/artigos/o-cultivoe-o-mercado-do-mamao,937a9e665b182410VgnVCM100000b272010aRCRD#>">>, Acessado em: 26/11/2018, 2016.</a>
- [48] SELF, S. G.; LIANG, K.-Y. Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Tests Under Nonstandard Conditions. v. 82, n. 398, p. 605–610, 1987.
- [49] SILVA, G. C. B. M. Avaliação comparativa de características físicas, químicas e fisiológicas de mamões(Carica papaya L.) ev.Golden e ev.Aliança. 2016. p. 29. Tese - Federal de Viçosa, 2016.
- [50] SOUZA, G. S. Introdução aos modelos de regressão linear e não-linear. 1998.
- [51] STAPLETON, J. H. Linear statistical models. John Wiley, 1995.
- [52] VERBEKE, G.; MOLENBERGHS, G. Linear Mixed Models for Longitudinal Data. 2009.
- [53] WEST, B. T.; WELCH, K. B.; GALECKI, A. T. Linear mixed models: A practical gruide using statistical software. Chapman Hall, 2007.
- [54] WILKS, S. S. The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses. Ann. Math. Statist., v. 9, n. 1, p. 60–62, 1938.
- [55] WU, D.; SUN, D.-W. Colour measurements by computer vision for food quality control A review. Trends in Food Science and Technology, v. 29, n. 1, p. 5–20, 2013.
- [56] YAMANOUCHI, T. K. Seleção de modelos lineares mistos utilizando critérios de informação. Piracicaba: ESALQ, 2017.

## ANEXOS

#### Anexo A: Principais comandos no Software R na análise dos dados referente a tonalidade.

```
library(lattice)
library(ggplot2)
library(lme4)
library(nlme)
```

```
dados <- read.csv("papaya.csv")
```

# Gráficos individuais

xyplot(tonalidade ~tempo, groups=fruto, dados, type=c("l"), xlab = "Tempo(dias)", ylab = "Tonalidade(média)", xlim = c(0,15))

xyplot(tonalidade tempo | fruto, groups = fruto, dados, xlab = "Tempo (dias)", ylab = "Tonalidade (média)", type=c("l"), layout=c(5,4))

# Modelos de efeitos fixos

modelo1 <- lm(tonalidade tempo, dados) modelo2 <- lm(tonalidade tempo+I(tempo^2), dados)

# gráfico do intervalo de confiança individual dos coeficientes dos modelos

```
modelo2.lis <- lmList(tonalidade tempo+I(tempo^2)|fruto, dados)
plot(intervals(modelo2.lis))
```

```
\# Ajuste de modelos mistos
```

M1 <- lmer(tonalidade ~1 + tempo + I(tempo 2) + (1 + tempo + I(tempô) |fruto), dados) M2 <- lmer(tonalidade ~1 + tempo + I(tempo 2) + (1 + tempo |fruto), dados) M3 <- lmer(tonalidade ~1 + tempo + I(tempo 2) + (1 + tempo + I(tempo 2)|fruto), dados, M4 <- lmer(tonalidade ~1 + tempo + I(tempo 2) + (1 + tempo + I(tempo 2)|fruto), dados, REML = F) M5 <- lmer(tonalidade ~1 + tempo + (1 + tempo + I(tempo 2) |fruto), dados, REML = F) M6 <- lmer(tonalidade ~1 + (1 + tempo + I(tempo 2) |fruto), dados, REML = F) anova(M1,M2, test="LRT", method="REML") anova(M2,M3, test="LRT")

anova(M4,M5)

anova(M5,M6)

#Seleção das matrizes de variâncias <br/>e covariâncas dos efeitos aleatórios

# Não estruturada
M.ne <- lme(tonalidade ~1 + tempo + I(tempô), data=dados, random=list(fruto=pdSymm(~1 + tempo + I(tempo^2))), method = "REML")</p>

# Matriz diagonal heterogênea (CVH)
M.diag <- lme(tonalidade ~1 + tempo + I(tempo ^2), data=dados, random=list(fruto=pdDiag(~1+tempo+I(tempo ^2))), method = "REML")</p>

# Bloco diagonal (NE.CV) Model <- lme(tonalidade~1 + tempo + I(tempo^2), data=dados, random=list(fruto=pdBlocked (list(pdIdent(form=~1), pdSymm(form=~-1+tempo+I(tempo^2))))), method = "REML")

#Seleção das matrizes de variâncias <br/>e covariâncas dos erros aleatórios

# Matriz diagonal heterogênea (CVH)
M.cv <- update(Model, weight=varIdent(form=~1|tempo), control=lmeControl(opt="optim"))</pre>

# Autorregressiva de ordem 1 M.ar <- update(Model, weight=varIdent(form=~1|tempo), corr=corAR1(form=~1|fruto), control=lmeControl(opt="optim"))

anova(M.cv, M.ar)

# Diagnostico

 $\begin{array}{l} modelo\_ final<-lme(tonalidade~1+tempo+I(tempo^2), \ data=dados, \ random=list(fruto=pdBlocked(list(pdIdent(form=~1), \ pdSymm(form=~-1+tempo+I(tempo^2))))), corr=corAR1 \\ (form=~1|fruto), \ weight=varIdent(form=~tempo), \ method=~"REML") \end{array}$ 

```
# tempo x residuo
plot(tempo, resid(modelo_ final), ylim=c(-4,4), col="blue", xlab = "Tempo", ylab="Resíduo",
main = "(a)")
abline(h=0, col="red")
abline(h=c(-3,3), lty=2)
```

```
# observado x ajustado
plot(tonalidade, fitted(modelo_ final), col="blue", xlab = "Valores observados",
ylab="Valores ajustados", main="(b)")
abline(0,1)
```

# ajustado x resíduo plot(fitted(modelo\_ final), resid(modelo\_ final), col="blue", xlab = "Valores ajustados", ylab="Resíduo", ylim = c(-4,4)) abline(h=0, col="red") abline(h=c(-3,3), lty=2)

#normalidade dos efeitos aleatórios e dos resíduos

b0AR<-random.effects(modelo\_ final)[[1]] b1AR<-random.effects(modelo\_ final)[[2]] b2AR<-random.effects(modelo\_ final)[[3]]

qqPlot(b0AR,ylab="Quantil teórico da distribuição normal", xlab=expression(b[0]), main="(a)",grid=F) qqPlot(b1AR,ylab="Quantil teórico da distribuição normal", xlab=expression(b[1]), main="(b)",grid=F) qqPlot(b2AR,ylab="Quantil teórico da distribuição normal", xlab=expression(b[2]), main="(c)",grid=F)

hnp(residuals(modelo\_ final), xlab="Quantis teóricos", ylab="Resíduos", scale=T, half=F, print=F, main="(a)")

### Anexo B: Principais comandos no Software R na análise dos dados referente a intensidade.

library(nlme)
library(lattice)
library(ggplot2)
library(gridExtra)
library(lme4)
library(hnp)

```
dados <- read.csv("papaya.csv")
```

```
# Gráficos individuais
```

g1 <- ggplot(data=dados, aes(x=tempo, y=intensidade, group=fruto), xlim=c(0,20)) + geom\_ line() + geom\_ point() + scale\_ x\_ continuous(name=expression(Tempo)) + scale\_ y\_ continuous(name=expression(Íntensidade))

g2 <- ggplot(data = dados, aes(x = tempo, y = intensidade)) + geom\_ line()+ geom\_ point() + facet\_ wrap(~fruto) + scale\_ x\_ continuous(name=expression(Tempo)) + scale\_ y\_ continuous(name=expression('Intensidade'))

grid.arrange(g1,g2, ncol=1, nrow=2)

# Modelos de efeitos fixos

$$\label{eq:stars} \begin{split} & \text{modelo3} <- \text{lm}(\text{intensidade~tempo} + I(\text{tempo}^2) + I(\text{tempo}^3), \, \text{dados}) \\ & \text{modelo4} <- \text{lm}(\text{intensidade~tempo} + I(\text{tempo}^2) + I(\text{tempo}^3) + I(\text{tempo}^4), \, \text{dados}) \\ & \text{anova}(\text{modelo3}, \, \text{modelo4}, \, \text{test} = \text{"LRT"}) \end{split}$$

#Intervalos de confianças individuais para os coeficientes

modelo 3.lis <- lmList(intensidade tempo<br/>+I(tempo^2)+I(tempo^3), dados) plot(intervals(modelo 3.lis, level=0.95))

 $\#\,$  Ajuste de modelos para seleção dos efeitos ale<br/>atórios

M1 <- lmer(intensidade tempo + I(tempo 2) + I(tempo 3) + (tempo + I(tempo 2) + I(tempo 3)|fruto), dados, REML=T) M2 <- lmer(intensidade tempo + I(tempo 2) + I(tempo 3) + (tempo + I(tempo 2)|fruto), dados, REML=T) M3 <- lmer(intensidade tempo + I(tempo 2) + I(tempo 3) + (tempo |fruto), dados, REML=T) M4 <- lmer(intensidade tempo + I(tempo 2) + I(tempo 3) + (1|fruto), dados, REML=T) anova(M1,M2,test="LRT") anova(M1,M3,test="LRT") anova(M2,M3,test="LRT")

 $\#\,$  Ajuste de modelos para seleção dos efeitos fixos

M5 <- lmer(intensidade ~tempo + I(tempo^2) + I(tempo^3) + (tempo + I(tempo^2) + I(tempo ^3)|fruto), dados, REML=F) M6 <- lmer(intensidade ~tempo + I(tempo^2) + (tempo + I(tempo^2) + I(tempo^3)|fruto), dados, REML=F) M7 <- lmer(intensidade ~tempo + (tempo + I(tempo^2) + I(tempo^3)|fruto), dados, REML=F) M8 <- lmer(intensidade ~1 + (tempo + I(tempo^2) + I(tempo^3)|fruto), dados, REML=F) anova(M5, M6) anova(M6, M7) anova(M7, M8)

# Verificando as variâncias, covariâncias e correlações

M <- lme(intensidade~1 , data=dados, random=~(tempo + I(tempo^2)+ I(tempo^3))|fruto, control=lmeControl(maxIter=150, msMaxIter=100, niterEM=100, msMaxEval=300, msVerbose = F, returnObject=TRUE), method = "REML") getVarCov(M) intervals(M) splom(`random.effects(M), xlab = , varnames = c("Intercepto", "Tempo", "Tempo", "Tempo"))

#~Seleção para matrizes de variâncias <br/>e covariâncias dos efeitos aleatórios

# Não estruturada M.ne <- update(M, random=list(fruto=pdSymm(~(tempo + I(tempo^2) + I(tempo^3)))), method="REML", control=lmeControl(maxIter=150, msMaxIter=100, niterEM=100, msMaxEval=300, msVerbose = F, returnObject=TRUE))

# Bloco diagonal (NE.CV) Model <- update(M,random=list(fruto=pdBlocked(list(pdIdent(form=~1), pdSymm(form=~-1+tempo + I(tempo^2) + I(tempo^3))))), method="REML")

# Seleção para matrizes de variâncias covariâncias dos erros

M <- lme(intensidade~1, data=dados, random=~(tempo + I(tempo^2) + I(tempo^3))|fruto, control=lmeControl(maxIter=150, msMaxIter=100, niterEM=100, msMaxEval=300, msVerbose = F, returnObject=TRUE), method = "REML") Model <- update(M, random=list(fruto=pdBlocked(list(pdIdent(form=~1), pdSymm(form=~-1+tempo + I(tempo^2) + I(tempo^3))))), method="REML")

# Diagonal (CV) M.sc <- update(Model, corr=corCompSymm(form=~1|fruto), control=lmeControl(opt="optim"))

# Diagonal heterogênea(CVH)
M.cv <- update(Model, weight=varIdent(form= 1|tempo), control=lmeControl(maxIter=150, msMaxIter=100, niterEM=100, msMaxEval=300, returnObject=TRUE))</p>

# Autorregressiva de ordem 1 M.ar <- update(Model, corr=corAR1(form=~1|fruto), weights = varIdent(form=~tempo), control=lmeControl(maxIter=150, msMaxIter=100, niterEM=100, msMaxEval=300, returnObject=TRUE))

# Diagnóstico do modelo model\_ final <- M.cv

# tempo x residuo
plot(tempo, resid(model\_ final), ylim=c(-4,4), col="blue", xlab = "Tempo", ylab="Resíduo",
main="(a)")
abline(h=0, col="red")
abline(h=c(-3,3), lty=2)

# observado x ajustado
plot(intensidade, fitted(model\_ final), col="blue", xlab = "Valores observados",
ylab="Valores ajustados", xlim=c(31,39), ylim=c(31,39), main="(b)")
abline(0,1)

# ajustado x resíduo plot(fitted(model\_ final), resid(model\_ final), col="blue", xlab = "Valores ajustados", ylab="Resíduo", ylim = c(-4,4)) abline(h=0, col="red") abline(h=c(-3,3), lty=2)

# normalidade dos resíduos com envelope simulado hnp(residuals(model\_ final), xlab="Quantis teóricos", ylab="Resíduos", scale=T, alf=F, print=F)

#Gráfico das predições do modelo final

novo <- expand.grid(fruto = unique(dados\$ fruto),tempo = seq(0,14,1)) preditos <- predict(model\_ final, newdata = novo) dados.preditos <- data.frame(pred = preditos, novo) funi <- function(x) {  $cof[1] + cof[2] * x + coef[3] * x^2 + coef[4] * x^3$  } dados\$ fruto = as.factor(dados\$ fruto) ggplot(data = dados, aes(x = tempo, y = intensidade)) + geom\_ point() + geom\_ line(data = dados.preditos, aes(y = pred, x = tempo)) + facet\_ wrap(~fruto)+ xlab("Tempo") + ylab("Intensidade")