



João Pedro Serenini

## **Análise de um experimento longitudinal sobre a evolução do cancro cístico via modelos mistos**

Maringá – Paraná  
17 de abril de 2018

João Pedro Serenini

**Análise de um experimento longitudinal sobre a evolução  
do cancro cítrico via modelos mistos**

Orientadora: Prof.<sup>a</sup> Dr.<sup>a</sup> Terezinha Aparecida  
Guedes

Universidade Estadual de Maringá - UEM

Departamento de Estatística - DES

Mestrado em Bioestatística

Maringá – Paraná

17 de abril de 2018

---

## Lista de ilustrações

---

Figura 1 – Lesão de cancro cítrico foliar e no fruto. . . . .	9
Figura 2 – Colonia em agar apresentando o pigmento amarelo e a aparência geral mucóide. . . . .	10
Figura 3 – (a) Distribuição log-normal para $\mu = 0$ e $\sigma = 10; 1, 5; 1; 0, 5; 0, 25; 0, 125;$ (b) Distribuição log-normal para $\mu = 1$ e $\sigma = 10; 1, 5; 1; 0, 5; 0, 25; 0, 125;$ (c) Comparação entre as distribuições log-normal e normal . . . . .	27
Figura 4 – (a) Distribuição Gama para $\mu = 0$ e $\sigma = 0, 5; 1; 3; 5; 9; 7, 5;$ (b) Distribuição Gama para $\mu = 1$ e $\sigma = 0, 5; 1; 3; 5; 9; 7, 5;$ (c) Distribuição Gama para $\mu = 2$ e $\sigma = 0, 5; 1; 3; 5; 9; 7, 5.$ . . . . .	34
Figura 5 – Gráfico de dispersão dos dados . . . . .	42
Figura 6 – Gráfico de dispersão por variedade . . . . .	43
Figura 7 – Histograma da variável resposta diâmetro das lesões para as 12 variedades de laranja doce. . . . .	44
Figura 8 – <i>Box-plot</i> dos dados por variedade num total de 2880 observações. . . . .	46
Figura 9 – <i>Box-plot</i> dos dados por dias após a inoculação para cada variedade. . . . .	47
Figura 10 – Histograma da variável resposta diâmetro das lesões para os 8 dias após a inoculação. . . . .	48
Figura 11 – <i>Box-plot</i> dos dados por Folha. . . . .	49
Figura 12 – Perfil Médio em cada tempo para cada variedade . . . . .	50
Figura 13 – (a) Valores Ajustados <i>vs</i> Quantis Residuais; (b) Dias Após a Inoculação <i>vs</i> Quantis Residuais; (c) Quantis Residuais <i>vs</i> Densidade Estimada; (d) Quantis Teóricos <i>vs</i> Quantis Amostrais. . . . .	54
Figura 14 – (a) Valores Ajustados <i>vs</i> Quantis Residuais; (b) Dias Após a Inoculação <i>vs</i> Quantis Residuais; (c) Quantis Residuais <i>vs</i> Densidade Estimada; (d) Quantis Teóricos <i>vs</i> Quantis Amostrais. . . . .	55
Figura 15 – <i>Worm plot</i> dos resíduos quantílicos aleatorizados para o modelo Log-Normal. . . . .	56
Figura 16 – <i>Worm plot</i> dos resíduos quantílicos aleatorizados para o modelo Gama. . . . .	57

Figura 17 – Gráfico <i>detrended transformed Owen's</i> dos resíduos quantílicos aleatorizados para o modelo Log-Normal. . . . .	58
Figura 18 – Gráfico <i>detrended transformed Owen's</i> dos resíduos quantílicos aleatorizados para o modelo Gama. . . . .	58
Figura 19 – Gráficos de dispersão com linha de igualdade e Gráficos de Bland Altman para os modelo log-normal e gama. . . . .	63
Figura 20 – Gráfico de dispersão dos dados . . . . .	66
Figura 21 – Gráfico de dispersão por variedade . . . . .	67
Figura 22 – Histograma da variável resposta diâmetro das lesões para as 12 variedades de laranja doce. . . . .	68
Figura 23 – <i>Box-plot</i> dos dados por variedade num total de 2880 observações. . . .	70
Figura 24 – <i>Box-plot</i> dos dados por dias após a inoculação para cada variedade. . .	71
Figura 25 – Histograma da variável resposta diâmetro das lesões para os 8 dias após a inoculação. . . . .	72
Figura 26 – <i>Box-plot</i> dos dados por folha. . . . .	73
Figura 27 – Perfil Médio em cada tempo para cada variedade . . . . .	74
Figura 28 – (a) Valores Ajustados <i>vs</i> Quantis Residuais; (b) Dias Após a Inoculação <i>vs</i> Quantis Residuais; (c) Quantis Residuais <i>vs</i> Densidade Estimada; (d) Quantis Teóricos <i>vs</i> Quantis Amostrais. . . . .	77
Figura 29 – (a) Valores Ajustados <i>vs</i> Quantis Residuais; (b) Dias Após a Inoculação <i>vs</i> Quantis Residuais; (c) Quantis Residuais <i>vs</i> Densidade Estimada; (d) Quantis Teóricos <i>vs</i> Quantis Amostrais. . . . .	78
Figura 30 – <i>Worm plot</i> dos resíduos quantílicos aleatorizados para o modelo Log-Normal. . . . .	79
Figura 31 – <i>Worm plot</i> dos resíduos quantílicos aleatorizados para o modelo Gama. .	80
Figura 32 – Gráfico <i>detrended transformed Owen's</i> dos resíduos quantílicos aleatorizados para o modelo Log-Normal. . . . .	81
Figura 33 – Gráfico <i>detrended transformed Owen's</i> dos resíduos quantílicos aleatorizados para o modelo Gama. . . . .	81
Figura 34 – Gráficos de dispersão com linha de igualdade e Gráficos de Bland Altman para os modelo Log-Normal e Gama. . . . .	86



---

## Lista de tabelas

---

Tabela 1 – Variedades avaliadas . . . . .	14
Tabela 2 – 1º Ensaio . . . . .	16
Tabela 3 – 2º Ensaio . . . . .	16
Tabela 4 – Medidas Resumo para cada variedade. . . . .	45
Tabela 5 – Modelos propostos para a distribuição Log-normal . . . . .	51
Tabela 6 – Modelos propostos para a distribuição Gama. . . . .	51
Tabela 7 – Medidas do AIC, Deviance Global e BIC para os modelos Log-normal. . . . .	52
Tabela 8 – Medidas do AIC, Deviance Global e BIC para os modelos Gama. . . . .	52
Tabela 9 – Medidas descritivas dos resíduos para o modelo Log-Normal. . . . .	56
Tabela 10 – Medidas descritivas dos resíduos para o modelo Gama. . . . .	56
Tabela 11 – Estimativas dos parâmetros do modelo Log-Normal para $\mu$ . . . . .	59
Tabela 12 – Estimativas dos parâmetros do modelo Log-Normal para $\sigma$ . . . . .	60
Tabela 13 – Estimativas dos parâmetros do modelo Gama para $\mu$ . . . . .	60
Tabela 14 – Estimativas dos parâmetros do modelo Gama para $\sigma$ . . . . .	61
Tabela 15 – Estimativas das variâncias da parte aleatória do modelo log-normal para o parâmetro $\mu$ . . . . .	62
Tabela 16 – Estimativas das variâncias da parte aleatória do modelo log-normal para o parâmetro $\sigma$ . . . . .	62
Tabela 17 – Estimativas das variâncias da parte aleatória do modelo gama para o parâmetro $\mu$ . . . . .	62
Tabela 18 – Estimativas das variâncias da parte aleatória do modelo gama para o parâmetro $\sigma$ . . . . .	62
Tabela 19 – Valores de $R^2$ Cox & Snell. . . . .	62
Tabela 20 – Valores de $R^2$ Cragg & Uhler's. . . . .	62
Tabela 21 – Medidas Resumo para cada variedade. . . . .	69
Tabela 22 – Modelos propostos para a distribuição Log-normal . . . . .	75
Tabela 23 – Modelos propostos para a distribuição Gama. . . . .	75
Tabela 24 – Medidas do AIC, Deviance Global e BIC para os modelos Log-normal. . . . .	75

Tabela 25 – Medidas do AIC, Deviance Global e BIC para os modelos Gama. . . . .	76
Tabela 26 – Medidas descritivas dos resíduos para o modelo log-normal. . . . .	79
Tabela 27 – Medidas descritivas dos resíduos para o modelo gama. . . . .	79
Tabela 28 – Estimativas dos parâmetros do modelo Log-Normal para $\mu$ . . . . .	82
Tabela 29 – Estimativas dos parâmetros do modelo Log-Normal para $\sigma$ . . . . .	83
Tabela 30 – Estimativas dos parâmetros do modelo Gama para $\mu$ . . . . .	83
Tabela 31 – Estimativas dos parâmetros do modelo Gama para $\sigma$ . . . . .	84
Tabela 32 – Estimativas das variâncias da parte aleatória do modelo Log-Normal para o parâmetro $\mu$ . . . . .	84
Tabela 33 – Estimativas das variâncias da parte aleatória do modelo Log-Normal para o parâmetro $\sigma$ . . . . .	85
Tabela 34 – Estimativas das variâncias da parte aleatória do modelo Gama para o parâmetro $\mu$ . . . . .	85
Tabela 35 – Estimativas das variâncias da parte aleatória do modelo Gama para o parâmetro $\sigma$ . . . . .	85
Tabela 36 – Valores de $R^2$ Cox & Snell. . . . .	85
Tabela 37 – Valores de $R^2$ Cragg & Uhler's. . . . .	85

---

# Sumário

---

<b>INTRODUÇÃO</b>	<b>7</b>
<b>1 Materiais</b>	<b>14</b>
1.1 Implantação do Experimento	14
<b>2 Métodos</b>	<b>17</b>
2.0.1 Modelo Linear Misto (LMM)	17
2.0.2 Modelo Linear Generalizado Misto (GLMM)	20
2.0.3 Estruturas da Matriz de Covariância	22
2.0.4 Algoritmos de estimação dos parâmetros via pacote <i>gamlss</i>	23
2.0.5 Seleção do Modelo	24
2.1 Adequabilidade do modelo	25
2.2 Distribuição Log-Normal	26
2.2.1 Estimação dos Parâmetros da distribuição Log-Normal	28
2.2.2 Modelo Generalizado Misto Log-Normal	31
2.3 Distribuição Gama	32
2.3.1 Estimação dos Parâmetros da distribuição Gama	36
2.3.2 Modelo Generalizado Misto Gama	39
<b>3 Resultados 1º Ensaio</b>	<b>41</b>
3.1 Análise Descritiva	42
3.2 Ajuste	50
<b>4 Resultados 2º Ensaio</b>	<b>65</b>
4.1 Análise Descritiva	65
4.2 Ajuste	74
<b>5 Conclusão</b>	<b>87</b>
 <b>Referências</b>	 <b>88</b>

---

## Introdução

---

O estudo de dados longitudinais ou medidas repetidas envolvem a realização de duas ou mais observações nas unidades experimentais. Normalmente nestes tipos de experimentos as medidas são feitas de modo sistemático em cada unidade experimental, admitindo-se correlação não nula entre as observações feitas em ocasiões distintas. Este tipo de estudo pode apresentar uma estrutura hierárquica, uma vez que as medidas repetidas são aninhadas dentro do indivíduo. Portanto pode-se fazer a suposição de que as observações entre os indivíduos sejam independentes e que o aninhamento neles possua dependência correlacionada com os erros, sendo apropriado a especificação de um modelo para os valores médios em cada uma das ocasiões, além de uma estrutura para a matriz de covariância entre as medidas feitas ao longo do tempo.

Em termos gerais, o objetivo de estudos com medidas repetidas é caracterizar padrões da variável resposta sobre o tempo, e definir as relações existentes entre tais padrões e as covariáveis (WARE, 1985).

Uma vantagem em estudos com medidas repetidas é permitir uma avaliação global das mudanças na variável resposta. No entanto as análises estatísticas são mais difíceis, envolvem um maior custo computacional, um maior custo experimental, além de ser comum dados faltantes.

Na literatura destacam-se diferentes análises de dados longitudinais ou medidas repetidas. Wallace e Green (2002) apresentam um estudo que busca analisar as habilidades latentes perceptivas de crianças com 0, 3, 6 e 9 meses de idade através de modelos lineares mistos. Harvey (1987) analisa dados longitudinais através de modelos multiníveis. Já Diggle (2002) demonstra a utilização de modelos lineares generalizados com efeitos aleatórios, na análise de dados longitudinais. Lindsey (1993) descreve desde conceitos básicos a avançados utilizados na modelagem de dados com medidas repetidas, dados longitudinais contínuos e discretos, dados super dispersos, dados com heterogeneidade de variância e não normalidade, além de discutir problemas com dados faltantes. Raudenbush (2002) descreve alguns métodos para se lidar com dados hierárquicos. Longford

(1995) ilustra diversos problemas longitudinais apresentando para cada caso o melhor modelo a ser utilizado em sua solução.

Um estudo interessante que utiliza dados longitudinais foi proposto por Franco (2016), que procura encontrar associações biológicas entre espécies de laranja doce no desenvolvimento do cancro cítrico, uma doença que causa grandes prejuízos ao citricultor.

A laranja doce (*Citrus sinensis* (L.) OSbeck) surgiu na antiguidade pelo cruzamento entre a cimboa e a tangerina, e é utilizada desde o consumo *in natura*, em sucos e doces, até na produção de rações e biocombustível com seu bagaço (SILVA et al., 2015).

Um dos grandes problemas na indústria produtora de laranjas são as doenças infecciosas, as principais são: huanglongbing, cancro cítrico, clorose variegada dos citros, gomose de phytophthora, leprose dos citros, mancha de alternária, pinta preta, podridão floral dos citros, rubelose, entre outras (KIMATI et al., 1997). No presente trabalho é estudado o cancro cítrico, causado pela bactéria *Xanthomonas citri* subsp. *citri* (SCHAAD et al., 2006).

As lesões causados por tal bactéria aparecem nas folhas, ramos e frutos, iniciando com lesões necróticas, geralmente circundadas por um halo amarelo. Com o aumento da gravidade dos sintomas elas podem gerar rachaduras nos frutos, servindo de porta de entrada para outros patógenos, além de ocasionar queda do mesmo com o passar do tempo (LARANJEIRA et al., 2005). As manchas se tornam evidentes na superfície inferior das folhas entre 7 e 10 dias após o início da infecção, logo depois atingem também a superfície superior.



Figura 1 – Lesão de cancro cítrico foliar e no fruto.

A bactéria *X. citri* subsp. *citri* é um bastonete, gram-negativo com um único flagelo polar (2). As colônias em meio de cultura costumam apresentar coloração amarelada devido à produção de pigmentos xantomonadina (2).

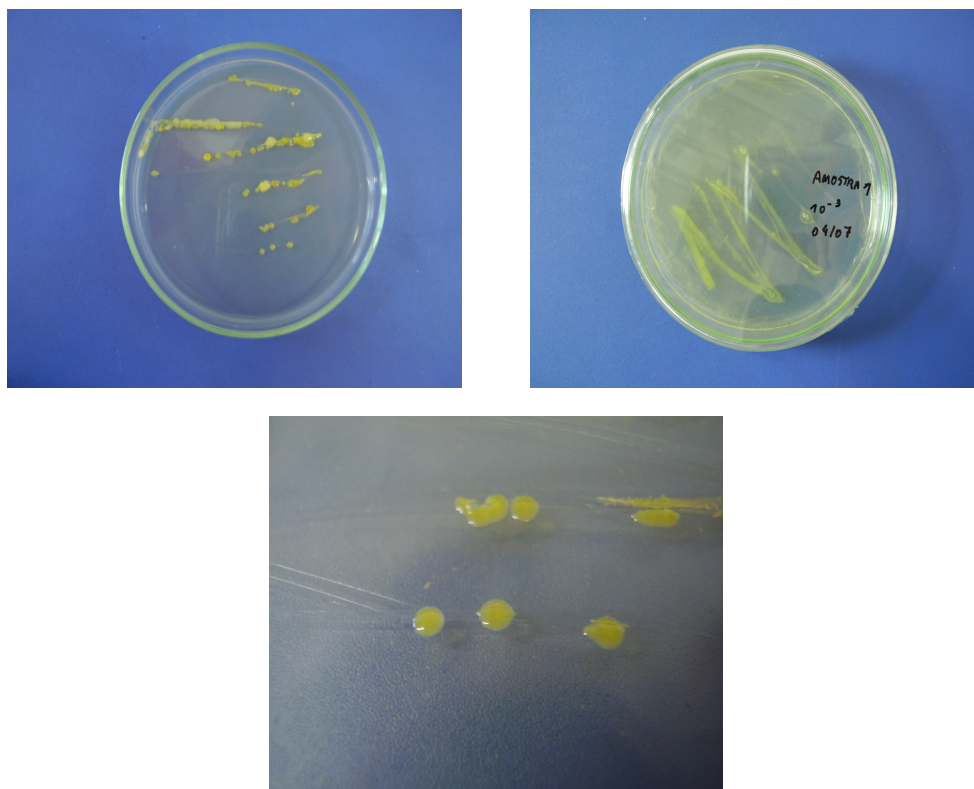


Figura 2 – Colônia em agar apresentando o pigmento amarelo e a aparência geral mucóide.

A suscetibilidade a infecção da *X. citri* pode variar amplamente entre as espécies cítricas, onde as que apresentam maior tempo de crescimento vegetativo são, em geral mais vulneráveis a infecção do patógeno, sendo que a resistência a bactéria está diretamente relacionada a juvenilidade do tecido (GOTTWALD; GRAHAM, 1992), (GRAHAM et al., 1992). De acordo com Gottwald et al. (1993), cultivares de citros com maior frequência, tamanho e duração do crescimento de folhas e frutos são mais suscetíveis a *X. citri*, quando comparado com cultivares menos vigorosos ou cujas folhas amadurecem mais rápido.

As condições ideais de infecção pela *X. citri* ocorrem entre temperaturas de 25°C a 30°C, com lâmina de água presente na superfície da folha (GOTO, 1992). Além disso, a *X. citri* cresce facilmente em meios de cultura utilizados em laboratório, podendo ser vistas após 48 horas de incubação. Os sintomas em plantas inoculadas, começam a ser visíveis após 10 dias (LARANJEIRA et al., 2005).

Para avaliações da patogenicidade de *X. citri* são realizados testes diagnósticos ou qualquer outro procedimento que tenha como objetivo a rápida obtenção de sintomas de cancro cítrico, é sugerido a inoculação do hospedeiro por ferimentos independentemente da concentração de inoculo utilizada ( $10^6$  a  $10^8$  UFC/mL) (JR; FILHO, 2006).

Para medir a suscetibilidade do tecido foliar, uma importante ferramenta é a avalia-



ção por meio de diâmetros de lesão, que verifica a interação entre a *Xanthomonas citri* × variedades (NOCITI et al., 2006). Estudos evidenciam que há correlação na comparação dos diâmetros de lesões em plantas mantidas em condições de campo e em casa de vegetação (GRAHAM; GOTTWALD et al., 1990). Da mesma forma, a técnica de inoculação das folhas é melhor realizada por ferimentos com agulhas, comparada a inoculação por aspersão ou infiltração (GRAHAM; GOTTWALD et al., 1990), (GRAHAM et al., 1992), (JR; FILHO, 2006), (NOCITI et al., 2006).

Muitos pesquisadores já utilizaram ferramentas estatísticas na análise de resultados obtidos em campo. Gonçalves-Zuliani et al. (2016) utilizou métodos não paramétricos para comparar médias populacionais de algumas medidas de interesse. Brugnara et al. (2012) utilizou regressões lineares para avaliar danos causados por doenças. Nociti et al. (2006) avalia a agressividade entre linhagens diferentes de um mesmo gênero de bactérias através, de regressão linear. Franco (2016) avaliou a resistência de genótipos de laranja doce (*Citrus sinensis*), variedade Pêra a bactéria *X. citri* subsp. *citri*, em condições parcialmente controladas (casa de vegetação), através de análise de variância e aplicação do teste de médias Scott-knott à 5% de significância.

A análise de regressão consiste de uma análise estatística utilizada para estimar a relação entre variáveis, avaliando a significância de dependência entre elas. Tal método é muito utilizado em análises empíricas, no entanto, em estudo com dados longitudinais envolvendo medidas repetidas, um modelo de regressão linear simples não é recomendado, pois ignora a existência de correlação, tratando as observações de forma independente, o que por sua vez acaba gerando estimadores dos parâmetros menos confiáveis, com erros-padrão normalmente viesados.

Um dos principais modelos estatísticos existentes para analisar dados longitudinais, que incorpora a dependência e a estrutura de correlação dos erros, são os modelos mistos (WALLACE; GREEN, 2002). Segundo Pinheiro (1994) observações feitas no mesmo indivíduo não possuem correlação nula e os modelos mistos são capazes de incorporar tal correlação na modelagem.

Um modelo misto é uma extensão de um modelo linear. Tal modelo é dividido em duas partes, uma de efeitos fixos (já presente nos modelos lineares) e outra de efeitos aleatórios, onde esta última tem a função de identificar a variação entre os indivíduos do problema em questão. Nesse sentido, enquanto os coeficientes fixos são constantes em todos os indivíduos, os aleatórios variam para cada um, o que por sua vez incorpora as diferenças existentes entre eles. Portanto, como tais modelos podem fazer com que os coeficientes da regressão variem entre os indivíduos, isso faz com que o modelo possua dois componentes de parametrização, um intra-indivíduo e outro entre indivíduo.

Como os modelos mistos permitem uma melhor descrição da correlação entre os indivíduos inerentemente associadas com a repetição das medidas, pode-se obter melhores



inferências estatísticas a respeito das performances médias dos indivíduos ([WALLACE; GREEN, 2002](#)).

Um dos principais motivos para se trabalhar com modelos mistos em relação aos lineares está em resolver o problema de não independência dos dados. Contudo, se os efeitos fixos e aleatórios do modelo não forem corretamente especificados, continuará violando suposição de independência. Se, por exemplo, fossem analisados dados com um modelo que não inclui o efeito aleatório "sujeito", então não há formas de se incorporar múltiplas respostas por sujeito. Portanto, para se resolver a não independência dos dados é necessário estipular corretamente seus efeitos fixos e aleatórios ([WINTER, 2013](#)).

De acordo com [Littell, Pendergast e Natarajan \(2000\)](#) os modelos mistos foram desenvolvidos por geneticistas para avaliar o potencial genético de touros. Com o avanço tecnológico computacional a aplicação de tais modelos espalhou-se por várias áreas de investigação. Antes de tais avanços as análises eram feitas adaptando-se métodos para modelos com efeitos fixos, o que as limitava, pois as estruturas de covariâncias não eram modeladas ([BARBOSA, 2009](#)).

Em tais modelos a estimação dos parâmetros é mais trabalhosa que a realizada em modelos lineares que contém apenas efeitos fixos, pois em tais modelos além dos parâmetros de efeitos fixos, também é de interesse conhecer as estimativas dos parâmetros de variância e covariância dos efeitos aleatórios e dos erros. Em modelos lineares que contém apenas efeitos fixos, os parâmetros fixos normalmente não podem ser estimados independente dos componentes de variância, pois as estimativas de efeitos fixos requerem estimativas dos componentes de variância e as estimativas para os componentes de variância exigem estimativas dos efeitos fixos. Consequentemente, os procedimentos de estimativa globais requerem uma iteração entre equações de efeitos fixos e equações de componentes de variância para se chegar a uma solução final ([WALLACE; GREEN, 2002](#)). Para se obter tais estimativas os dois principais métodos utilizados na estimação em modelos mistos são: Máxima Verossimilhança e Máxima Verossimilhança Restrita.

Na utilização de modelos mistos se destacam duas principais metodologias, a dos modelos lineares mistos, os quais são utilizados caso variável de interesse provenha de uma distribuição normal, e a de modelos lineares generalizados mistos, utilizados caso a variável de interesse provenha de qualquer distribuição pertencente a família exponencial (2.13).

A avaliação da resistência genética pela inoculação da folha com agulha e a avaliação do sintoma pela medida do diâmetro da lesão, permite realizar estudos mais rápidos, com um menor custo e maior número de plantas ou repetições, além de se ocupar um menor espaço, quando comparado com experimentos realizados no campo. No entanto, estudos em campo são sempre necessários, pois no campo podem ser avaliados aspectos relacionados ao crescimento do hospedeiro, sendo que em casa de vegetação o

crecimento é restrito (VILORIA et al., 2004).

A variável resposta abordada no presente estudo é o diâmetro da lesão nas folhas da planta causado pelo cancro cítrico, que foi avaliado em centímetros. Além disso o banco de dados apresenta mais três variáveis: a variedade das plantas, os dias após a inoculação da bactéria nas folhas da planta e o sujeito, onde esse último é dado pelo tipo de variedade em cada dia, após a inoculação. Tendo isso em vista, se deseja avaliar o comportamento da variável resposta com relação a esses outros três fatores, desejando-se compreender a influência de cada um deles no comportamento do diâmetro da lesão.

O capítulo 1 apresenta como foi realizado a implantação do experimento, desde quais variedades foram analisadas, como foram coletadas as amostras e o que foi utilizado na avaliação estatística. O capítulo 2 descreve os métodos pertinentes no trabalho, expondo o que são modelos lineares e generalizados mistos, assim como as possíveis estruturas da matriz de covariância que podem ser adotadas para eles, além de descrever detalhadamente as distribuições utilizadas na modelagem. O capítulo 3 apresenta toda a modelagem estatística utilizada no trabalho, desde a análise descritiva até a seleção do modelo final. O capítulo 4 apresenta todos os resultados e conclusões obtidas no trabalho.

## Capítulo 1

# Materiais

## 1.1 Implantação do Experimento

Para avaliar a suscetibilidade de laranjas doce (*Citrus sinensis*) ao cancro cítrico provocado pela bactéria *Xanthomonas citri* subsp. *citri* (SCHAAD et al., 2006), Franco (2016) executou um experimento envolvendo 12 variedades, conforme Tabela (1).

Tabela 1 – Variedades avaliadas

	Abrev.	Variedade
1	VALEN	Valência
2	Valen	Valência 2
3	Puka	Valência Puka
4	Paloma	Valência Paloma
5	Pera Ori	Perâ Oriçanga
6	Irradiada	Perâ Irradiada
7	Itapetininga	Perâ Itapetininga
8	Pera mga	Perâ Maringá
9	Pera IAC	Perâ IAC
10	Sanguine	Sanguinea Mombuca
11	Hamlin	Hamlin
12	Prec. Ori.	Precoce Oriçanga

Segundo Franco (2016) a estirpe de *Xanthomonas citri* subsp. *citri* utilizada na inoculação das plantas foi a Xcc 306, proveniente da coleção do Fundo de Defesa da Citricultura (Fundecitrus) e armazenada no laboratório de Biotecnologia do Núcleo de Pesquisa em Biotecnologia Aplicada (NBA) da Universidade Estadual de Maringá (UEM).

O experimento foi conduzido em casa de vegetação nas instalações do Núcleo

de Pesquisa em Biotecnologia Aplicada (NBA) (23°23'57,8"S; 51°57'5,3"O e aproximadamente 500 m de altitude).

O inóculo foi preparado a partir de uma cultura pura de *X. citri* (isolado Xcc 306) e mantida em tampão fosfato (0,075 M, pH 7,0) em geladeira.

Para reativar o organismo, a bactéria foi semeada em Placas-de-Petri contendo meio de cultura MGY (*Manitol Glutamato Yeast*). O inóculo foi preparado em tampão fosfato (0,075 M, pH 7,0), ajustado a  $10^8$  unidades formadoras de colônia/mL com auxílio de um espectrofotômetro ajustado para leitura a 600 nm (FRANCO, 2016).

As amostras foram obtida a partir da inoculação da bactéria em 6 folhas por planta, num total de 5 plantas por variedade, as quais receberam adubações periódicas de N,P,K. O diâmetro das lesões provocadas foi avaliado aos 18, 24, 33, 39, 46, 61, 68 e 76 dias após a inoculação (DAI), sendo que a presença de halo amarelo ao redor da necrose foi desconsiderado. Todas as 2880 amostras foram obtidas no ano de 2016. Foi utilizado como porta-enxerto o limão cravo (*Citrus limonia*).

O método de inoculação utilizado foi o ferimento do limbo foliar com agulha de  $0,55 \times 0,20$  mm, perfurando-se imediatamente após a imersão da mesma na suspensão bacteriana. Em cada folha foram realizadas oito perfurações subepidérmicas (atravesando o limbo foliar) e equidistantemente distribuídas (FRANCO, 2016).

Vinte e quatro horas após as inoculações foi realizado o molhamento em toda a casa de vegetação, porém sempre se atentando para não molhar as folhas inoculadas, também procurou-se manter sempre a casa de vegetação sob temperatura entre 13° e 35°C.

Foi realizada a quantificação bacteriana nas lesões de *X. citri* após cada avaliação. Para isto, as lesões foram cortadas em discos de aproximadamente 2,0 cm de diâmetro e maceradas 1 lesão por planta em  $1000\mu\text{l}$  de solução salina (8g de NaCl; 1 litro de água destilada). Então feitas as diluições seriadas, foram plaqueadas  $50\mu\text{l}$  da solução diluída a  $10^{-4}$  em meio MGY e após 48 horas procedeu-se contagem das UFCs (Unidades formadoras de colônia) de *Xanthomonas* por placa (FRANCO, 2016).

O conjunto de dados é do tipo longitudinal, formado por um total de 2880 observações contendo 4 variáveis, sendo elas, os variedades de laranja doce utilizadas no experimento (Geno), o número de plantas em cada variedade (folhas), os dias após a inoculação (DAI) e o diâmetro da lesão (Diam) sendo essa última a variável resposta, que é do tipo contínua, medida em centímetros.

O conjunto de dados é composto por observações de dois ensaios, sendo que o primeiro foi conduzido de fevereiro a abril de 2016 e o segundo de junho a agosto de 2016.

Tabela 2 – 1º Ensaio

	Variedades	Folhas	DAI	Diam
1	PeraOri	1	18	1,35
2	PeraOri	1	18	1,58
3	PeraOri	1	18	1,70
4	PeraOri	1	18	1,60
5	PeraOri	1	18	1,55
6	PeraOri	1	18	1,58
7	PeraOri	2	18	2,09
:	:	:	:	:
2874	Irradiada	59	75	1,41
2875	Irradiada	60	75	1,73
2876	Irradiada	60	75	2,12
2877	Irradiada	60	75	1,96
2878	Irradiada	60	75	1,61
2879	Irradiada	60	75	2,19
2880	Irradiada	60	75	1,99

Tabela 3 – 2º Ensaio

	Variedades	Folhas	DAI	Diam
1	PeraOri	1	18	2,00
2	PeraOri	1	18	2,03
3	PeraOri	1	18	2,02
4	PeraOri	1	18	1,75
5	PeraOri	1	18	1,96
6	PeraOri	1	18	1,95
7	PeraOri	2	18	1,61
:	:	:	:	:
2874	Irradiada	59	75	2,18
2875	Irradiada	60	75	2,25
2876	Irradiada	60	75	2,30
2877	Irradiada	60	75	2,26
2878	Irradiada	60	75	2,29
2879	Irradiada	60	75	2,18
2880	Irradiada	60	75	2,32

As Tabelas (2) e (3) apresentam um esboço do conjunto de dados, dispondo as sete primeiras e últimas observações de cada ensaio.

O estudo da resistência de espécies cítricas, que envolvem grande número de variedades podem ser facilitadas através de avaliações prévias em condições controladas (FRANCO, 2016). Nesse sentido o estudo tem por objetivo avaliar a resistência de 12 variedades de laranja doce (*Citrus sinensis*) a bactéria *X. citri* subsp. *citri*, em condições parcialmente controladas (casa de vegetação), por inoculação da folha com agulha e avaliação do diâmetro da lesão através de modelos generalizados mistos.

## Capítulo 2

# Métodos

### 2.0.1 Modelo Linear Misto (LMM)

Modelos mistos para dados contínuos com resposta normal, veem sendo intensivamente desenvolvidos desde a publicação do artigo de (LAIRD; WARE, 1982). Tais modelos são recomendados para descrever dados longitudinais, pois permitem descrever a correlação intra indivíduos inerentemente associada com as medidas repetidas, isso permite obter inferências estatísticas melhores a respeito do comportamento médio dos indivíduos no experimento.

Um modelo linear misto com  $N$  unidades experimentais e  $n_i$  observações feitas na  $i$ -ésima unidade experimental, pode ser escrito da seguinte forma matricial:

$$y_i = X_i\beta + Z_ib_i + e_i, \quad i = 1, \dots, N, \quad (2.1)$$

em que  $y_i$  é uma matriz ( $n_i \times 1$ ) da variável resposta da  $i$ -ésima unidade experimental ou sujeito,  $X_i$  é uma matriz ( $n_i \times p$ ) de posto completo das variáveis explicativas dos efeitos fixos,  $\beta$  é uma matriz ( $p \times 1$ ) dos parâmetros fixos,  $Z_i$  é uma matriz ( $n_i \times q$ ) de posto completo das variáveis explicativas dos efeitos aleatórios,  $b_i$  é um vetor ( $q \times 1$ ) de parâmetros de efeitos aleatórios.  $e_i$  é uma matriz ( $n_i \times 1$ ) dos erros aleatórios, sendo que:

$$b_i \sim N_q(0, D) \quad (2.2)$$

$$e_i \sim N_{n_i}(0, R_i), \quad i = 1, 2, \dots, N, \quad (2.3)$$

onde cada  $b_i$  e  $e_i$  são mutuamente independentes,  $D$  é uma matriz de variâncias e covariâncias ( $q \times q$ ),  $R_i$  é uma matriz de variâncias e covariâncias ( $n_i \times n_i$ ) que depende de  $i$ .

De forma geral, assumindo uma estrutura não-estruturada (2.25), para qualquer conjunto de variáveis  $Z_1, Z_2, \dots, Z_n$ , a matriz de variâncias e covariâncias dos efeitos aleatórios pode ser definida como:

$$D = Var \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_N \end{bmatrix} = \begin{bmatrix} \sigma^2(b_1) & \sigma(b_1, b_2) & \dots & \sigma(b_1, b_N) \\ \sigma(b_2, b_1) & \sigma^2(b_2) & \ddots & \vdots \\ \vdots & \ddots & \ddots & \sigma(b_{N-1}, b_N) \\ \sigma(b_N, b_1) & \dots & \sigma(b_N, b_{N-1}) & \sigma^2(b_N) \end{bmatrix}; \quad (2.4)$$

em que  $\sigma^2(b_i)$  é a variância de  $b_i$  e  $\sigma(b_i, b_j)$  é a covariância de  $b_i$  de  $b_j$ . Quando os  $b_i$ 's forem independentes e tiverem variâncias iguais tal matriz será diagonal.

Tipicamente a matriz de variâncias e covariâncias dos erros  $R_i$  é definida como segue:

$$R_i = Var \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_N \end{bmatrix} = \begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \dots & \sigma^2 \end{bmatrix}; \quad (2.5)$$

em que  $\sigma^2$  é a variâncias dos erros. Como será visto tanto a estrutura da matriz  $D$  quanto da matriz  $R_i$  podem ser especificadas de diversas formas, sendo utilizadas individualmente ou em conjunto, que é a forma mais comum, dado por (2.8).

Em um contexto envolvendo dados longitudinais com medidas repetidas o modelo (2.1) pode ser interpretado da seguinte forma:

$Y_i$  é uma matriz contendo os valores de  $y_{ij}$  para o  $i$ -ésimo indivíduo no  $j$ -ésimo momento:

$$y_{ij}^T = [y_{11}, y_{12}, \dots, y_{1n_1}, y_{21}, y_{22}, \dots, y_{2n_2}, \dots, y_{Ln_L}], \quad (2.6)$$

em que o sujeito  $i$  é observado  $n_i$  vezes, sendo  $L$  o total de indivíduos, logo o total de observações é igual a  $N = \sum_{i=1}^L n_i$ .

Tem-se que  $X$  é uma matriz de especificação de dimensão  $(N \times p)$  da parte fixa do modelo, contendo  $N$  valores para os  $p$  preditores assumidos como medidos sem erros,  $\beta$  é um vetor  $(p \times 1)$  de parâmetros desconhecidos da população,  $\beta_0, \dots, \beta_{p-1}$ , onde estes são os parâmetros de predição dos sujeitos, ou seja, os parâmetros que descrevem seus comportamentos.

Tem-se também  $Z$  é uma matriz de incidência  $(n_i \times q)$  das co-variáveis do  $i$ -ésimo sujeito no  $j$ -ésimo momento, pertencente a parte aleatória do modelo e que contém  $q$

preditores. Logo  $Z$  é uma matriz bloco diagonal ( $N \times Lq$ ) da parte aleatória, onde a sua diagonal é composta pelas matrizes  $Z_1, Z_2, \dots, Z_L$ .

Tem-se que  $b_i = [b_{0i}, b_{1i}, \dots, b_{(q-1)i}]$  é um vetor ( $q \times 1$ ) desconhecido associado a  $Z_i$ . Cada  $b_i$  é um parâmetro de efeito aleatório específico e representa as diferenças nos coeficientes dos  $q$  preditores para o indivíduo  $i$ . Por fim  $e$  é um vetor ( $N \times 1$ ) contendo os erros aleatórios intra-indivíduos,  $e_{ij}$ :

$$e^T = [e_{11}, e_{12}, \dots, e_{1n_1}, e_{21}, e_{22}, \dots, e_{2n_2}, \dots, e_{Ln_L}]. \quad (2.7)$$

Em tais modelos os erros aleatórios são independentes e utilizados na determinação dos componentes da variância. A partir dos erros das estimativas da média intra-individual, que apresentam dependência entre si, é determinada a covariância. A inclusão da parte aleatória em um modelo se deve ao fato de que parte da variabilidade dos dados pode ser provocada por fatores que são medidos nas co-variáveis, entre e intra-indivíduos. Isso acontece muitas vezes por conta de efeitos de confundimento desconhecidos, omitidos ou não avaliados durante o estudo.

De acordo com [Wallace e Green \(2002\)](#) os parâmetros de um modelo que precisam ser estimados são os de efeitos fixos  $\beta$ , as variâncias e covariâncias dos efeitos aleatórios  $D$  e as variâncias e covariâncias dos erros  $R$ . Os efeitos aleatórios  $b$  não são parâmetros do modelo e geralmente não são de interesse do pesquisador, pois eles são específicos a cada indivíduo particularmente.

A matriz de variância covariância de  $y_i$  é escrita como:

$$\text{cov}(y_i) = V_i = Z_i D Z_i^T + R_i. \quad (2.8)$$

Vê-se com isso que a correlação entre as observações é influenciada pela variabilidade entre os indivíduos  $Z_i D Z_i^T$ , e/ou, pela matriz de covariância intra-indivíduos  $R_i$ .

Segundo [Barbosa \(2009\)](#) o modelo linear misto é geralmente especificado em termos das respostas condicionadas aos efeitos aleatórios, de modo que assumindo que  $y_i$  é uma matriz de medidas repetidas para o  $i$ -ésimo indivíduo, é satisfeito:

$$y_i | b_i \sim N(X_i \beta + Z_i b_i, R_i) \quad (2.9)$$

$$b_i \sim N(0, D). \quad (2.10)$$

Além disso:



$$\begin{aligned} y_i &\sim N(X_i\beta, Z_i D Z_i^T + R_i) \\ y_i &\sim N(X_i\beta, V_i). \end{aligned} \quad (2.11)$$

A média de  $y_i$  é dada por  $\mathbb{E} = X_i\beta$  e representa a média sobre os efeitos aleatórios não correspondentes as medidas longitudinais individuais, pois as correspondentes a tais medidas é dada pelos próprios efeitos aleatórios  $b_i$  (MANCO, 2013).

## 2.0.2 Modelo Linear Generalizado Misto (GLMM)

Um modelo generalizado misto é um modelo estatístico que busca descrever o comportamento de uma variável aleatória com relação a uma ou mais variáveis explicativas, onde estas podem ser tanto de efeitos fixos como aleatórios. A grande diferença desses modelos com relação a um modelo linear misto, é que nos generalizados a variável resposta não precisa estar necessariamente associada a distribuição normal, o que amplia a abrangência de tais modelos.

Para se especificar um modelo generalizado misto, a variável resposta condicionada aos efeitos aleatórios ( $y_i|b$ ) deve seguir uma distribuição pertencente a família exponencial:

$$y_i|b \sim f_{Y_i|b}(y_i|\theta) \quad (2.12)$$

$$f_{Y_i|b}(y_i|\theta) = b(y_i) \exp\{\eta(\theta)^T T(y_i) - a(\theta)\}. \quad (2.13)$$

Tipicamente se deseja relacionar os parâmetros da distribuição com os preditores do modelo, para isso é feito uma transformação da média  $\mu_i$ , para que a mesma se relacione com tais preditores:

$$\mathbb{E}(y_i|b) = \mu_i \quad (2.14)$$

$$g(\mathbb{E}(y_i|b)) = g(\mu_i) = \eta_i = X_i\beta + Z_i b, \quad (2.15)$$

em que  $g(\cdot)$  é uma função conhecida, chamada função de ligação, responsável por relacionar a média de  $y_i$  com os preditores lineares,  $X_i$  é a  $i$ -ésima linha da matriz modelo dos efeitos fixos,  $\beta$  é o vetor de parâmetros dos efeitos fixos,  $Z_i$  é a  $i$ -ésima linha da matriz modelo dos efeitos aleatórios e  $b$  é o vetor dos efeitos aleatórios. A variância de  $y_i|b$  é dada por  $Var(y_i|b) = \phi^2 v(\mu_i)$ , em que  $v$  é uma função de variância conhecida, dada pela distribuição de  $y_i$  e  $\phi$  é o parâmetro de escala da distribuição, o qual precisa ser estimado.

Em tal contexto pode-se escrever a média, variância, covariância e correlação como:

- Média:

$$\begin{aligned}\mathbb{E}[y_i] &= \mathbb{E}[\mathbb{E}[y_i|b]] \\ &= \mathbb{E}[\mu_i] \\ &= \mathbb{E}[g^{-1}(X_i\beta + Z_i b)];\end{aligned}\tag{2.16}$$

- Variância:

$$\begin{aligned}Var(y_i) &= Var(\mathbb{E}[y_i|b]) + \mathbb{E}[Var(y_i|b)] \\ &= Var(\mu_i) + \mathbb{E}[\phi^2 v(\mu_i)] \\ &= Var(g^{-1}(X_i\beta + Z_i b)) + \mathbb{E}[\phi^2 v(g^{-1}(X_i\beta + Z_i b))];\end{aligned}\tag{2.17}$$

- Covariância e Correlação:

$$\begin{aligned}cov(y_i, y_j) &= cov(\mathbb{E}[y_i|b], \mathbb{E}[y_j|b]) + \mathbb{E}[cov(y_i, y_j|b)] \\ &= cov(\mu_i, \mu_j) + \mathbb{E}[0] \\ &= cov(g^{-1}(X_i\beta + Z_i b), g^{-1}(X_j\beta + Z_j b)) \\ cov(y_i, y_j) &= \frac{cov(g^{-1}(X_i\beta + Z_i b), g^{-1}(X_j\beta + Z_j b))}{\sqrt{Var(y_i)}\sqrt{Var(y_j)}}.\end{aligned}\tag{2.18}$$

Nos GLMM's tais como nos modelos lineares mistos existem dois tipos de efeitos aleatórios. Colocando de forma simples, se um efeito aleatório é um elemento de  $b$ , ele é tido como um efeito  $D$ , ou seja, em tal modelo está sendo modelado a estrutura de covariância da matriz  $D$ , caso contrário, está sendo modelado a estrutura de covariâncias dos erros  $R$ . Modelos sem efeitos  $D$  também são conhecidos como modelos marginais. Um mesmo modelo pode conter os dois tipos de efeitos aleatórios, quando isso ocorre a matriz de covariância de  $Y$  é escrita como:

$$cov(y) = V = ZDZ^T + R$$

Nos modelos generalizados mistos, assim como nos lineares, são estimados os parâmetros de efeitos fixos  $\beta$ , as variâncias e covariâncias dos efeitos aleatórios  $D$ . Porém, diferentemente dos modelos lineares nos GLMM's a função de verossimilhança, tipicamente, não possui uma expressão de forma fechada, o que gera integrais que não podem ser resolvidas analiticamente. Por conta disso, diversas aproximações podem ser utilizadas afim de contornar tais dificuldades. Uma delas é feita pela maximização da função de verossimilhança penalizada  $l_p$  (RIGBY; STASINOPOULOS, 2005), dada por

$$l_p(\beta, b) = l - \frac{1}{2} \sum_{i=1}^N b_{i\mu}^T (D_{i\mu}^{-1}) b_{i\mu} + b_{i\sigma}^T (D_{i\sigma}^{-1}) b_{i\sigma}, \quad (2.19)$$

onde  $l = \sum_{i=1}^n \ln(f(y_i|\mu, \sigma^2))$  é a função log-verossimilhança dos dados,  $b$  são os parâmetros de efeitos aleatórios e  $D_i^{-1}$  é a matriz inversa de  $D_i$  a matriz bloco diagonal ( $q \times q$ ) de variâncias e covariâncias dos efeitos aleatórios  $b_i$ . Pode-se mostrar que maximizar tal função é equivalente a maximizar a função de verossimilhança hierárquica ((PAWITAN, 2001), página 429 e (LEE; NELDER, 1996)) definida como:

$$l_h = l_p + \frac{1}{2} \sum_{i=1}^N \log |D_i| - q \log(2\pi). \quad (2.20)$$

### 2.0.3 Estruturas da Matriz de Covariância

Uma parte importante no estudo de modelos mistos se resume a modelar a matriz de covariância  $V_i$  (2.8). Diggle (2002) nota que a matriz de covariância deve incluir:

- i) A variação devida a efeitos aleatórios, quando as unidades de investigação formam uma amostra aleatória da população de interesse;
- ii) a variação que pode ser explicada por correlação serial, em que se esperam observações próximas mais fortemente correlacionadas que observações mais distantes;
- iii) a variação devida a erros de medida.

Modelar a estrutura de covariância refere-se a representar  $V_i$  como uma função de uma determinada quantidade de parâmetros (LITTELL; PENDERGAST; NATARAJAN, 2000). Como variância de  $y_i$  é feita a partir de dois componentes,  $ZDZ^T$  e  $R$ , é possível modelar a estrutura de  $D$  ou  $R$  ou de ambas. Entre as principais estruturas da matriz de covariância disponíveis na literatura, algumas estão listadas abaixo em uma dimensão ( $4 \times 4$ ):

- Simetria Composta (CS):

$$V_i = \begin{bmatrix} \sigma^2 + \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma^2 + \sigma_1^2 & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & \sigma^2 + \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma^2 + \sigma_1^2 \end{bmatrix}; \quad (2.21)$$

- Estrutura AR(1):

$$V_i = \sigma^2 \begin{bmatrix} 1 & \phi & \phi^2 & \phi^3 \\ \phi & 1 & \phi & \phi^2 \\ \phi^2 & \phi & 1 & \phi \\ \phi^3 & \phi^2 & \phi & 1 \end{bmatrix}; \quad (2.22)$$

- Estrutura ARMA(1,1):

$$V_i = \sigma^2 \begin{bmatrix} 1 & \gamma & \gamma\phi & \gamma\phi^2 \\ \gamma & 1 & \gamma & \gamma\phi \\ \gamma\phi & \gamma & 1 & \gamma \\ \gamma\phi^2 & \gamma\phi & \gamma & 1 \end{bmatrix}; \quad (2.23)$$

- Uniforme:

$$V_i = \sigma^2 \begin{bmatrix} 1 & \phi & \phi & \phi \\ \phi & 1 & \phi & \phi \\ \phi & \phi & 1 & \phi \\ \phi & \phi & \phi & 1 \end{bmatrix}; \quad (2.24)$$

- Não-estruturada (NE):

$$V_i = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} & \sigma_{24} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 & \sigma_{34} \\ \sigma_{14} & \sigma_{24} & \sigma_{34} & \sigma_4^2 \end{bmatrix}; \quad (2.25)$$

Para a maioria dessas estruturas, a covariância entre duas observações no mesmo indivíduo depende apenas do comprimento do intervalo de tempo entre as medições, além disso a variância é considerada constante sobre o tempo. A covariância pode ser caracterizada em termos da variância e as correlações expressadas como função do intervalo de tempo entre as medições (LITTELL; PENDERGAST; NATARAJAN, 2000).

#### 2.0.4 Algoritmos de estimação dos parâmetros via pacote *gamlss*

- CG: É uma generalização do algoritmo de Cole e Green (1992) que utiliza a primeira e segunda derivada transversal, esperada ou aproximada, da função de verossimilhança com respeito ao parâmetro  $\theta = (\mu, \sigma, \nu, \tau)$  da função densidade. No caso totalmente paramétrico, o algoritmo CG corresponde ao Método Escore de Fisher;

- RS: Para muitas funções de densidades de probabilidade  $f(y|\theta)$  onde os parâmetros  $\theta$  são ortogonais, como no caso da log-normal, em que os valores esperados da matriz de derivadas de segunda ordem da função de verossimilhança são zeros, exceto na diagonal principal, o RS é utilizado por ser mais simples. Tal algoritmo é uma generalização do algoritmo [Rigby e Stasinopoulos \(1996\)](#) que ajusta a média e dispersão de modelos aditivos (e não usa derivadas transversais), e portanto é mais adequado;
- mixed: É uma mistura dos dois algoritmos explicados acima, o qual começa com o algoritmo RS e termina com o CG.

No presente trabalho apenas o algoritmo RS foi utilizado. Tal algoritmo é entendido como o trabalho em conjunto de três componentes: a iteração externa, a iteração interna e o retroajuste modificado. Existe uma estrutura de aninhamento entre as três componentes, isto é, a iteração externa repetidamente utiliza os resultados da iteração interna que, por sua vez, utiliza os resultados do algoritmo de retroajuste modificado. A convergência do algoritmo RS é obtida quando a convergência das três componentes é alcançada.

### 2.0.5 Seleção do Modelo

O modelo final, considerado mais adequado, capaz de explicar satisfatoriamente a variável resposta com relação as preditoras, será determinado a partir da análise dos resíduos, da medida do Desvio Global Ajustado (*Global Deviance*) e o Critério de Informação de Akaike Generalizado (GAIC), onde quanto menor o valor dessas medidas menor o erro do modelo e portanto melhor o ajuste. Tais medidas são definidas como segue:

$$GD = -2l(\hat{\theta}) \quad (2.26)$$

$$GAIC(\omega) = GD + \omega df. \quad (2.27)$$

No qual  $\omega$  é uma penalidade fixa para cada grau de liberdade utilizado no modelo,  $df$  representa o total de graus de liberdade utilizados no modelo. Temos que quando  $\omega = 2$   $GAIC = AIC$ , enquanto que quando  $\omega = \log(n)$   $GAIC = BIC$  que é o Critério Bayesiano de Schwarz ([RIGBY; STASINOPOULOS, 2010](#)).

Já para análise de resíduos será utilizado os resíduos quantílicos aleatorizados, proposto por [Dunn e Smyth \(1996\)](#). Tais resíduos devem possuir distribuição normal padrão, para variável resposta com qualquer distribuição de probabilidade, caso o modelo seja adequado. Portanto dado uma distribuição densidade de probabilidade  $f(y|\theta)$  ajustado a variável resposta  $y_i, i = 1, 2, \dots, n$ , os resíduos quantílicos aleatorizados são determinados por:

$$\hat{r}_i = \Phi^{-1}(u_i), \quad (2.28)$$

onde  $\Phi^{-1}$  é a função inversa da distribuição acumulada da distribuição normal padrão e os  $u_i$ 's são os resíduos quantílicos, definidos da seguinte forma:

- Caso contínuo: Quando a variável resposta segue uma distribuição contínua,  $u_i = F(y_i|\theta_i)$ , em que  $F(y_i|\theta_i)$  é a distribuição de probabilidade acumulada do modelo.
- Caso discreto: Caso a variável resposta siga uma distribuição discreta  $u_i$  é um valor aleatório de uma distribuição uniforme no intervalo  $[F(y_i - 1|\hat{\theta}_i), F(y_i|\hat{\theta}_i)]$ .

Uma ferramenta que pode ser utilizada para verificar a adequabilidade dos resíduos é o *Worm plot*, em que esperamos ver os pontos (que representam os resíduos) o mais próximo possível da linha central e com 95% deles dentro das curvas elípticas que delimitam a região de confiança. Caso mais de 5% dos pontos estejam fora das duas curvas elípticas, isso indica que o modelo ajustado não é adequado para explicar a variável resposta (BUUREN; FREDRIKS, 2001).

## 2.1 Adequabilidade do modelo

Existem diversas medidas para se avaliar a qualidade do ajuste, no presente trabalho foi utilizado o  $R^2$  generalizado de Nagelkerke, dado por (NAGELKERKE, 1991):

$$R^2 = 1 - \left( \frac{L(0)}{L(1)} \right)^{2/n}.$$

Tal definição algumas vezes é referida como  $R^2$  de Cox & Snell. Outra formulação possível é conhecida como Nagelkerke/Cragg & Uhler's:

$$R^2 = \frac{1 - \left( \frac{L(0)}{L(1)} \right)^{2/n}}{1 - L(0)^{2/n}}.$$

Onde  $L(0)$  é o valor da verossimilhança para o modelo nulo (apenas uma constante é ajustada para todos os parâmetros) e  $L(1)$  para o modelo ajustado em questão.

Tais pseudos  $R^2$  não podem ser interpretados como na regressão linear, ou seja, eles não indicam o quanto da variabilidade dos dados está sendo explicada pelo modelo, eles são apenas valores que indicam a qualidade do ajuste, em que quanto mais próximos de 1 melhor o ajuste do modelo.

## 2.2 Distribuição Log-Normal

A distribuição log-normal é uma distribuição apropriada para dados contínuos, positivos e assimétricos, cuja assimetria ocorre frequentemente quando as médias são pequenas, as variâncias são grandes e os valores não são negativos (LIMPERT; STAHEL; ABBT, 2001), onde o logaritmo da variável aleatória em questão é normalmente distribuído. Sua função densidade de probabilidade pode ser definida como (STASINOPOULOS; RIGBY; AKANTZILIOTOU, 2008):

$$f(y|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{y} \exp \left\{ -\frac{(\log(y) - \mu)^2}{2\sigma^2} \right\}, \quad (2.29)$$

para  $y > 0$ , onde  $\mu > 0$  e  $\sigma > 0$ . Tem-se  $\mathbb{E}(Y) = \omega^{1/2}e^\mu$  e  $Var(Y) = \omega(\omega - 1)e^{2\mu}$ , com  $\omega = \exp(\sigma^2)$ , cujo suporte é dado por  $supp(f) = x \in (0, \infty)$ .

Escrevendo tal função na forma da família exponencial (2.13) tem-se:

$$\begin{aligned} f(y|\mu, \sigma^2) &= \exp \left( \log \left( \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{y} \exp \left\{ -\frac{(\log(y) - \mu)^2}{2\sigma^2} \right\} \right) \right) \\ &= \exp \left( \log \left( \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{y} \right) + \left( -\frac{\log^2(y)}{2\sigma^2} + \frac{2\mu \log(y)}{2\sigma^2} - \frac{\mu^2}{2\sigma^2} \right) \right) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{y} + \exp \left\{ -\frac{\log^2(y)}{2\sigma^2} + \frac{\mu \log(y)}{\sigma^2} - \frac{\mu^2}{2\sigma^2} \right\}, \end{aligned} \quad (2.30)$$

em que  $b(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{y}$ ,  $\eta(\mu, \sigma^2) = \left( \frac{-1}{2\sigma^2}, \frac{\mu}{\sigma^2} \right)^T$ ,  $T(y) = (\log^2(y), \log(y))^T$  e  $a(\mu, \sigma^2) = \frac{\mu^2}{2\sigma^2}$ .

Como tal distribuição possui uma relação próxima com a distribuição normal, onde se  $\ln(Y)$  é normalmente distribuída, tem-se que  $Y$  é log-normalmente distribuída, o parâmetro  $\mu$  da equação (2.29) pode ser interpretado como a média do logaritmo da variável aleatória, enquanto o parâmetro  $\sigma$  pode ser interpretado como sendo o desvio-padrão do logaritmo da variável aleatória. Adicionalmente,  $\mu$  é o parâmetro de escala, enquanto que  $\sigma$  é do de forma.

Na Figura (3) é mostrado alguns dos possíveis formatos que a distribuição log-normal (2.29) pode assumir, conforme a variação dos seus dois parâmetros, destacando seu formato simétrico e assimétrico a direita.

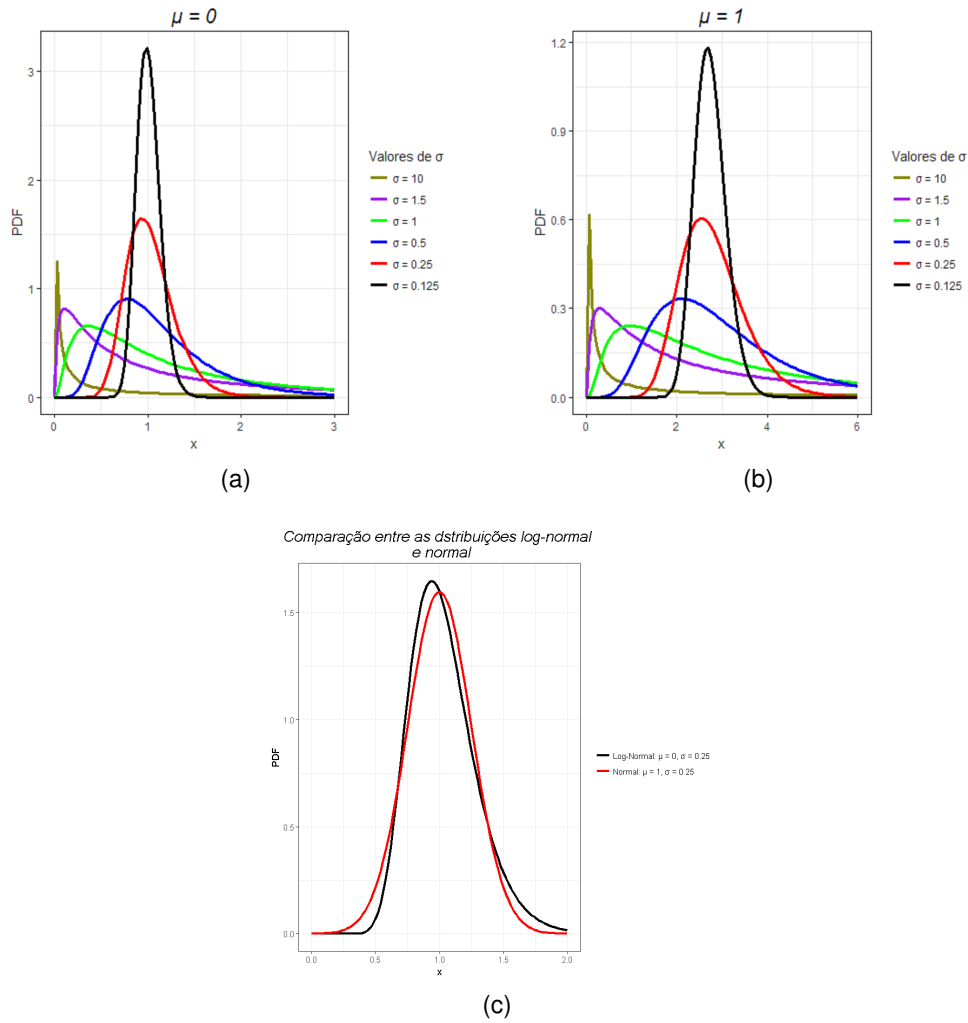


Figura 3 – (a) Distribuição log-normal para  $\mu = 0$  e  $\sigma = 10; 1, 5; 1; 0, 5; 0, 25; 0, 125$ ; (b) Distribuição log-normal para  $\mu = 1$  e  $\sigma = 10; 1, 5; 1; 0, 5; 0, 25; 0, 125$ ; (c) Comparação entre as distribuições log-normal e normal

Na Figura (3) pode-se notar uma importante propriedade da distribuição log-normal, que é o fato de que para valores de  $\sigma$  menores que 0, 3, a distribuição log-normal é quase indistinguível da distribuição normal (GINOS, 2009).

O momento de ordem  $k$  de uma variável aleatória  $Y$  com distribuição  $LN(\mu, \sigma)$  é determinado pela igualdade  $\mu_k = \mathbb{E}[Y^k]$ , como segue:

$$\begin{aligned}
 \mu_k &= \mathbb{E}[Y^k] \\
 &= \int_{-\infty}^{\infty} y^k f_{\ln N}(y; \mu, \sigma) dy \\
 &= \int_{-\infty}^{\infty} y^k \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{y} \exp \left\{ -\frac{(\log(y) - \mu)^2}{2\sigma^2} \right\} dy \\
 &= e^{k\mu + \frac{k^2\sigma^2}{2}}.
 \end{aligned} \tag{2.31}$$



O valor esperado de uma realização da variável aleatória  $Y$  com distribuição  $LN(\mu, \sigma^2)$  é o valor da equação (2.31) quando  $k = 1$ , isto é:

$$\mu_1 = \mathbb{E}[Y] = e^{\mu + \sigma^2/2}. \quad (2.32)$$

Com base na igualdade  $Var(Y) = \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2$ , pode-se determinar a variância de  $Y$ :

$$\begin{aligned} \mu_2 = \mathbb{E}[Y^2] &= e^{2\mu + \frac{2^2\sigma^2}{2}} \\ &= e^{2\mu + 2\sigma^2}. \end{aligned} \quad (2.33)$$

Portanto:

$$\begin{aligned} Var(Y) &= \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2 \\ &= e^{2\mu + 2\sigma^2} - (e^{\mu + \sigma^2/2})^2 \\ &= (e^{\sigma^2} - 1)e^{2\mu + \sigma^2}. \end{aligned} \quad (2.34)$$

Uma importante propriedade da distribuição log-normal é que se duas variáveis aleatórias independentes  $Y_1$  e  $Y_2$ , são log-normalmente distribuídas, então o produto  $Y_1 \times Y_2$  é distribuído por  $LN(\mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2})$ . Tal propriedade multiplicativa deriva da propriedade aditiva da distribuição normal (GINOS, 2009).

### 2.2.1 Estimação dos Parâmetros da distribuição Log-Normal

A estimação dos parâmetros dos modelos mistos geralmente é feita por meio dos métodos de máxima verossimilhança e máxima verossimilhança restrita, com o respaldo de que o método de máxima verossimilhança, assume que não existe erro na estimativa da média e portanto a subestima ao estimar a variância. Já com método de máxima verossimilhança restrita não ocorre o mesmo, produzindo estimativas da variância não viciadas, removendo o vício existente ao se estimar a média (DIGGLE, 2002).

Um dos grandes problemas na estimação de modelos mistos é que, além de ser necessário estimar os  $\beta's$ , também é preciso estimar os parâmetros de variância e covariância em  $D$  e  $R_i$ . A estimação dos  $\beta's$  é baseada no inverso da matriz  $V$ , conseqüentemente a estimação é geralmente baseada em métodos baseados na verossimilhança, assumindo que  $b_i$  e  $e_i$  são normalmente distribuídos.

Os métodos comumente utilizados para estimar os parâmetros de uma distribuição log-normal são o de Máxima Verossimilhança e o Método dos Momentos.

Para calcular os estimadores de Máxima Verossimilhança, primeiramente determina-se a verossimilhança da função log-normal para uma variável aleatória  $Y_i$ , com  $i = 1, 2, \dots, n$  (número de observações), que é o produto das probabilidades individuais dos  $Y_i$ 's:

$$\begin{aligned} L(\mu, \sigma^2 | Y) &= \prod_{i=1}^n [f(y_i | \mu, \sigma^2)] \\ &= \prod_{i=1}^n \left[ \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{y_i} \exp \left\{ -\frac{(\log(y_i) - \mu)^2}{2\sigma^2} \right\} \right] \\ &= (2\pi\sigma^2)^{-n/2} \prod_{i=1}^n \left[ y_i^{-1} \exp \left[ \sum_{i=1}^n \frac{-(\log(y_i) - \mu)^2}{2\sigma^2} \right] \right]. \end{aligned} \quad (2.35)$$

A função log-verossimilhança da log-normal é dada pela aplicação do logaritmo natural na função de verossimilhança:

$$\begin{aligned} l(\mu, \sigma^2 | Y) &= \ln \left( (2\pi\sigma^2)^{-n/2} \prod_{i=1}^n y_i^{-1} \exp \left[ \sum_{i=1}^n \frac{-(\log(y_i) - \mu)^2}{2\sigma^2} \right] \right) \\ &= -\frac{n}{2} \ln(2\pi\sigma^2) - \sum_{i=1}^n \ln(y_i) - \frac{\sum_{i=1}^n (\ln(y_i) - \mu)^2}{2\sigma^2} \\ &= -\frac{n}{2} \ln(2\pi\sigma^2) - \sum_{i=1}^n \ln(y_i) - \frac{\sum_{i=1}^n [\sum_{i=1}^n \ln(y_i)^2 - 2 \ln(y_i)\mu + \mu^2]}{2\sigma^2} \\ &= -\frac{n}{2} \ln(2\pi\sigma^2) - \sum_{i=1}^n \ln(y_i) - \frac{\sum_{i=1}^n (\ln(y_i))^2}{2\sigma^2} + \frac{\sum_{i=1}^n 2 \ln(y_i)\mu}{2\sigma^2} - \frac{\sum_{i=1}^n \mu^2}{2\sigma^2} \\ &= -\frac{n}{2} \ln(2\pi\sigma^2) - \sum_{i=1}^n \ln(y_i) - \frac{\sum_{i=1}^n (\ln(y_i))^2}{2\sigma^2} + \frac{\sum_{i=1}^n 2 \ln(y_i)\mu}{2\sigma^2} - \frac{n\mu^2}{2\sigma^2}. \end{aligned} \quad (2.36)$$

Maximizando a função  $l(\mu, \sigma^2 | Y)$  encontra-se os estimadores de  $\mu$  e  $\sigma$ , para isso deriva-se tal função e a resolve com relação a  $\mu$ :

$$\begin{aligned} U_\mu(\mu, \sigma^2 | Y) &= \frac{\partial l}{\partial \mu} \\ &= \frac{\sum_{i=1}^n \ln(Y_i)}{\hat{\sigma}^2} - \frac{2n\hat{\mu}}{2\hat{\sigma}^2} = 0 \\ \Rightarrow \frac{n\hat{\mu}}{\hat{\sigma}^2} &= \frac{\sum_{i=1}^n \ln(Y_i)}{\hat{\sigma}^2} \\ \Rightarrow \hat{\mu} &= \frac{\sum_{i=1}^n \ln(Y_i)}{n}. \end{aligned} \quad (2.37)$$

Como já dito anteriormente, o estimador de  $\mu$  nada mais é que a média do logaritmo dos dados. Com relação a  $\sigma^2$  temos:

$$\begin{aligned}
U_{\sigma}(\mu, \sigma^2|Y) &= \frac{\partial l}{\partial \sigma^2} \\
&= -\frac{n}{2\hat{\sigma}^2} - \frac{\sum_{i=1}^n (\ln(Y_i) - \hat{\mu})^2}{2} \cdot (-\hat{\sigma}^2)^{-2} \\
&= -\frac{n}{2\hat{\sigma}^2} + \frac{\sum_{i=1}^n (\ln(Y_i) - \hat{\mu})^2}{2(\hat{\sigma}^2)^2} = 0 \\
&\Rightarrow \frac{n}{2\hat{\sigma}^2} = \frac{\sum_{i=1}^n (\ln(Y_i) - \hat{\mu})^2}{2\hat{\sigma}^4} \\
&\Rightarrow \hat{\sigma}^2 = \frac{\sum_{i=1}^n (\ln(Y_i) - \hat{\mu})^2}{n}.
\end{aligned} \tag{2.38}$$

Tais soluções determinam o ponto  $(\hat{\mu}, \hat{\sigma}^2)$  que maximiza a função  $l(\mu, \sigma^2|Y)$ . A matriz de variâncias e covariâncias é obtida pela inversa da matriz de informação de Fisher:

$$I(\mu, \sigma^2|Y) = - \begin{bmatrix} U_{\mu\mu}(\mu, \sigma^2|Y) & U_{\mu\sigma}(\mu, \sigma^2|Y) \\ U_{\sigma\mu}(\mu, \sigma^2|Y) & U_{\sigma\sigma}(\mu, \sigma^2|Y) \end{bmatrix}, \tag{2.39}$$

em que  $U_{\mu\mu}(\mu, \sigma|Y)$ ,  $U_{\mu\sigma}(\mu, \sigma|Y) = U_{\sigma\mu}(\mu, \sigma|Y)$ ,  $U_{\sigma\sigma}(\mu, \sigma|Y)$ , são determinados respectivamente por:

$$\begin{aligned}
U_{\mu\mu}(\mu, \sigma^2|Y) &= \frac{\partial^2 l}{\partial \mu^2} \\
&= \frac{\partial l}{\partial (\mu)} \left[ \frac{\sum_{i=1}^n \ln(Y_i)}{\hat{\sigma}^2} - \frac{2n\hat{\mu}}{2\hat{\sigma}^2} \right] \\
&= -\frac{n}{\hat{\sigma}^2}
\end{aligned} \tag{2.40}$$

$$\begin{aligned}
U_{\mu\sigma}(\mu, \sigma^2|Y) &= U_{\sigma\mu}(\mu, \sigma^2|Y) = \frac{\partial^2 l}{\partial \mu \cdot \partial \sigma} = \frac{\partial^2 l}{\partial \sigma \cdot \partial \mu} \\
&= \frac{\partial l}{\partial (\sigma)} \left[ \frac{\sum_{i=1}^n \ln(Y_i)}{\hat{\sigma}^2} - \frac{2n\hat{\mu}}{2\hat{\sigma}^2} \right] \\
&= 0
\end{aligned} \tag{2.41}$$

$$\begin{aligned}
U_{\sigma\sigma}(\mu, \sigma^2|Y) &= \frac{\partial^2 l}{\partial (\sigma^2)^2} \\
&= \frac{\partial l}{\partial \sigma^2} \left[ -\frac{n}{2\hat{\sigma}^2} - \frac{\sum_{i=1}^n (\ln(Y_i) - \hat{\mu})^2}{2} \cdot (-\hat{\sigma}^2)^{-2} \right] \\
&= \frac{1}{2 \cdot (\hat{\sigma}^2)^3} \left[ -\sum_{i=1}^n (\ln(Y_i) - \hat{\mu})^2 \right].
\end{aligned} \tag{2.42}$$

Portanto a matriz de informação de Fisher é escrita como:

$$I(\mu, \sigma^2 | Y) = - \begin{bmatrix} -\frac{n}{\sigma^2} & 0 \\ 0 & \frac{1}{2 \cdot (\sigma^2)^3} [-\sum_{i=1}^n (\ln(Y_i) - \hat{\mu})^2] \end{bmatrix}. \quad (2.43)$$

### 2.2.2 Modelo Generalizado Misto Log-Normal

Em um contexto de medidas repetidas, define-se um modelo misto log-normal com efeitos aleatórios normais, com  $Y$  denotando as observações aleatórias e independentes, onde  $Y_i = (Y_{i1}, \dots, Y_{in_i})$  tem distribuição  $LN(\mu_{ij}, \sigma_{ij}^2)$ , com  $i = 1, \dots, N$  observações. Tem-se dois efeitos aleatórios  $b_{i\mu}$  e  $b_{i\sigma^2}$ , que são independentes e com distribuição Normal de média 0 e matrizes de covariâncias  $D_{b_\mu}$  e  $D_{b_{\sigma^2}}$ . Portanto pode-se escrever o modelo de regressão como (WALLACE; GREEN, 2002):

$$Y_{ij} | b_{i\mu} b_{i\sigma^2} \sim LN(\mu_{ij}, \sigma_{ij}^2), \quad (2.44)$$

$$i = 1, \dots, N, \quad j = 1, \dots, n_i$$

$$b_{i\mu} \sim N(0, D_{b_\mu}), \quad (2.45)$$

$$b_{i\sigma^2} \sim N(0, D_{b_{\sigma^2}}), \quad (2.46)$$

onde os parâmetros  $\mu_{ij}$  e  $\sigma_{ij}^2$  satisfazem as relações:

$$g_\mu = \eta_{ij\mu} = x_{ij\mu}^T \beta_\mu + Z_{ij\mu}^T b_{i\mu} \quad (2.47)$$

$$g_{\sigma^2} = \eta_{ij\sigma^2} = x_{ij\sigma^2}^T \beta_{\sigma^2} + Z_{ij\sigma^2}^T b_{i\sigma^2}, \quad (2.48)$$

em que:

- $x_{ij\mu} = (x_{ij\mu 1}, \dots, x_{ij\mu p_\mu})$  e  $x_{ij\sigma^2} = (x_{ij\sigma^2 1}, \dots, x_{ij\sigma^2 p_{\sigma^2}})$  são matrizes com as observações das variáveis preditoras, associadas ao  $i$ -ésimo indivíduo no  $j$ -ésimo momento dos efeitos fixos de  $\mu$  e  $\sigma^2$ ;
- $Z_{ij\mu} = (Z_{ij1\mu}, \dots, Z_{ijq_\mu\mu})^T$  e  $Z_{ij\sigma^2} = (Z_{ij1\sigma^2}, \dots, Z_{ijq_{\sigma^2}\sigma^2})^T$  são matrizes com as observações das variáveis preditoras, associadas ao  $i$ -ésimo indivíduo no  $j$ -ésimo momento dos efeitos aleatórios de  $\mu$  e  $\sigma^2$ ;
- $\beta_\mu = (\beta_0, \dots, \beta_{(p-1)_\mu})^T$  e  $\beta_{\sigma^2} = (\beta_0, \dots, \beta_{(p-1)_{\sigma^2}})^T$  são vetores dos efeitos fixos associados a cada um dos parâmetros da distribuição  $\mu$  e  $\sigma^2$ ;
- $b_{i\mu} = (b_{i1\mu}, \dots, b_{iq_\mu\mu})^T$  e  $b_{i\sigma^2} = (b_{i1\sigma^2}, \dots, b_{iq_{\sigma^2}\sigma^2})^T$  são vetores dos efeitos aleatórios associados ao  $i$ -ésimo indivíduo no  $j$ -ésimo momento de cada um dos parâmetros da distribuição  $\mu$  e  $\sigma^2$ ;

- $D_\mu$  e  $D_{\sigma^2}$  são matrizes  $q_\mu \times q_\mu$ ,  $q_{\sigma^2} \times q_{\sigma^2}$  bloco diagonal, que representam a variância de  $\gamma_{i\mu}$  e  $\gamma_{i\sigma^2}$ , respectivamente.

Em um contexto de regressão temos que a esperança e a variância de  $Y_i$  são expressas por:

$$\mathbb{E}(Y_{ij}) = e^{\mu_{ij} + \sigma_{ij}^2/2} = e^{g_\mu^{-1}(\eta_{ij\mu}) + \frac{(g_{\sigma^2}^{-1}(\eta_{ij\sigma^2}))^2}{2}} \quad (2.49)$$

e

$$Var(Y_{ij}) = (e^{\sigma_{ij}^2} - 1)e^{2\mu_{ij} + \sigma_{ij}^2} = (e^{(g_{\sigma^2}^{-1}(\eta_{ij\sigma^2}))^2} - 1)e^{2(g_\mu^{-1}(\eta_{ij\mu}) + (g_{\sigma^2}^{-1}(\eta_{ij\sigma^2}))^2)}. \quad (2.50)$$

Como se trata de um modelo misto generalizado, onde as observações não são necessariamente independentes, deve-se determinar a correlação e covariância existente entre as medidas observadas em um mesmo indivíduo e entre os indivíduos. Considerando as funções de ligação  $g_\mu$  e  $g_{\sigma^2}$ , a covariância entre  $Y_{ij}$  e  $Y_{ij'}$  é dada por:

$$\begin{aligned} Cov(Y_{ij}, Y_{kl}) &= \mathbb{E}(Y_{ij}Y_{kl}) - \mathbb{E}(Y_{ij})\mathbb{E}(Y_{kl}) \\ &= \mathbb{E}(\mu_{ij}\mu_{kl}) - \mathbb{E}(\mu_{ij})\mathbb{E}(\mu_{kl}) \\ &= \mathbb{E}(g_\mu^{-1}(\eta_{ij\mu})g_\mu^{-1}(\eta_{kl\mu})) - \mathbb{E}(g_\mu^{-1}(\eta_{ij\mu}))\mathbb{E}(g_\mu^{-1}(\eta_{kl\mu})). \end{aligned} \quad (2.51)$$

Logo a correlação entre  $Y_{ij}$  e  $Y_{ij'}$  é dada por:

$$\begin{aligned} Corr(Y_{ij}, Y_{kl}) &= \frac{Cov(Y_{ij}, Y_{kl})}{\sqrt{Var(Y_{ij})Var(Y_{kl})}} \\ &= \frac{\mathbb{E}(g_\mu^{-1}(\eta_{ij\mu})g_\mu^{-1}(\eta_{kl\mu})) - \mathbb{E}(g_\mu^{-1}(\eta_{ij\mu}))\mathbb{E}(g_\mu^{-1}(\eta_{kl\mu}))}{\sqrt{(e^{(g_{\sigma^2}^{-1}(\eta_{ij\sigma^2}))^2} - 1)e^{2(g_\mu^{-1}(\eta_{ij\mu}) + (g_{\sigma^2}^{-1}(\eta_{ij\sigma^2}))^2)} \sqrt{(e^{(g_{\sigma^2}^{-1}(\eta_{kl\sigma^2}))^2} - 1)e^{2(g_\mu^{-1}(\eta_{kl\mu}) + (g_{\sigma^2}^{-1}(\eta_{kl\sigma^2}))^2)}}} \end{aligned} \quad (2.52)$$

## 2.3 Distribuição Gama

A distribuição Gama é uma distribuição apropriada para dados contínuos, positivos e assimétricos, onde diversas distribuições como a exponencial e a qui-quadrado são casos particulares. Quando parametrizada em função de sua média  $\mu$ , sua função densidade de probabilidade denotada por  $GA(\mu, \sigma^2)$ , pode ser definida como:

$$f(y|\mu, \sigma^2) = \frac{1}{(\sigma^2\mu)^{1/\sigma^2}} \frac{y^{\frac{1}{\sigma^2}-1} e^{-y/(\sigma^2\mu)}}{\Gamma(1/\sigma^2)}, \quad (2.53)$$

para  $y > 0$ , onde  $\mu > 0$  e  $\sigma > 0$ . Tem-se  $\mathbb{E}(Y) = \mu$  e  $Var(Y) = \sigma^2\mu^2$ . Tal reparametrização é apresentada por [Johnson Samuel Kotz \(1995\)](#), obtida ao se determinar  $\sigma^2 = 1/\alpha$  e  $\mu = \alpha\beta$ . Adicionalmente,  $\mu$  é o parâmetro de escala, enquanto que  $\sigma$  é o de forma. O suporte da função é dado por  $supp(f) = x \in (0, \infty)$ . Além disso, para qualquer número real positivo  $\alpha$ , tem-se que  $\Gamma(\alpha)$  é definida como:

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx. \quad (2.54)$$

Escrevendo a função (2.53) na forma da família exponencial (2.13), tem-se:

$$\begin{aligned} f(y|\mu, \sigma^2) &= \exp \left( \log \left( \frac{1}{(\sigma^2\mu)^{1/\sigma^2}} \frac{y^{\frac{1}{\sigma^2}-1} e^{-y/(\sigma^2\mu)}}{\Gamma(1/\sigma^2)} \right) \right) \\ &= \exp \left( \log \left( \frac{1}{(\sigma^2\mu)^{1/\sigma^2}} \right) + \log \left( \frac{y^{\frac{1}{\sigma^2}-1}}{\Gamma(1/\sigma^2)} \right) - \frac{y}{(\sigma^2\mu)} - \log(\Gamma(1/\sigma^2)) \right) \\ &= \left( \frac{y^{\frac{1}{\sigma^2}-1}}{\Gamma(1/\sigma^2)} \right) \cdot \exp \left( -\frac{y}{(\sigma^2\mu)} - \log(\Gamma(1/\sigma^2)) + \log \left( \frac{1}{(\sigma^2\mu)^{1/\sigma^2}} \right) \right), \end{aligned} \quad (2.55)$$

em que  $b(y) = \left( \frac{y^{\frac{1}{\sigma^2}-1}}{\Gamma(1/\sigma^2)} \right)$ ,  $\eta(\mu, \sigma^2) = \left( \frac{1}{\mu}, \frac{1}{\sigma^2} \right)^T$ ,  $T(y) = -y$  e  $a(\mu, \sigma^2) = \log(\Gamma(1/\sigma^2)) + \log \left( \frac{1}{(\sigma^2\mu)^{1/\sigma^2}} \right)$ .

Na Figura 4 vê-se alguns dos possíveis formatos que a distribuição Gama (2.53) pode assumir, conforme a variação dos seus dois parâmetros:

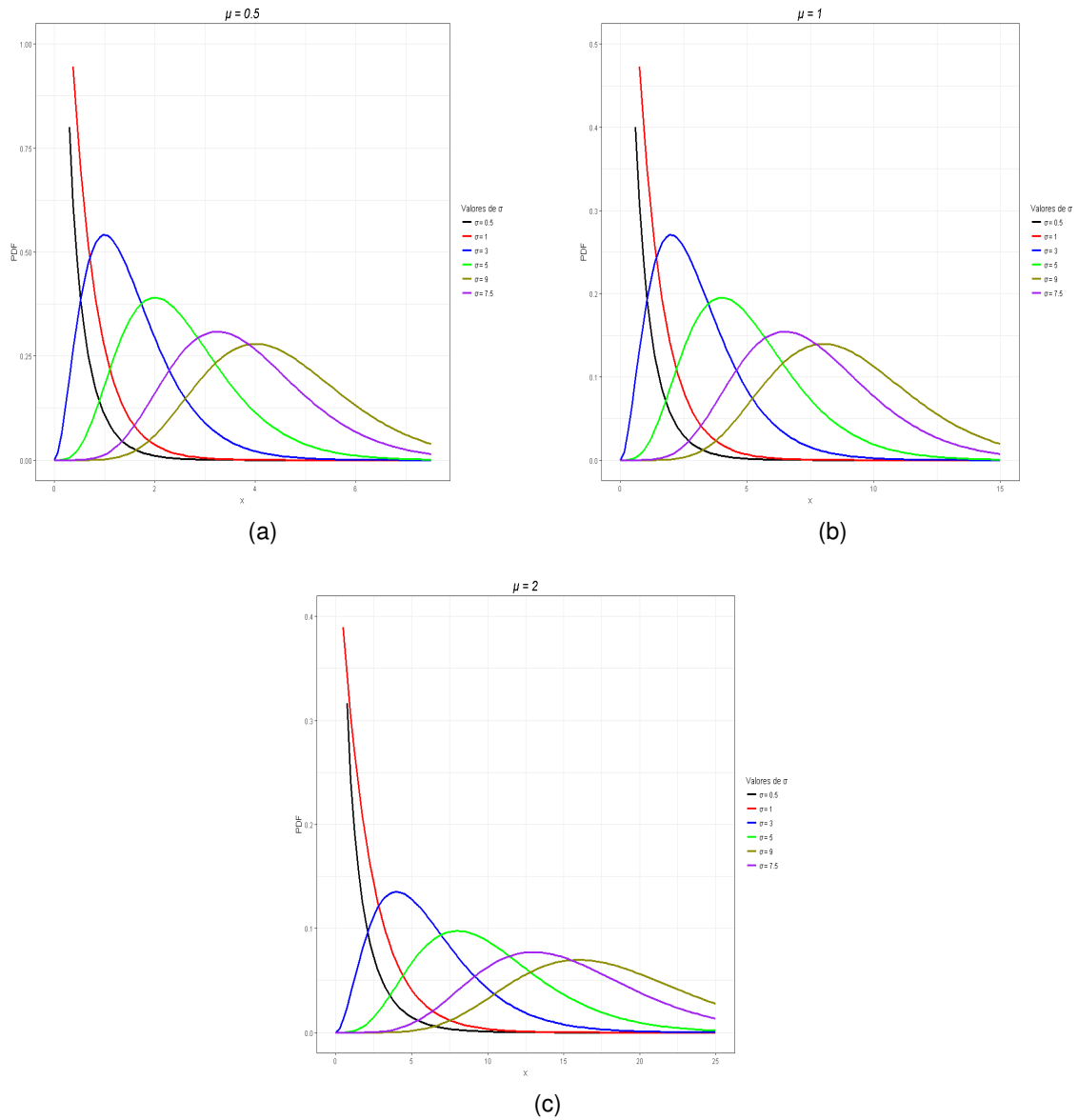


Figura 4 – (a) Distribuição Gama para  $\mu = 0$  e  $\sigma = 0, 5; 1; 3; 5; 9; 7, 5$ ; (b) Distribuição Gama para  $\mu = 1$  e  $\sigma = 0, 5; 1; 3; 5; 9; 7, 5$ ; (c) Distribuição Gama para  $\mu = 2$  e  $\sigma = 0, 5; 1; 3; 5; 9; 7, 5$ .

Na Figura (4) (c) pode-se notar uma propriedade da distribuição Gama, que é o fato de que a medida que  $\sigma$  aumenta, a distribuição fica mais simétrica em torno da média. Pode-se demonstrar que a medida que  $\sigma$  aumenta, a distribuição Gama se aproxima de uma distribuição normal de média  $\mu$  e variância  $\sigma^2\mu^2$ . Além disso, quando  $\sigma = 1$ , tem-se que a distribuição Gama é igual a exponencial.

O momento de ordem  $k$  de uma variável aleatória  $Y$  com distribuição  $GA(\mu, \sigma)$  é determinado pela igualdade  $\mu_k = \mathbb{E}[Y^k]$ , como segue:

$$\mu_k = M_Y^k(0) = \mathbb{E}[Y^k], \quad (2.56)$$

onde  $M_Y^k(t)$  é definida como segue:

$$\begin{aligned}
 M_Y^k(t) &= \frac{\partial^k}{\partial t^k} \mathbb{E}[e^{ty}] \\
 &= \frac{\partial^k}{\partial t^k} \int_{-\infty}^{\infty} e^{ty} f_{GA}(y; \mu, \sigma) dy \\
 &= \frac{\partial^k}{\partial t^k} \int_{-\infty}^{\infty} e^{ty} \frac{1}{(\sigma^2 \mu)^{1/\sigma}} \frac{y^{\frac{1}{\sigma^2}-1} e^{-y/(\sigma^2 \mu)}}{\Gamma(1/\sigma^2)} dy \\
 &= \frac{\partial^k}{\partial t^k} (1 - \mu \sigma^2 t)^{-1/\sigma^2}.
 \end{aligned} \tag{2.57}$$

O valor esperado de uma realização da variável aleatória  $Y$  com distribuição  $GA(\mu, \sigma)$  é o valor da equação (2.56), quando  $k = 1$ , ou seja, para se encontrar o valor esperado deve-se derivar a função geradora de momentos  $M_Y^k(t)$  (2.57) uma vez e então calcular o seu valor no ponto  $t = 0$ :

$$\begin{aligned}
 \mu_1 = \mathbb{E}[Y] &= \frac{\partial}{\partial t} (1 - \mu \sigma^2 t)^{-1/\sigma^2} \\
 &= \mu (1 - \mu \sigma^2 t)^{-\frac{1}{\sigma^2}-1} \\
 &= \mu (1 - \mu \sigma^2 \cdot 0)^{-\frac{1}{\sigma^2}-1} \\
 &= \mu.
 \end{aligned} \tag{2.58}$$

Com base na identidade  $Var(Y) = \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2$ , pode-se determinar a variância de  $Y$ :

$$\begin{aligned}
 \mu_2 = \mathbb{E}[Y^2] &= \frac{\partial^2}{\partial t^2} (1 - \mu \sigma^2 t)^{-1/\sigma^2} \\
 &= (\sigma^2 + 1) \mu^2 (1 - t \mu \sigma^2)^{-2-\frac{1}{\sigma^2}} \\
 &= (\sigma^2 + 1) \mu^2 (1 - 0 \cdot \mu \sigma^2)^{-2-\frac{1}{\sigma^2}} \\
 &= (\sigma^2 + 1) \mu^2.
 \end{aligned} \tag{2.59}$$

Portanto:

$$\begin{aligned}
 Var(Y) &= \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2 \\
 &= (\sigma^2 + 1) \mu^2 - \mu^2 \\
 &= \sigma^2 \mu^2.
 \end{aligned} \tag{2.60}$$



Uma importante propriedade da distribuição Gama é que a transformação  $Z_i = \text{Log}(Y_i)$  estabiliza a variância da distribuição Gama,  $\text{Var}(Y) = \sigma^2 \mu^2$ . De acordo com a expansão de Taylor de ordem 2, tem-se:

$$\begin{aligned} Z_i = \log(Y_i) &\approx \log(\mu_i) + \frac{1}{\mu_i}(Y_i - \mu_i) - \frac{1}{2\mu_i^2}(Y_i - \mu_i)^2 \\ \implies \mathbb{E}(Z_i) &\approx \log(\mu_i) - \frac{1}{2\mu_i^2}\text{Var}(Y_i) \\ &= \log(\mu_i) - \frac{1}{2}\sigma^2. \end{aligned} \quad (2.61)$$

De acordo com a expansão de Taylor de primeira ordem, tem-se:

$$\begin{aligned} \mathbb{E}(Z_i) &\approx \log(\mu_i) \\ \mathbb{E}(Z_i^2) &\approx \mathbb{E}\left(\left[\log(\mu_i) + \frac{1}{\mu_i}(Y_i - \mu_i)\right]^2\right) \\ &= (\log(\mu_i))^2 + 2\mathbb{E}\left(\log(\mu_i) \cdot \frac{1}{\mu_i}(Y_i - \mu_i)\right) + \mathbb{E}\left(\frac{1}{\mu_i^2}(Y_i - \mu_i)^2\right) \\ &= (\log(\mu_i))^2 + 0 + \frac{1}{\mu_i^2}\sigma^2\mu_i^2 = (\log(\mu_i))^2 + \sigma^2 \end{aligned} \quad (2.62)$$

$$\implies \text{Var}(Z_i) \approx (\log(\mu_i))^2 + \sigma^2 - (\log(\mu_i))^2 = \sigma^2. \quad (2.63)$$

Portanto para  $\sigma^2$  pequeno, tem-se  $\mathbb{E}(Z_i) \approx \log(\mu_i) - \sigma^2/2$  e  $\text{Var}(Z_i) \approx \sigma^2$ .

### 2.3.1 Estimação dos Parâmetros da distribuição Gama

Os métodos comumente utilizados para estimar os parâmetros de uma distribuição Gama, são o de Máxima Verossimilhança e o Método dos Momentos.

Para calcular as estimativas de Máxima Verossimilhança, primeiramente determina-se a verossimilhança da função Gama para uma variável aleatória  $Y_i$ , com  $(i = 1, 2, \dots, n)$ , que é o produto das probabilidades individuais dos  $Y_i$ 's:

$$\begin{aligned} L(\mu, \sigma^2 | Y) &= \prod_{i=1}^n [f(Y_i | \mu, \sigma)] \\ &= \prod_{i=1}^n \left[ \frac{1}{(\sigma^2 \mu)^{1/\sigma^2}} \frac{y^{\frac{1}{\sigma^2}-1} e^{-y/(\sigma^2 \mu)}}{\Gamma(1/\sigma^2)} \right]. \end{aligned} \quad (2.64)$$

A função log-verossimilhança da gama é dada pela aplicação do logaritmo natural na função de verossimilhança:

$$\begin{aligned}
 l(\mu, \sigma^2|Y) &= \ln \left( \prod_{i=1}^n \left[ \frac{1}{(\sigma^2 \mu)^{1/\sigma^2}} \frac{y_i^{\frac{1}{\sigma^2}-1} e^{-y_i/(\sigma^2 \mu)}}{\Gamma(1/\sigma^2)} \right] \right) \\
 &= \sum_{i=1}^n \left[ \ln \left( \frac{1}{(\sigma^2 \mu)^{1/\sigma^2}} \frac{y_i^{\frac{1}{\sigma^2}-1} e^{-y_i/(\sigma^2 \mu)}}{\Gamma(1/\sigma^2)} \right) \right] \\
 &= \sum_{i=1}^n \left[ \ln(1) - \frac{1}{\sigma^2} \ln(\sigma^2 \mu) + \left( \frac{1}{\sigma^2} - 1 \right) \ln(y_i) - \frac{y_i}{\mu \sigma^2} - \ln(\Gamma(1/\sigma^2)) \right] \\
 &= \sum_{i=1}^n \left[ -\frac{\ln(\sigma^2 \mu)}{\sigma^2} + \left( \frac{1}{\sigma^2} - 1 \right) \ln(y_i) - \frac{y_i}{\mu \sigma^2} - \ln(\Gamma(1/\sigma^2)) \right]. \quad (2.65)
 \end{aligned}$$

Maximizando a função  $l(\mu, \sigma^2|Y)$  encontra-se os estimadores de  $\mu$  e  $\sigma$ , para isso deriva-se tal função e a resolve, com relação a  $\mu$ :

$$\begin{aligned}
 U_\mu(\mu, \sigma^2|Y) &= \frac{\partial l(\mu, \sigma^2|Y)}{\partial \mu} \\
 &= \sum_{i=1}^n \left[ \frac{y_i - \mu}{\mu^2 \sigma^2} \right] = 0 \\
 \Rightarrow n\hat{\mu} &= \sum_{i=1}^n y_i \\
 \Rightarrow \hat{\mu} &= \frac{\sum_{i=1}^n y_i}{n}. \quad (2.66)
 \end{aligned}$$

Com relação a  $\sigma$  temos:

$$\begin{aligned}
 U_{\sigma^2}(\mu, \sigma^2|Y) &= \frac{\partial l(\mu, \sigma^2|Y)}{\partial \sigma^2} \\
 &= \sum_{i=1}^n \left[ \frac{\mu \ln(\mu \sigma^2) + \mu \psi^{(0)}\left(\frac{1}{\sigma^2}\right) - \mu - \mu \ln(y_i) + y_i}{\mu(\sigma^2)^2} \right] = 0. \quad (2.67)
 \end{aligned}$$

Tal equação (2.67) não possui solução analítica. Sendo que  $\psi^{(0)}(1/\sigma^2)$  é a  $n$ -ésima derivada da função digama, definida como a derivada do logaritmo da função Gama:

$$\begin{aligned}
 \psi(1/\sigma^2) &= \frac{\partial \ln(\Gamma(1/\sigma^2))}{\partial \sigma^2} \\
 &= \frac{\Gamma'(1/\sigma^2)}{\Gamma(1/\sigma^2)}. \quad (2.68)
 \end{aligned}$$

Tais soluções determinam o ponto  $(\hat{\mu}, \hat{\sigma})$  que maximiza a função  $l(\mu, \sigma|Y)$ . A matriz de variâncias e covariâncias é obtida pela inversa da matriz de informação de Fisher:

$$I(\mu, \sigma^2|Y) = - \begin{bmatrix} U_{\mu\mu}(\mu, \sigma^2|Y) & U_{\mu\sigma^2}(\mu, \sigma^2|Y) \\ U_{\sigma^2\mu}(\mu, \sigma^2|Y) & U_{\sigma^2\sigma^2}(\mu, \sigma^2|Y) \end{bmatrix}. \quad (2.69)$$

Em que  $U_{\mu\mu}(\mu, \sigma^2|Y)$ ,  $U_{\mu\sigma^2}(\mu, \sigma^2|Y) = U_{\sigma^2\mu}(\mu, \sigma^2|Y)$ ,  $U_{\sigma^2\sigma^2}(\mu, \sigma^2|Y)$ , são determinados respectivamente por:

$$\begin{aligned} U_{\mu\mu}(\mu, \sigma^2|Y) &= \frac{\partial^2 l}{\partial \mu^2} \\ &= \frac{\partial l}{\partial(\mu)} \sum_{i=1}^n \left[ \frac{y_i - \mu}{\mu^2 \sigma^2} \right] \\ &= \sum_{i=1}^n \left[ \frac{\mu - 2y_i}{\mu^3 \sigma^2} \right]. \end{aligned} \quad (2.70)$$

$$\begin{aligned} U_{\mu\sigma^2}(\mu, \sigma^2|Y) &= U_{\sigma^2\mu}(\mu, \sigma^2|Y) = \frac{\partial^2 l}{\partial \mu \cdot \partial \sigma^2} = \frac{\partial^2 l}{\partial \sigma^2 \cdot \partial \mu} \\ &= \frac{\partial l}{\partial(\sigma^2)} \sum_{i=1}^n \left[ \frac{y_i - \mu}{\mu^2 \sigma^2} \right] \\ &= \sum_{i=1}^n \left[ -\frac{2(y_i - \mu)}{\mu^2 \sigma^3} \right] \end{aligned} \quad (2.71)$$

$$\begin{aligned} U_{\sigma^2\sigma^2}(\mu, \sigma^2|Y) &= \frac{\partial^2 l}{\partial(\sigma^2)^2} \\ &= \frac{\partial l}{\partial(\sigma^2)} \sum_{i=1}^n \left[ \frac{\mu \ln(\mu \sigma^2) + \mu \psi^{(0)}\left(\frac{1}{\sigma^2}\right) - \mu - \mu \ln(y_i) + y_i}{\mu(\sigma^2)^2} \right] \\ &= -\frac{2\mu \sigma^2 \psi^{(0)}\left(\frac{1}{\sigma^2}\right) + \mu \psi^{(1)}\left(\frac{1}{\sigma^2}\right) + \sigma^2 (2\mu \ln(\mu \sigma^2) - 3\mu - 2\mu \ln(y_i) + 2y_i)}{\mu(\sigma^2)^4}. \end{aligned} \quad (2.72)$$

Portanto, a matriz de informação de Fisher é escrita como:

$$I(\mu, \sigma^2|Y) = - \begin{bmatrix} \sum_{i=1}^n \left[ \frac{\mu - 2y_i}{\mu^3 \sigma^2} \right] & \sum_{i=1}^n \left[ -\frac{2(y_i - \mu)}{\mu^2 \sigma^3} \right] \\ \sum_{i=1}^n \left[ -\frac{2(y_i - \mu)}{\mu^2 \sigma^3} \right] & -\frac{2\mu \sigma^2 \psi^{(0)}\left(\frac{1}{\sigma^2}\right) + \mu \psi^{(1)}\left(\frac{1}{\sigma^2}\right) + \sigma^2 (2\mu \ln(\mu \sigma^2) - 3\mu - 2\mu \ln(y_i) + 2y_i)}{\mu(\sigma^2)^4} \end{bmatrix}. \quad (2.73)$$

### 2.3.2 Modelo Generalizado Misto Gama

Em um contexto de medidas repetidas, define-se um modelo misto Gama com efeitos aleatórios normais, com  $Y$  denotando as observações aleatórias e independentes, onde  $Y_i = (Y_{i1}, \dots, Y_{in_i})$  tem distribuição  $GA(\mu_{ij}, \sigma_{ij}^2)$ , com  $i = 1, \dots, N$  observações e  $j = 1, \dots, n_i$  momentos. Sendo dois efeitos aleatórios,  $b_{i\mu}$  e  $b_{i\sigma^2}$ , que como visto são independentes e com distribuição Normal com média 0 e matrizes de covariâncias  $D_\mu$  e  $D_{\sigma^2}$ . Portanto, pode-se escrever o modelo de regressão como (WALLACE; GREEN, 2002):

$$Y_{ij}|b_{i\mu}b_{i\sigma^2} \sim GA(\mu_{ij}, \sigma_{ij}^2), \quad (2.74)$$

$$i = 1, \dots, N, \quad j = 1, \dots, n_i$$

$$b_{i\mu} \sim N(0, D_\mu), \quad (2.75)$$

$$b_{i\sigma^2} \sim N(0, D_{\sigma^2}), \quad (2.76)$$

em que os parâmetros  $\mu_{ij}$  e  $\sigma_{ij}$  satisfazem as relações:

$$g_\mu = \eta_{ij\mu} = x_{ij\mu}^T \beta_\mu + Z_{ij\mu}^T b_{i\mu} \quad (2.77)$$

$$g_\sigma = \eta_{ij\sigma} = x_{ij\sigma}^T \beta_{\sigma^2} + Z_{ij\sigma}^T b_{i\sigma^2}. \quad (2.78)$$

em que:

- $x_{ij\mu} = (x_{ij\mu 1}, \dots, x_{ij\mu p_\mu})$  e  $x_{ij\sigma^2} = (x_{ij\sigma^2 1}, \dots, x_{ij\sigma^2 p_{\sigma^2}})$  são matrizes com as observações das variáveis preditoras, associadas ao  $i$ -ésimo indivíduo no  $j$ -ésimo momento dos efeitos fixos de  $\mu$  e  $\sigma^2$ ;
- $Z_{ij\mu} = (Z_{ij1\mu}, \dots, Z_{ijq_\mu\mu})^T$  e  $Z_{ij\sigma^2} = (Z_{ij1\sigma^2}, \dots, Z_{ijq_{\sigma^2}\sigma^2})^T$  são matrizes com as observações das variáveis preditoras, associadas ao  $i$ -ésimo indivíduo no  $j$ -ésimo momento dos efeitos aleatórios de  $\mu$  e  $\sigma^2$ ;
- $\beta_\mu = (\beta_0, \dots, \beta_{(p-1)_\mu})^T$  e  $\beta_{\sigma^2} = (\beta_0, \dots, \beta_{(p-1)_{\sigma^2}})^T$  são vetores dos efeitos fixos associados a cada um dos parâmetros da distribuição  $\mu$  e  $\sigma^2$ ;
- $b_{i\mu} = (b_{i1\mu}, \dots, b_{iq_\mu\mu})^T$  e  $b_{i\sigma^2} = (b_{i1\sigma^2}, \dots, b_{iq_{\sigma^2}\sigma^2})^T$  são vetores dos efeitos aleatórios associados ao  $i$ -ésimo indivíduo no  $j$ -ésimo momento de cada um dos parâmetros da distribuição  $\mu$  e  $\sigma^2$ ;
- $D_\mu$  e  $D_{\sigma^2}$  são matrizes  $q_\mu \times q_\mu$ ,  $q_{\sigma^2} \times q_{\sigma^2}$  bloco diagonal, que representam a variância de  $\gamma_{i\mu}$  e  $\gamma_{i\sigma^2}$ , respectivamente.

Em um contexto de regressão temos que a esperança e a variância de  $Y_i$  são expressas por:

$$\mathbb{E}(Y_{ij}) = \mu_{ij} = g_{\mu}^{-1}(\eta_{ij\mu}) \quad (2.79)$$

e

$$Var(Y_{ij}) = \sigma_{ij}^2 \mu_{ij}^2 = (g_{\sigma^2}^{-1}(\eta_{ij\sigma^2}))(g_{\mu}^{-1}(\eta_{ij\mu}))^2. \quad (2.80)$$

Como se trata de um modelo misto, em que as observações não são necessariamente independentes, devemos determinar a correlação e covariância existente entre as medidas observadas em um mesmo indivíduo e entre os indivíduos. Considerando as funções de ligação  $g_{\mu}$  e  $g_{\sigma^2}$ , a covariância entre  $Y_{ij}$  e  $Y_{ij'}$  é dada por:

$$\begin{aligned} Cov(Y_{ij}, Y_{kl}) &= \mathbb{E}(Y_{ij}Y_{kl}) - \mathbb{E}(Y_{ij})\mathbb{E}(Y_{kl}) \\ &= \mathbb{E}(\mu_{ij}\mu_{kl}) - \mathbb{E}(\mu_{ij})\mathbb{E}(\mu_{kl}) \\ &= \mathbb{E}(g_{\mu}^{-1}(\eta_{ij\mu})g_{\mu}^{-1}(\eta_{kl\mu})) - \mathbb{E}(g_{\mu}^{-1}(\eta_{ij\mu}))\mathbb{E}(g_{\mu}^{-1}(\eta_{kl\mu})). \end{aligned} \quad (2.81)$$

Logo a correlação entre  $Y_{ij}$  e  $Y_{kl}$  é dada por:

$$\begin{aligned} Corr(Y_{ij}, Y_{kl}) &= \frac{Cov(Y_{ij}, Y_{kl})}{\sqrt{Var(Y_{ij})Var(Y_{kl})}} \\ &= \frac{\mathbb{E}(g_{\mu}^{-1}(\eta_{ij\mu})g_{\mu}^{-1}(\eta_{kl\mu})) - \mathbb{E}(g_{\mu}^{-1}(\eta_{ij\mu}))\mathbb{E}(g_{\mu}^{-1}(\eta_{kl\mu}))}{\sqrt{(g_{\sigma^2}^{-1}(\eta_{ij\sigma^2}))(g_{\mu}^{-1}(\eta_{ij\mu}))^2} \sqrt{(g_{\sigma^2}^{-1}(\eta_{kl\sigma^2}))(g_{\mu}^{-1}(\eta_{kl\mu}))^2}} \end{aligned} \quad (2.82)$$

---

## Capítulo 3

---

---

# Resultados 1<sup>o</sup> Ensaio

---

As distribuições Log-Normal  $LN(\mu, \sigma^2)$  e Gama  $GA(\mu, \sigma^2)$  foram utilizadas para analisar o banco de dados obtido por (FRANCO, 2016) cujo objetivo é avaliar a resistência de 12 variedades de laranja doce.

O conjunto de dados contém 2880 observações do diâmetro médio das lesões (12 Genótipos  $\times$  8 Dias  $\times$  60 Sujeitos  $\times$  6 Repetições), considerando todas as covariáveis e o efeito do tempo. Para se obter as estimativas dos parâmetros foram utilizado o *software* R com o pacote `gamlss`, em que este permite que a parte sistemática do modelo seja expandida permitindo modelar não apenas o parâmetro de média, mas também outros parâmetros da distribuição. A função de verossimilhança penalizada  $l_p$  (2.19) é utilizada para ajustar os modelos, pois nem sempre a função de verossimilhança possui forma fechada, gerando integrais que não podem ser resolvidas analiticamente.

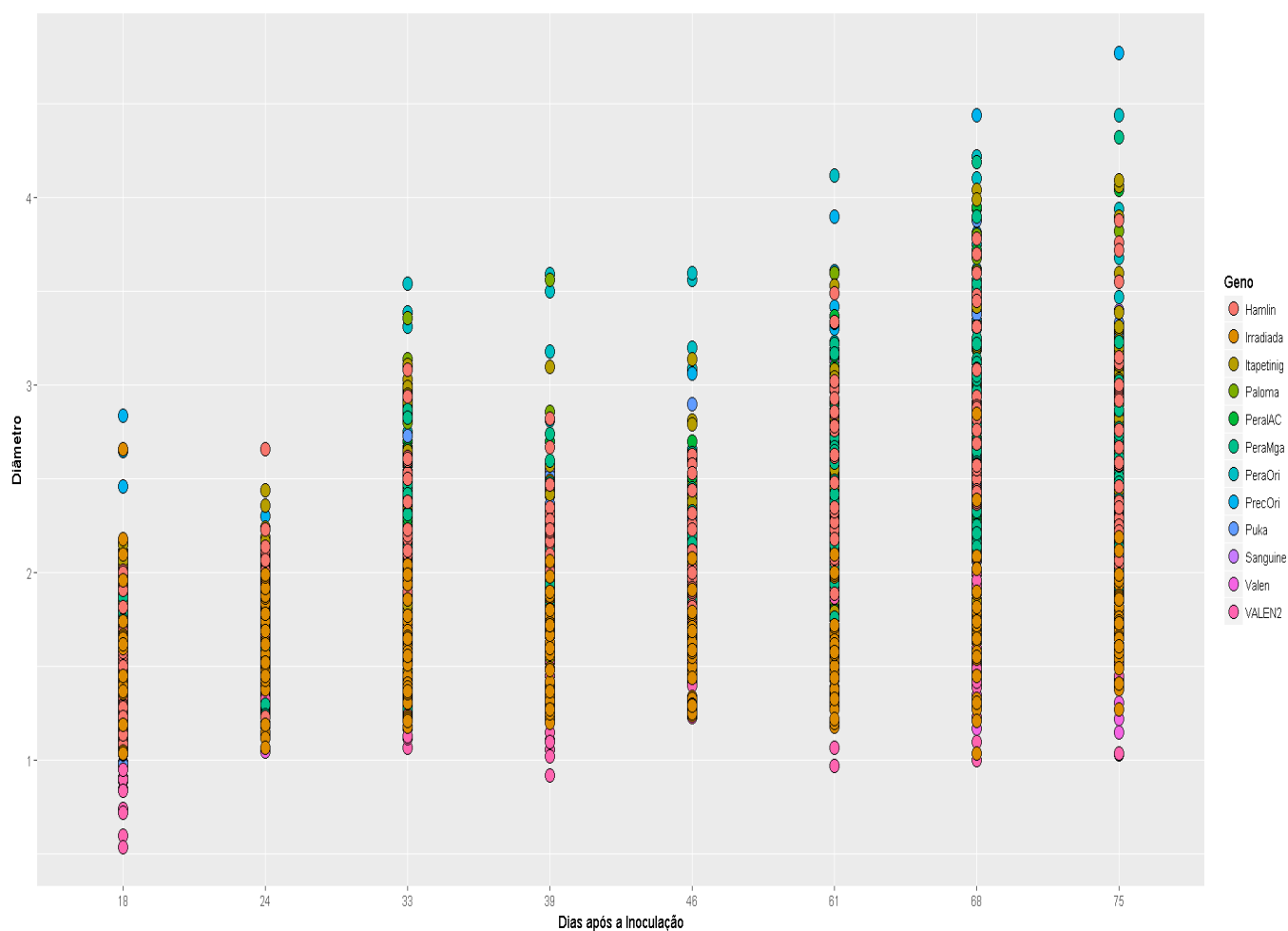


Figura 5 – Gráfico de dispersão dos dados

Na Figura (5) é apresentado o gráfico de dispersão dos dados com cada cor representando uma variedade diferente, onde pode ser notado como o diâmetro da lesão varia em relação a cada variedade e no decorrer do tempo.

### 3.1 Análise Descritiva

Realizando uma breve análise descritiva, nota-se na Figura (6) que aparentemente as variedades PrecOri, PeraOri, PeraMga possuem diâmetros de lesões maiores que as demais variedades, enquanto que as variedades VALEN, Valen e Irradiada possuem os menores. Além disso se nota que conforme o tempo passa há um aumento no diâmetro, onde tal crescimento em certo ponto se estabiliza e até mesmo diminui em algumas variedades.

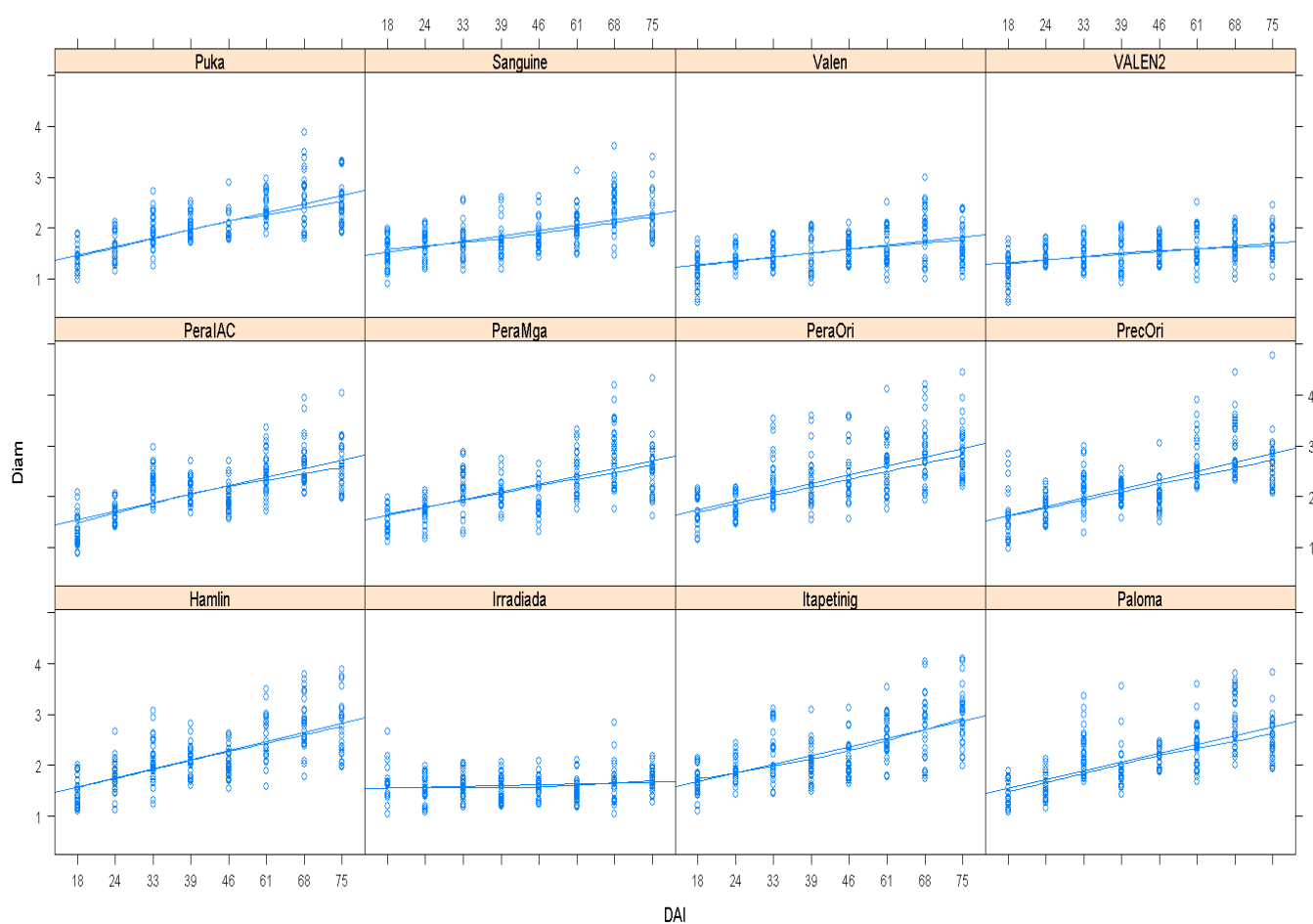


Figura 6 – Gráfico de dispersão por variedade

Na Figura (7) nota-se diferenças no comportamento do diâmetro das lesões em cada variedade, em que pode ser notado tanto assimetria a esquerda em algumas variedades quanto a direita. Tais diferenças comportamentais em relação a variável resposta fortalece a justificativa de se utilizar modelos mistos, pois estes são capazes de captar melhor tal variabilidade entre indivíduos.



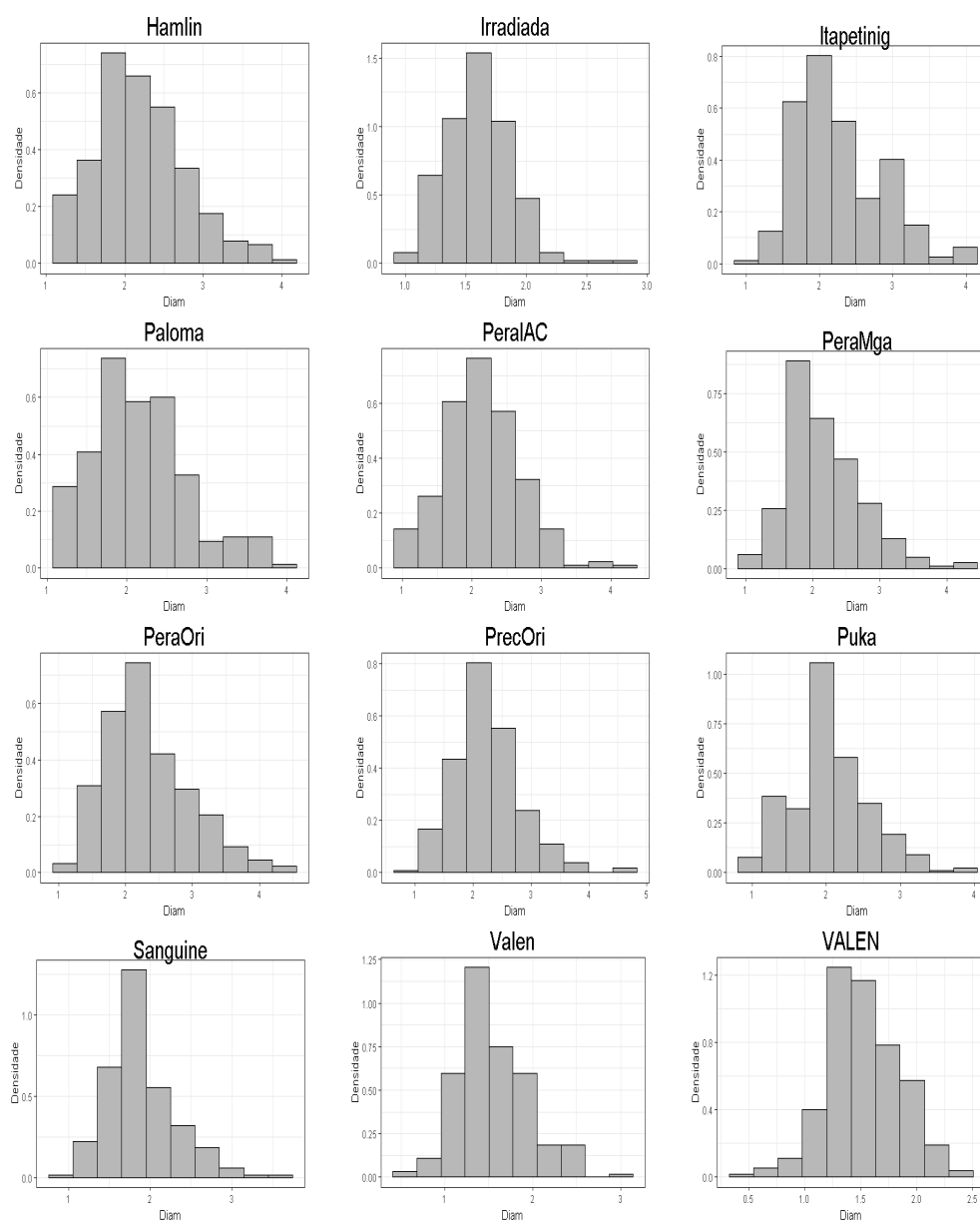


Figura 7 – Histograma da variável resposta diâmetro das lesões para as 12 variedades de laranja doce.

Na Tabela (4) tem-se um resumo da variável resposta diâmetro das lesões em relação a cada variedade:

Tabela 4 – Medidas Resumo para cada variedade.

	Variedades	n	Média	Desvio Padrão	Min.	Max.	Coefficiente de variação
1	PeraOri	240	2,34	0,62	1,16	4,44	0,27
2	Valen	240	1,55	0,40	0,54	3,00	0,26
3	PrecOri	240	2,24	0,59	0,99	4,77	0,26
4	Sanguine	240	1,90	0,43	0,91	3,61	0,22
5	Paloma	240	2,15	0,59	1,07	3,82	0,27
6	PeralAC	240	2,13	0,54	0,89	4,04	0,26
7	Puka	240	2,06	0,53	0,97	3,88	0,26
8	VALEN	240	1,51	0,33	0,54	2,50	0,22
9	Itapetinig	240	2,28	0,59	1,10	4,09	0,26
10	PeraMga	240	2,17	0,56	1,11	4,32	0,26
11	Hamlin	240	2,20	0,58	1,09	3,88	0,26
12	Irradiada	240	1,61	0,28	1,04	2,85	0,17

A partir da Tabela (4) nota-se que as médias dos variedades variam relativamente bastante, indo de 1,51cm para o variedade VALEN a 2,34cm para a variedade PeraOri, com desvios-padrões iguais a 0,62cm e 0,33cm respectivamente. Além disso percebe-se uma grande diferença entre os valores máximo e mínimo para cada variedade. Vê-se também que os valores dos coeficientes de variação oscilam de 0,17 a 0,27 representando o quanto, em média, os desvios atingem o valor da média amostral.

A Figura (8) apresenta o box-plot para o comportamento da variável resposta tamanho do diâmetro por tipo variedade:

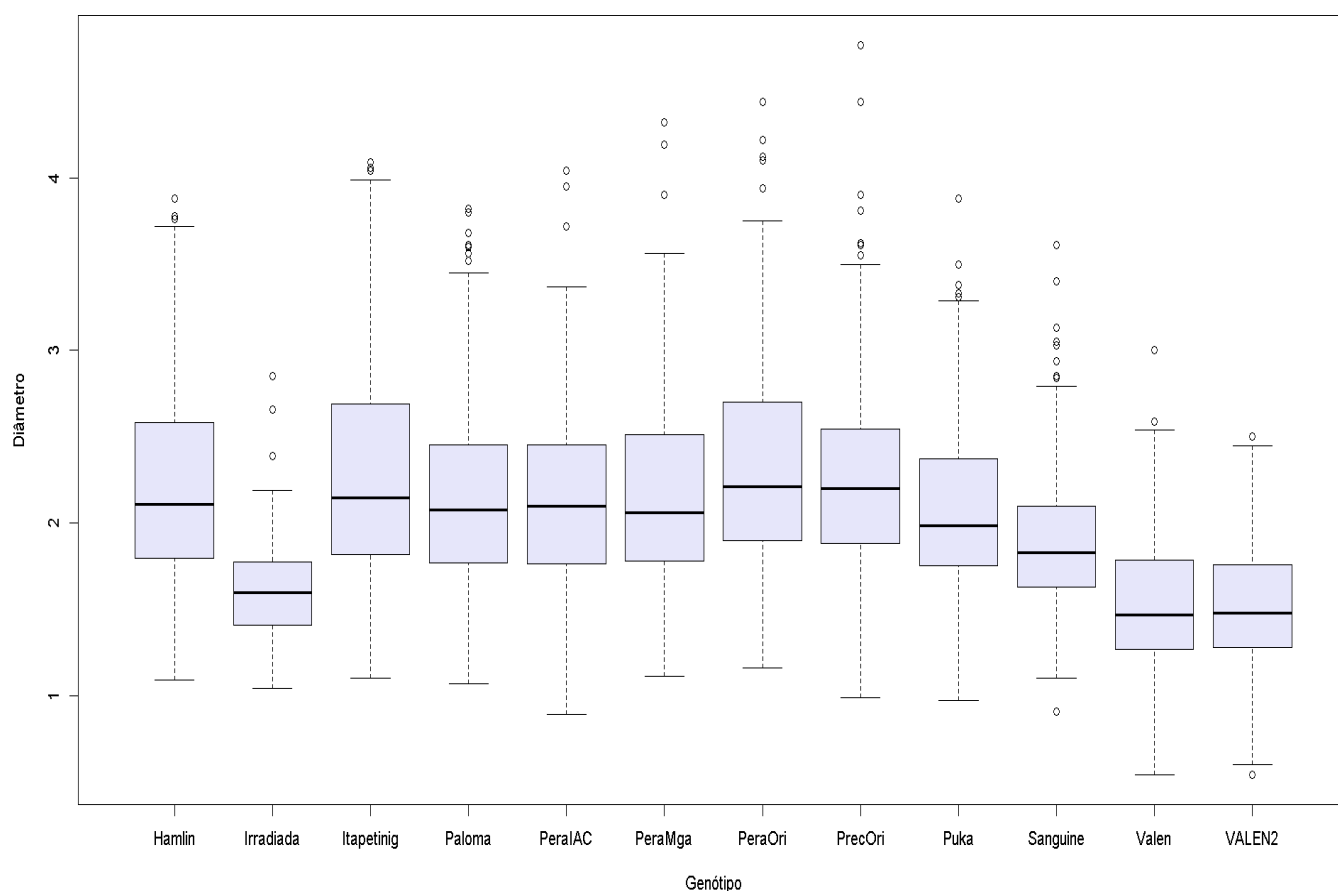


Figura 8 – *Box-plot* dos dados por variedade num total de 2880 observações.

Pela Figura (8) aparentemente a variedade Irradiada e as duas variedades de Valência possuem níveis médios do tamanho do diâmetro menor em relação as outras variedades, enquanto que a pera Ori. possui o maior nível médio. As outras variedades possuem níveis médios aparentemente próximos. Todas as variedades possuem medidas candidatas a outliers, tendo a Prec. Ori a maior medida e VALEN a menor. Não se nota grandes evidências de assimetrias a não ser na Itapetinig.

A Figura (9) apresenta o box-plot do comportamento do diâmetro por dias após a inoculação:

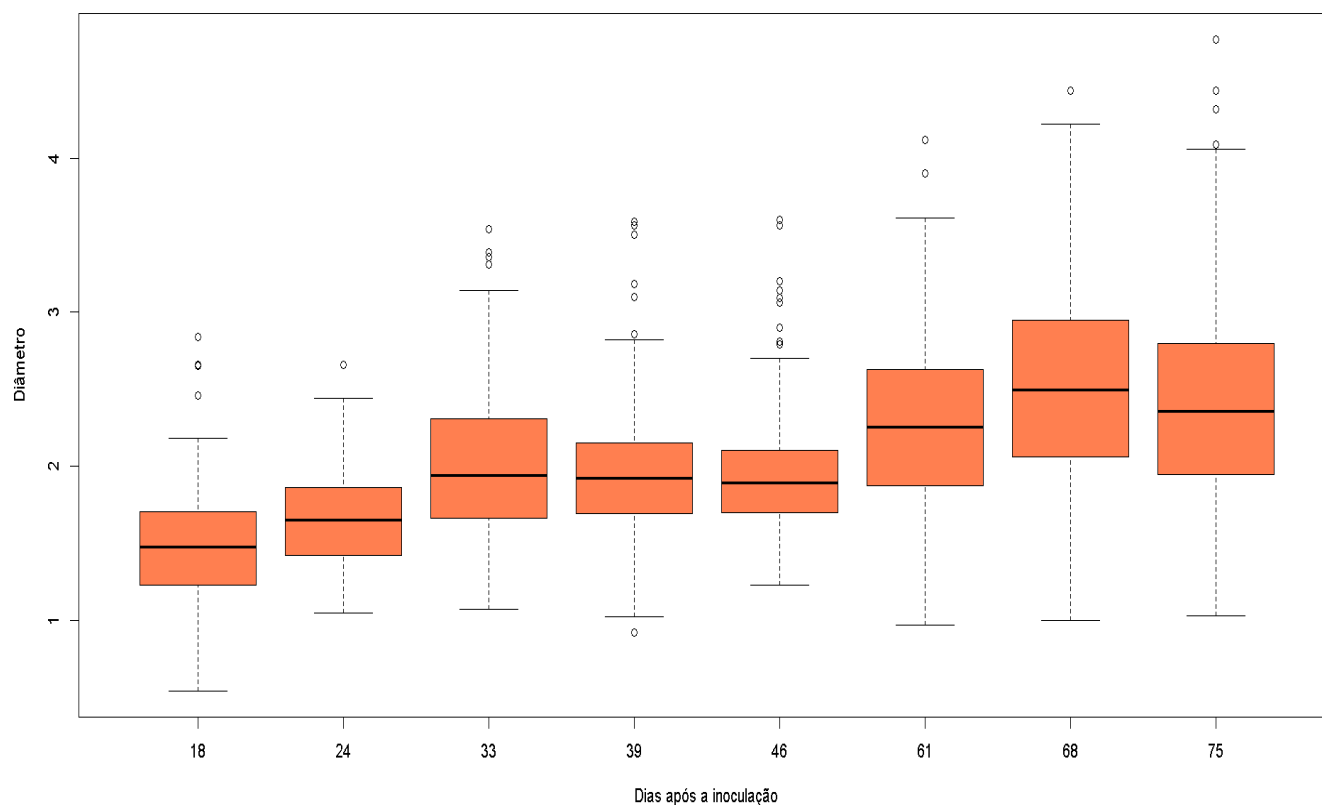


Figura 9 – *Box-plot* dos dados por dias após a inoculação para cada variedade.

Na Figura (9) nota-se um comportamento aparentemente crescente do diâmetro conforme passa-se o tempo, se estabilizando durante os dias 39 e 46, voltando a crescer 61º dia, e diminuindo no 75º. Em todos os dias vê-se medidas candidatas a *outliers*.

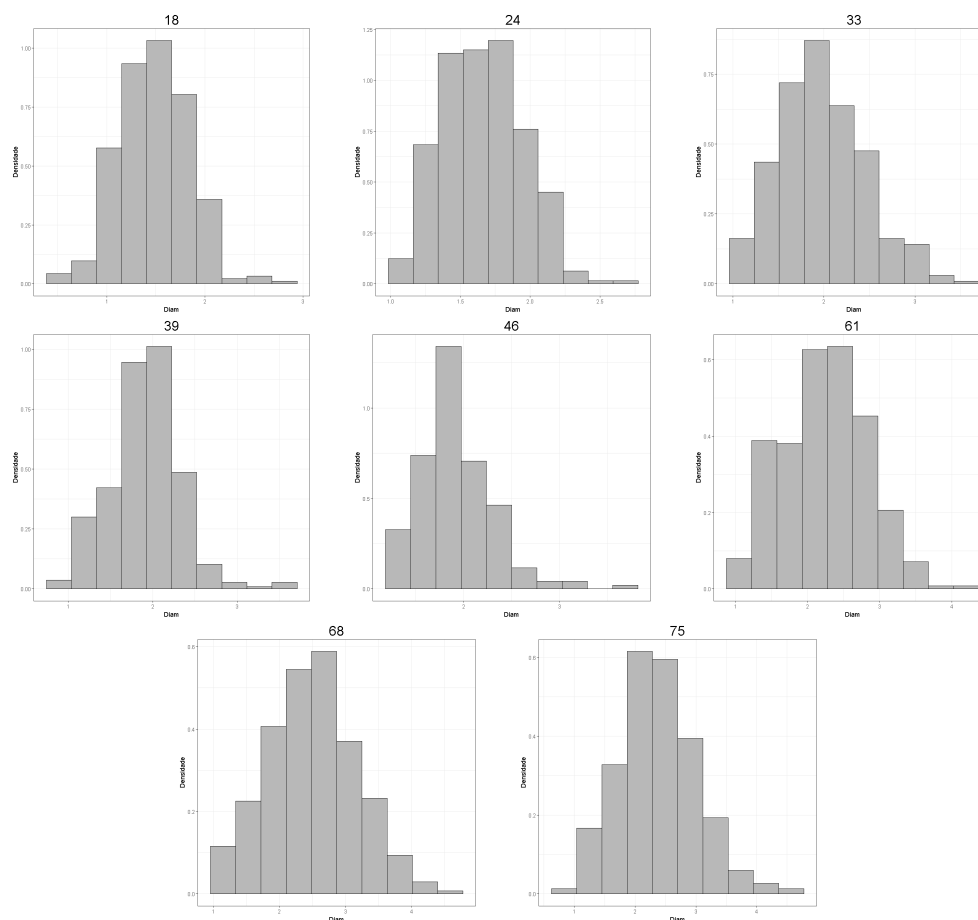


Figura 10 – Histograma da variável resposta diâmetro das lesões para os 8 dias após a inoculação.

Na Figura (10) é apresentado os histogramas para os dias após a inoculação. Podem ser notadas diferenças no comportamento do diâmetro das lesões em alguns dias, havendo assimetria a direita em alguns dias, o que pode indicar que em certos períodos o diâmetro da lesão cresceu menos que em outros.

A Figura (11) apresenta o do comportamento do diâmetro por sujeito:

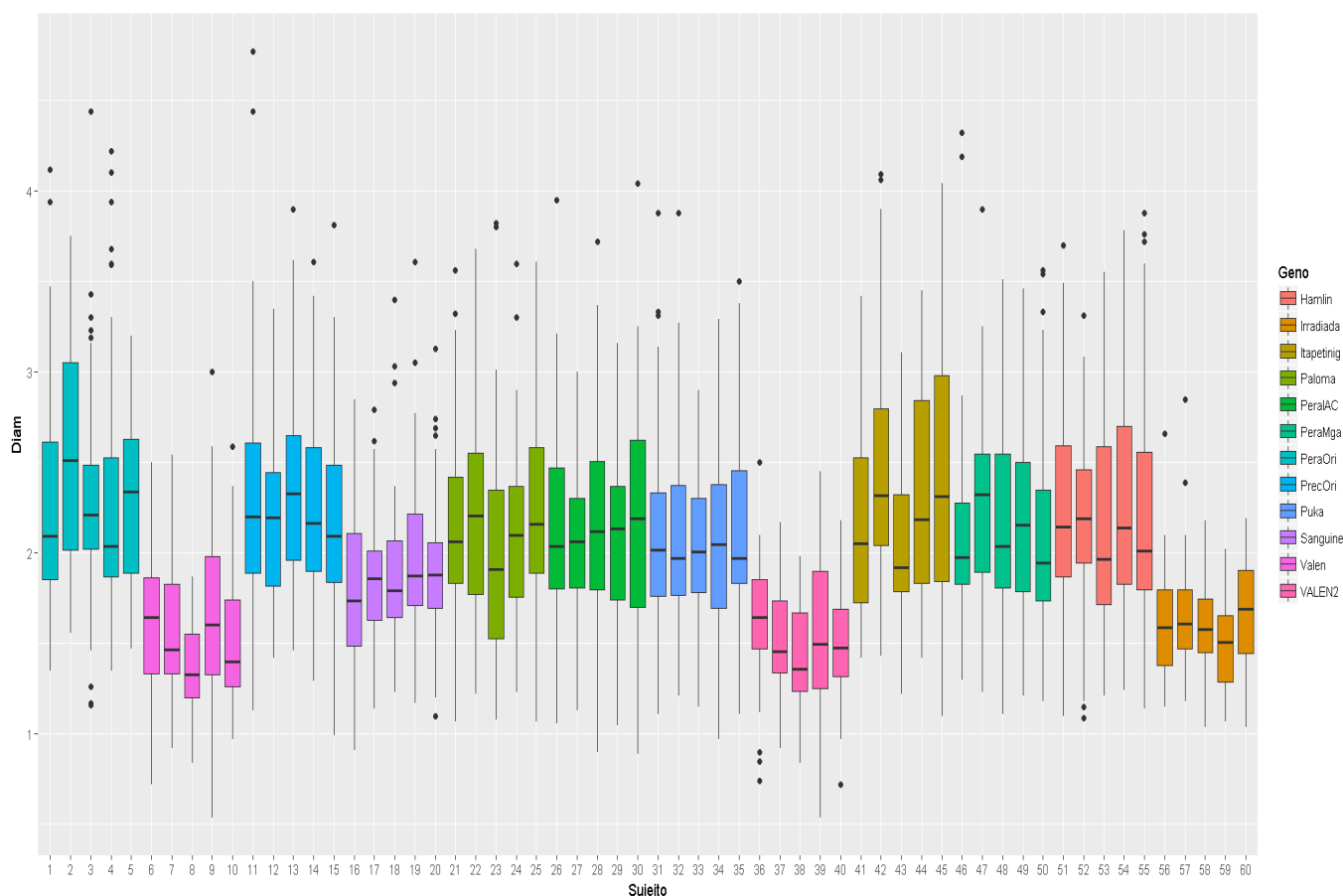


Figura 11 – Box-plot dos dados por Folha.

Perceba um comportamento aparentemente diferente do diâmetro para todos os sujeitos, possuindo variâncias e medianas diferentes de acordo com as variedades e os sujeitos pertencentes a elas. Em todos os sujeitos vê-se medidas candidatas a *outliers*. Como esperado pela Figura (8) as plantas das duas variedades de Valência e Irradiada aparentemente possuem os menores diâmetros, enquanto que a PrecOri e Itapetining possuem os maiores.

O Gráfico de perfis, Figura (12), é uma ferramenta descritiva muito importante na análise de dados longitudinais com medidas repetidas. Segundo [Singer e Andrade \(1986\)](#) eles são essencialmente gráficos de dispersão, em que os pontos associados a uma mesma unidade amostral são unidos por segmentos de reta, onde em geral os perfis médios são sobrepostos a eles. A importância de tal gráfico se deve ao fato de auxiliar:

- A identificar possíveis correlações intra-indivíduos;
- identificar *outliers*;
- na construção do tipo de modelo e de sua estrutura de covariância.

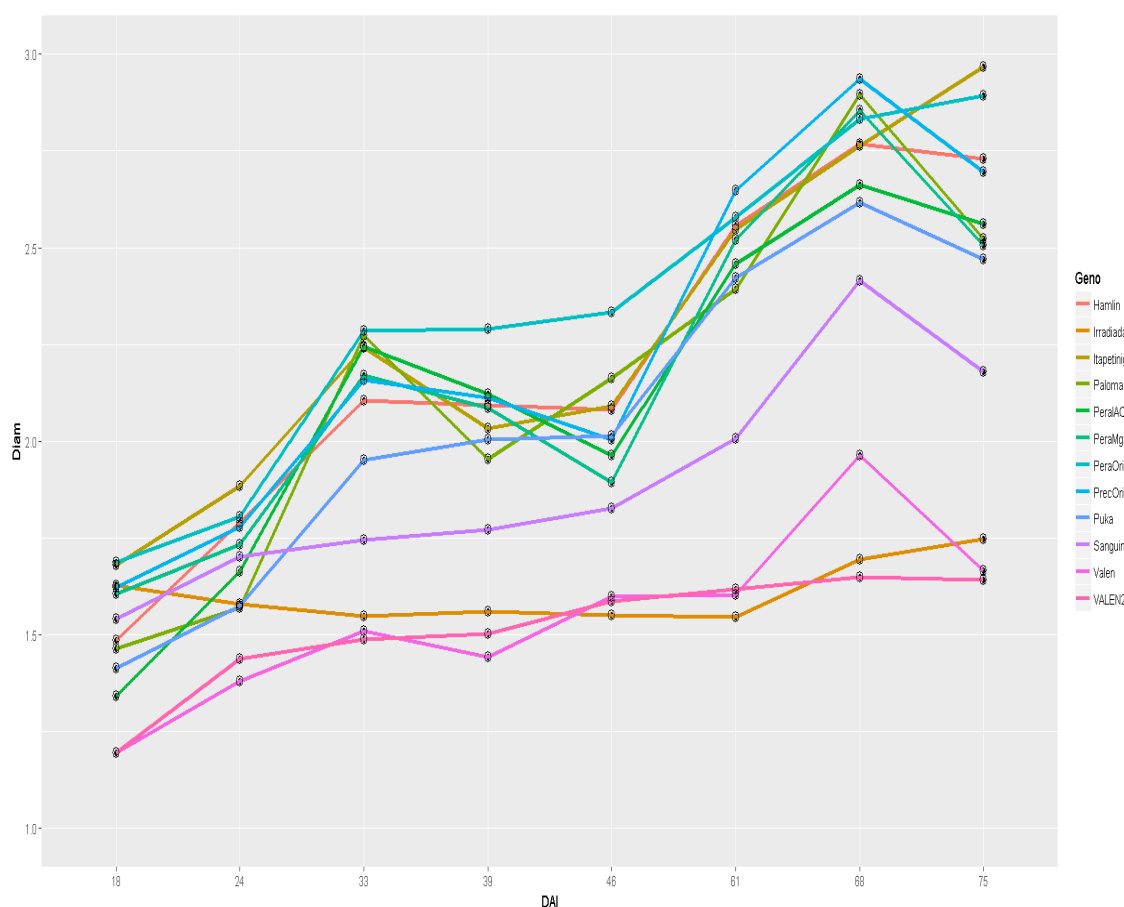


Figura 12 – Perfil Médio em cada tempo para cada variedade

Na Figura (12) nota-se que os primeiros dias são os que possuem menor diâmetro, enquanto que o dia 68 possui aparentemente os maiores, em média. Vê-se que, aparentemente, todas as variedades apresentam um crescimento no diâmetro conforme se passa os dias após a inoculação, com exceção de algumas variedades que acabam tendo um decaimento ao meio do processo, principalmente entre os dias 33 a 46, e ao final entre 68 e 75, tais decaimentos não são explicados pelo pesquisador, mas podem ser fatores que influenciem a adequabilidade de um modelo posterior. Nota-se também que as variedades VALEN, Valen, e Irradiada são as que possuem menor diâmetro em média. Enquanto que as variedades PeraOri, PrecOri e Itapetinig possuem os maiores em média, onde a Itapetinig e PeraOri, tiveram crescimento no diâmetro mesmo após o 68º dia.

## 3.2 Ajuste

Para se ajustar os dados a um modelo misto log-normal e Gama foi utilizado o pacote GAMLSS do *software* R, onde a estimação dos parâmetros fixos são feitos pela maximização da função de máxima verossimilhança penalizada  $l_p$  (2.19).

Tal função pode não possuir uma solução analítica, ou caso exista sua obtenção é

de grande dificuldade, para contornar tal dificuldade foram implementados três algoritmos no pacote GAMLSS para a obtenção dos estimadores, Rigby e Stasinopoulos (RS), Cole e Green (CG) e um algoritmo denominado *mixed* (STASINOPOULOS; RIGBY; AKANTZILI-OTOU, 2008). No presente trabalho apenas o algoritmo RS foi utilizado.

Para o ajuste dos modelos lineares generalizados mistos com base nas distribuições log-normal e gama foram utilizados o *software* R (R Core Team, 2017) e o pacote Gamlss (RIGBY; STASINOPOULOS, 2005), buscou-se analisar o comportamento do cancro cítrico em 12 variedade de laranja doce, tendo em vista analisar a relação da variável resposta "Diâmetro da Lesão", aos possíveis fatores de influência, variedade e dias após a inoculação, com estes pertencentes a parte fixa do modelo e as folhas pertencentes a parte aleatória.

Os modelos foram ajustados na variável resposta  $y_{ijk}$ ,  $i = 1, 2, \dots, 12$  (variedades),  $j = 1, 2, \dots, 8$  (dias após a inoculação),  $k = 1, 2, \dots, 60$  (folhas). A seleção do modelo foi baseada no Critério de Informação de Akaike (AIC), na Deviance Global (GD) e no Critério bayesiano de Schwarz (BIC). Os modelos ajustados estão descritos nas Tabelas (5) e (6).

Tabela 5 – Modelos propostos para a distribuição Log-normal

	ID	Modelo
1	$LN_0$	$\eta = \beta_0$
2	$LN_1$	$\eta_i = \beta_0 + \alpha_i Variedade$
3	$LN_2$	$\eta_{ij} = \beta_0 + \alpha_i Variedade + \theta_j DAI$
4	$LN_3$	$\eta_{ijk} = \beta_0 + \alpha_i Variedade + \theta_j DAI + u_{0k}$
5	$LN_4$	$\eta_{ijk} = \beta_0 + \alpha_i Variedade + \theta_j DAI + u_{1k} Variedade$
6	$LN_5$	$\eta_{ijk} = \beta_0 + \alpha_i Variedade + \theta_j DAI + u_{0k} + u_{1k} Variedade$

Tabela 6 – Modelos propostos para a distribuição Gama.

	ID	Modelo
1	$GA_0$	$\eta = \beta_0$
2	$GA_1$	$\eta_{ijk} = \beta_0 + \alpha_i Variedade$
3	$GA_2$	$\eta_{ijk} = \beta_0 + \alpha_i Variedade + \theta_j DAI$
4	$GA_3$	$\eta_{ijk} = \beta_0 + \alpha_i Variedade + \theta_j DAI + u_{0k}$
5	$GA_4$	$\eta_{ijk} = \beta_0 + \alpha_i Variedade + \theta_j DAI + u_{1k} Variedade$
6	$GA_5$	$\eta_{ijk} = \beta_0 + \alpha_i Variedade + \theta_j DAI + u_{0k} + u_{1k} Variedade$

Em tais modelos temos que  $\alpha_i$  representa os efeitos devido a  $i$ -ésima variedade,  $\theta_j$  os efeitos devido ao  $j$ -ésimo dia após a inoculação,  $u_k$  é o efeito da parte aleatória para a  $k$ -ésima folha.

Para ambas as Tabelas (5) e (6) tem-se que o primeiro modelo é dado apenas pelo intercepto da parte fixa, no segundo é acrescentado o efeito das variedades a parte fixa,



tendo como *baseline* a variedade Hamlin. No terceiro modelo é acrescentado o efeitos do dias após a inoculação a parte fixa. No quarto modelo é acrescentado o efeito das folhas por meio de um intercepto aleatório para cada uma delas. No quinto modelo é retirado o intercepto aleatório das folhas e adicionado um coeficiente angular aleatório, para cada folha em cada variedade. Finalmente no sexto modelo é acrescentado novamente o intercepto aleatório para cada folha.

Os valores de todas as medidas de decisão de cada modelo são apresentados nas Tabelas (7) e (8):

Tabela 7 – Medidas do AIC, Deviance Global e BIC para os modelos Log-normal.

	Modelo	AIC	GD	BIC
1	$LN_0$	4816,892	4812,892	4828,823
2	$LN_1$	3971,412	3945,412	4048,964
3	$LN_2$	2441,357	2401,357	2560,668
4	$LN_3$	2402,035	2293,953	2724,421
5	$LN_4$	2388,757	2294,917	2668,659
6	$LN_5$	2388,757	2294,916	2668,661

Através de tais medidas de decisão e do algoritmo *step GAIC* (STASINOPOULOS; RIGBY et al., 2007) implementado no pacote *Gamlss*, foi escolhido o modelo  $LN_3$  como o melhor modelo para a distribuição Log-Normal, pois entre os que explicam melhor os dados ele é o mais parcimonioso.

Tabela 8 – Medidas do AIC, Deviance Global e BIC para os modelos Gama.

	Modelo	AIC	GD	BIC
1	$GA_0$	4835,042	4831,042	4846,973
2	$GA_1$	3967,097	3941,097	4044,65
3	$GA_2$	2456,278	2416,278	2575,589
4	$GA_3$	2417,004	2312,309	2729,286
5	$GA_4$	2404,507	2313,313	2676,519
6	$GA_5$	2404,505	2313,279	2676,613

Através de tais medidas de decisão e do algoritmo *step GAIC*, foi escolhido o modelo  $GA_3$  como o melhor modelo para a distribuição Gama, pois entre os que explicam melhor os dados ele é o mais parcimonioso.

Para os dois modelos escolhidos foram testadas diversas estruturas de covariância, porém nenhuma melhorou os mesmos de forma significativa, isso pode ser um indicativo de que o intercepto aleatório extraiu grande parte da correlação existente. Portanto para os dois modelos selecionados a forma da matriz de covariância é a Não-estruturada (2.25).

Os modelos escolhidos para cada distribuição ( $LN_3$ ,  $GA_3$ ) possuem a mesma função de ligação e formulação podendo ser escritos como segue:

$$\log(\mu) = \eta_{ijk\mu} = x_{ij\mu}^T \beta_\mu + Z_{k\mu}^T b_{k\mu} \quad (3.1)$$

$$\log(\sigma) = \eta_{ijk\sigma} = x_{ij\sigma}^T \beta_{\sigma^2} + Z_{k\sigma^2}^T b_{k\sigma^2}. \quad (3.2)$$

$$\begin{aligned} \log(\mu) = & \beta_0 + \alpha_1 \text{Irradiada} + \alpha_2 \text{Itapetinig} + \alpha_3 \text{Paloma} + \alpha_4 \text{PeralAC} + \\ & \alpha_5 \text{PeraMga} + \alpha_6 \text{PeraOri} + \alpha_7 \text{PrecOri} + \alpha_8 \text{Puka} + \alpha_9 \text{Sanguine} + \\ & \alpha_{10} \text{Valen} + \alpha_{11} \text{VALEN2} + \theta_1 \text{DAI 24} + \theta_2 \text{DAI 33} + \dots + \\ & \theta_8 \text{DAI 75} + u_0 \text{Folha 1} + \dots + u_0 \text{Folha 60}, \end{aligned} \quad (3.3)$$

e

$$\begin{aligned} \log(\sigma) = & \beta_0 + \alpha_1 \text{Irradiada} + \alpha_2 \text{Itapetinig} + \alpha_3 \text{Paloma} + \alpha_4 \text{PeralAC} + \\ & \alpha_5 \text{PeraMga} + \alpha_6 \text{PeraOri} + \alpha_7 \text{PrecOri} + \alpha_8 \text{Puka} + \alpha_9 \text{Sanguine} + \\ & \alpha_{10} \text{Valen} + \alpha_{11} \text{VALEN2} + \theta_1 \text{DAI 24} + \theta_2 \text{DAI 33} + \dots + \\ & \theta_8 \text{DAI 75} + u_0 \text{Folha 1} + \dots + u_0 \text{Folha 60}. \end{aligned} \quad (3.4)$$

A validade de tais modelos podem ser confirmadas verificando os pressupostos dos mesmos, os quais foram avaliados através da análise de resíduos. Tal análise foi baseada nos resíduos quantílicos aleatorizados e nos gráficos de diagnósticos apresentados nas Figuras (13) e (14) para os modelos Log-Normal e Gama, respectivamente.

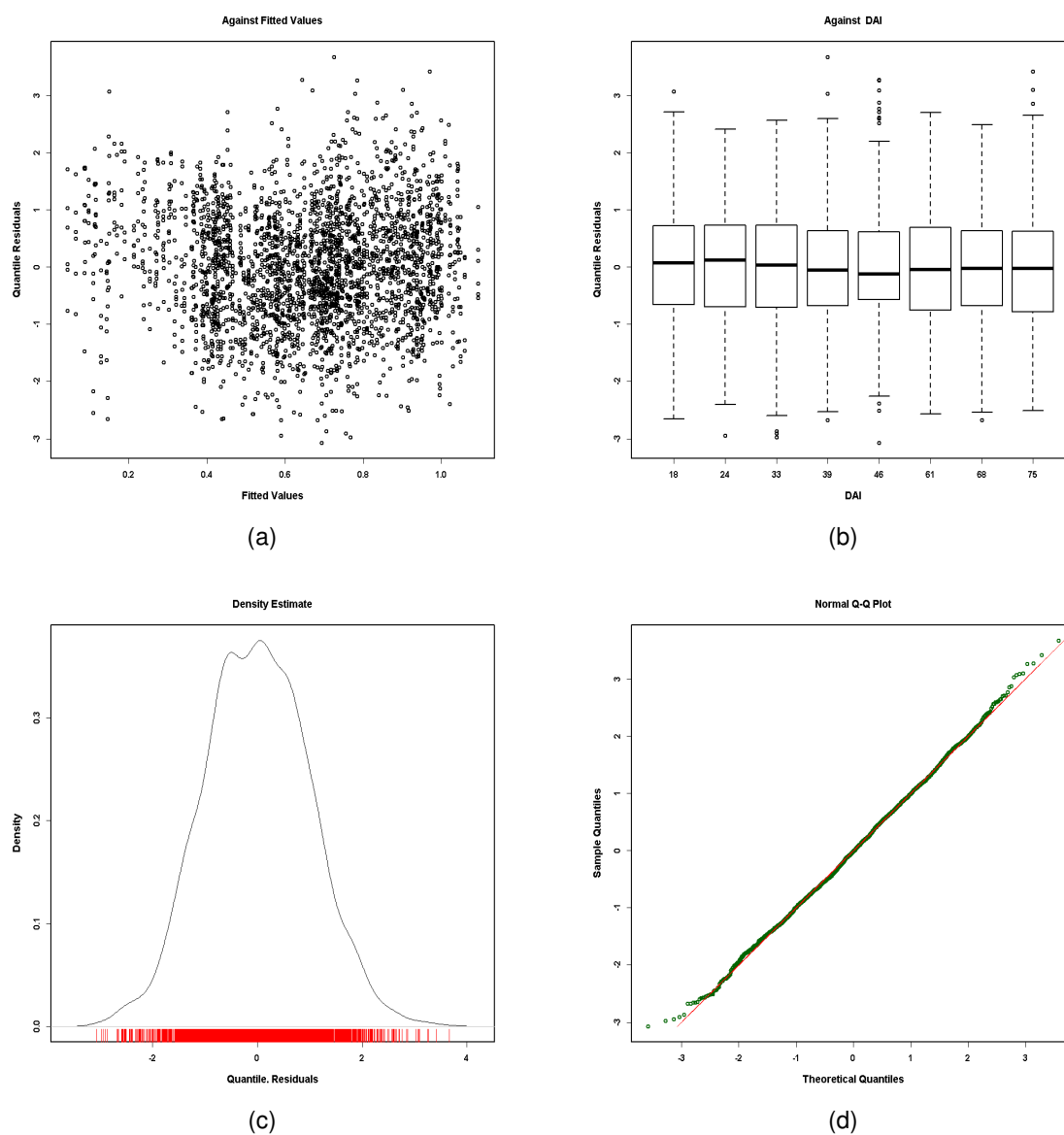


Figura 13 – (a) Valores Ajustados *vs* Quantis Residuais; (b) Dias Após a Inoculação *vs* Quantis Residuais; (c) Quantis Residuais *vs* Densidade Estimada; (d) Quantis Teóricos *vs* Quantis Amostrais.

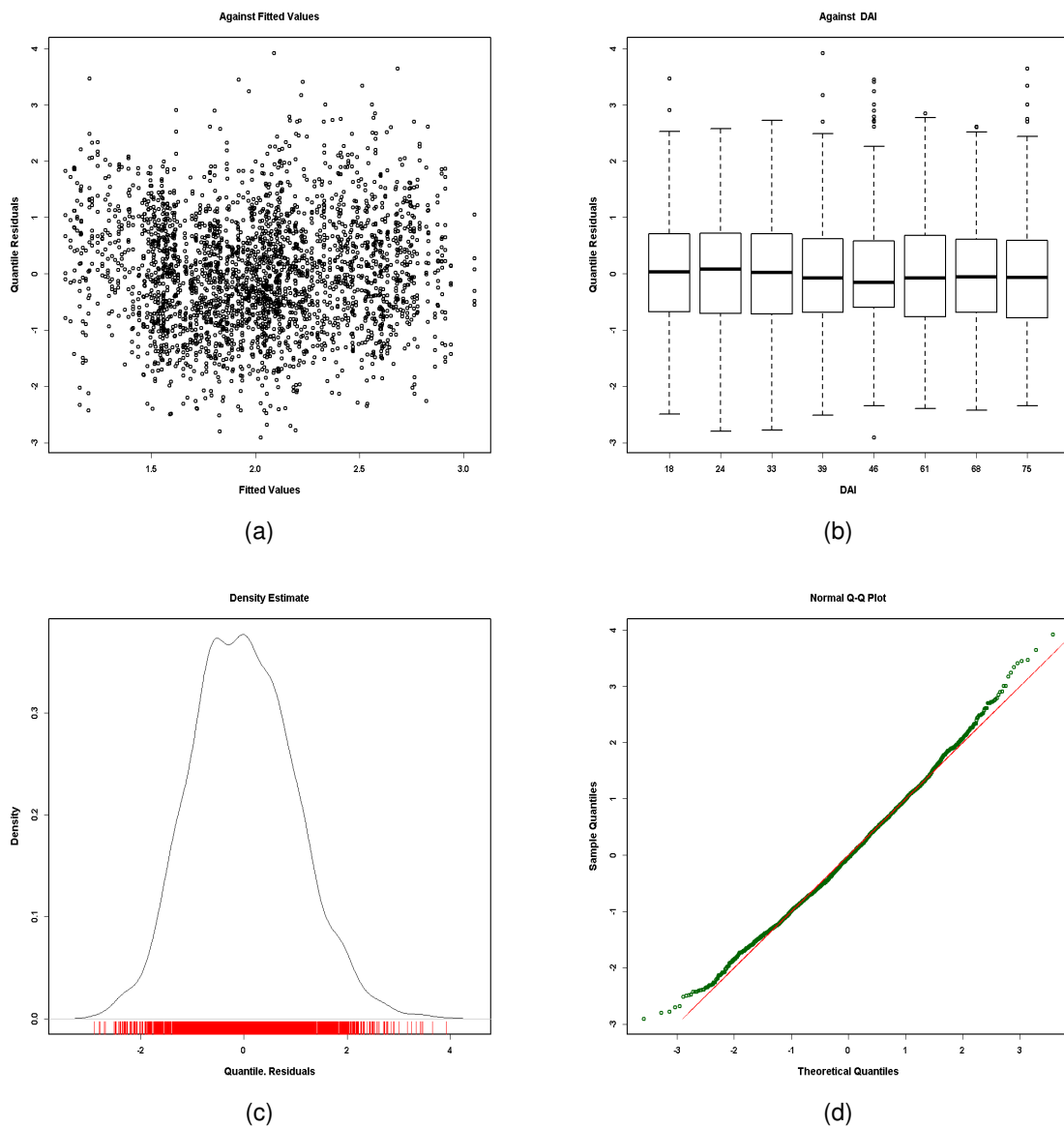


Figura 14 – (a) Valores Ajustados *vs* Quantis Residuais; (b) Dias Após a Inoculação *vs* Quantis Residuais; (c) Quantis Residuais *vs* Densidade Estimada; (d) Quantis Teóricos *vs* Quantis Amostrais.

Nas Figuras (13), (14) (a) e (b) dos valores ajustados *vs* quantis residuais e do índice *vs* quantis residuais, mostram que os resíduos estão distribuídos de forma aleatória em torno do zero, o que não revela nenhuma violação quanto a homoscedasticidade. Já os dois últimos quantis residuais *vs* densidade estimada e quantis teóricos *vs* quantis amostrais, indicam que os quantis residuais seguem uma distribuição Normal Padrão, que é requisito para um ajuste adequado.

Nas Tabelas (9) e (10) é apresentado as medidas descritivas dos resíduos quantílicos aleatorizadas para os modelos log-normal e gama, respectivamente.

Tabela 9 – Medidas descritivas dos resíduos para o modelo Log-Normal.

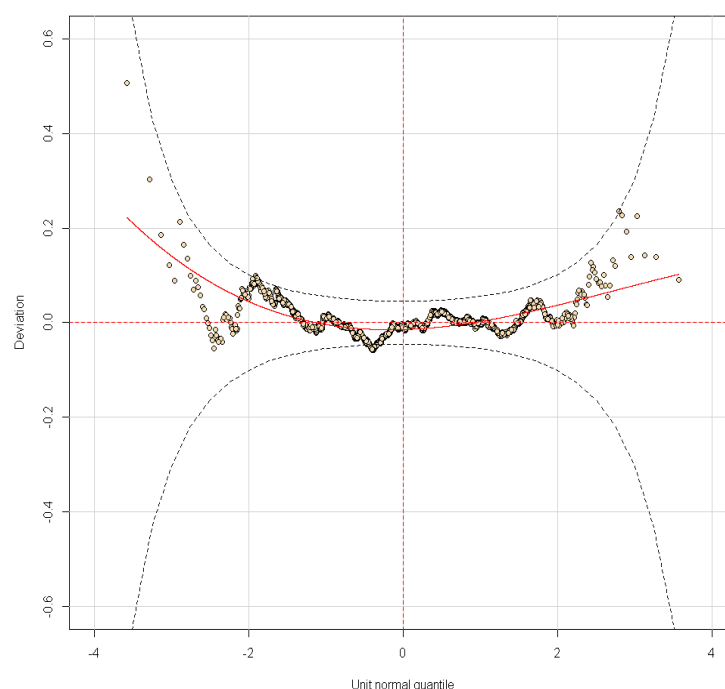
Variável	Parâmetro $Pr(> t )$
Média	0,000797188
Variância	1,000347
Coef. Assimetria	0,08083705
Coef. Curtose	2,966511

Tabela 10 – Medidas descritivas dos resíduos para o modelo Gama.

Variável	Parâmetro $Pr(> t )$
Média	-0,001039119
Variância	1,000024
Coef. Assimetria	0,2416004
Coef. Curtose	3,049964

Conforme já esperado após a análise gráfica, nota-se que as médias são próximas de 0, a variância de 1, assim como a assimetria e a curtose próximas de 0 e 3, respectivamente, conforme esperado para uma variável aleatória com distribuição Normal Padrão, o que indica um ajuste adequado.

As Figuras (15) e (16) apresentam o *Worm plot* dos resíduos quantílicos aleatorizados para os modelos log-normal e gama, respectivamente. Em tais gráficos espera-se ver os pontos o mais próximo possível da linha horizontal em torno do zero e que 95% deles estejam contidos dentro dos intervalos de confiança elípticos.

Figura 15 – *Worm plot* dos resíduos quantílicos aleatorizados para o modelo Log-Normal.

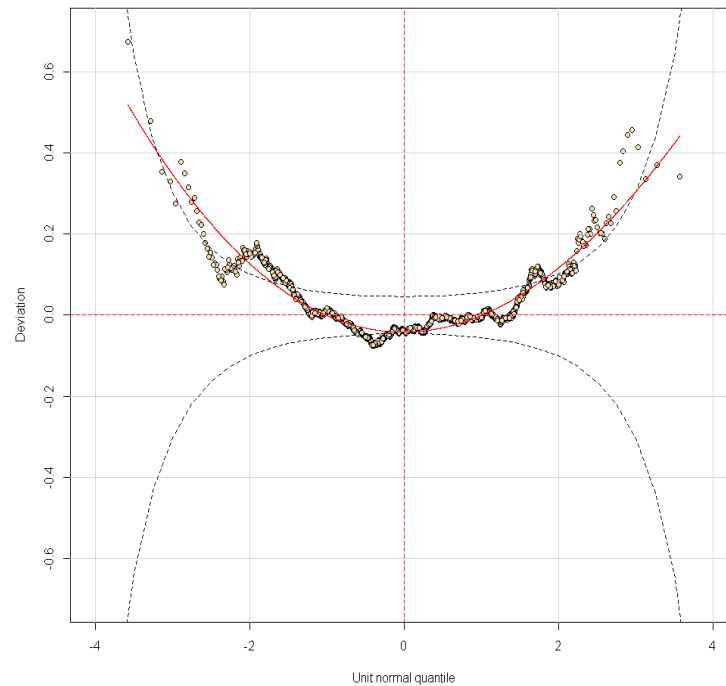


Figura 16 – *Worm plot* dos resíduos quantílicos aleatorizados para o modelo Gama.

Considerando tais figuras (15) e (16) nota-se que aparentemente, os resíduos do modelo log-normal estão melhores que o do modelo gama.

As Figuras (17) e (18) apresentam o gráfico *detrended transformed Owen's* (OWEN, 1995), o qual também é uma ferramenta de diagnóstico para checar se os resíduos quantílicos vem de uma distribuição normal, que é verdadeiro caso a linha horizontal esteja entre os intervalos de confiança, consequentemente a distribuição assumida para a variável resposta é razoável.

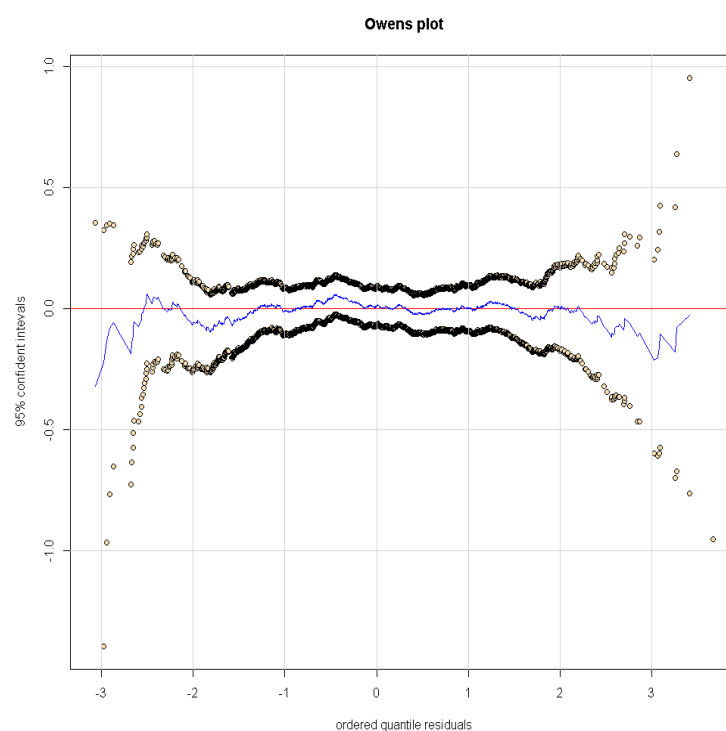


Figura 17 – Gráfico *detrended transformed Owen's* dos resíduos quantílicos aleatorizados para o modelo Log-Normal.

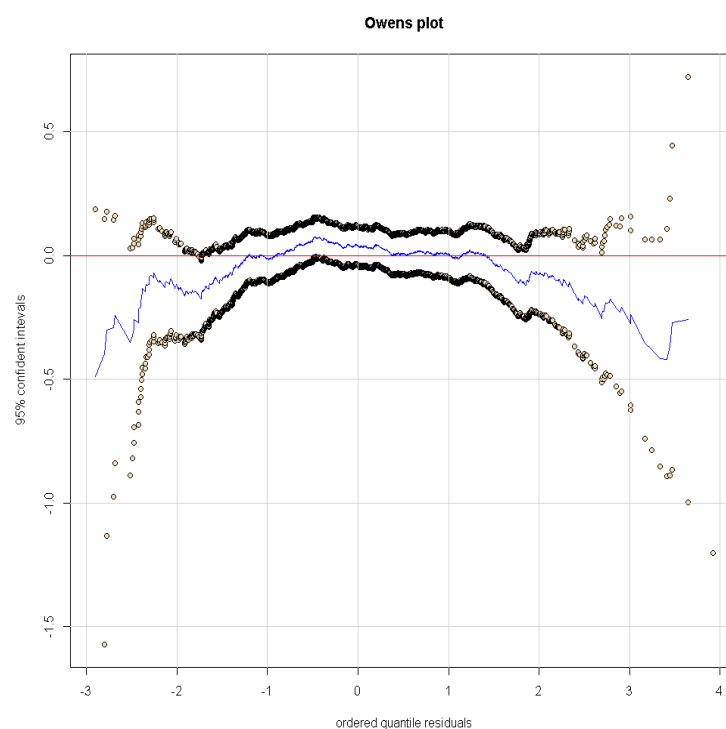


Figura 18 – Gráfico *detrended transformed Owen's* dos resíduos quantílicos aleatorizados para o modelo Gama.

Em tais figuras (17) e (18) nota-se novamente que, aparentemente, o modelo Log-

Normal se apresenta melhor que o modelo gama, pois a sua linha de tendência está mais próxima da linha horizontal em torno do 0.

No ajuste de tal modelo foi utilizado a função logarítmica como função de ligação do modelo log-normal, assim como para o modelo gama. Nas Tabelas (11) e (12), são apresentadas as estimativas dos parâmetros fixos do modelo log-normal, nas Tabelas (13) e (14) são apresentadas as estimativas dos parâmetros fixos do modelo gama, para os parâmetros  $\mu$  e  $\sigma$  respectivamente, com seus intervalos de confiança e erros padrões.

Tabela 11 – Estimativas dos parâmetros do modelo Log-Normal para  $\mu$ .

Variável	Parâmetro	Estimativa	Erro Padrão	valor t	Pr(> t )
Intercepto	$\beta_0$	0,45219	0,01658	27,278	$< 2e - 16$
Irradiada	$\alpha_1$	-0,30186	0,01769	-17,061	$< 2e - 16$
Itapetinig	$\alpha_2$	0,02746	0,01567	1,752	0,079875
Paloma	$\alpha_3$	-0,02626	0,01564	-1,679	0,093186
PeralAC	$\alpha_4$	-0,03066	0,01491	-2,057	0,039787
PeraMga	$\alpha_5$	-0,02102	0,01557	-1,350	0,177073
PeraOri	$\alpha_6$	0,06182	0,01633	3,787	0,000156
PrecOri	$\alpha_7$	0,01207	0,01519	0,794	0,427121
Puka	$\alpha_8$	-0,06376	0,01512	-4,216	$2,56e - 05$
Sanguine	$\alpha_9$	-0,14319	0,01593	-8,987	$< 2e - 16$
Valen	$\alpha_{10}$	-0,34847	0,01779	-19,586	$< 2e - 16$
VALEN2	$\alpha_{11}$	-0,35253	0,01783	-19,775	$< 2e - 16$
DAI 24	$\theta_1$	0,12558	0,01502	8,361	$< 2e - 16$
DAI 33	$\theta_2$	0,30327	0,01608	18,855	$< 2e - 16$
DAI 39	$\theta_3$	0,27364	0,01506	18,173	$< 2e - 16$
DAI 46	$\theta_4$	0,27809	0,01469	18,926	$< 2e - 16$
DAI 61	$\theta_5$	0,43113	0,01600	26,953	$< 2e - 16$
DAI 68	$\theta_6$	0,53634	0,01681	31,900	$< 2e - 16$
DAI 75	$\theta_7$	0,48684	0,01616	30,135	$< 2e - 16$



Tabela 12 – Estimativas dos parâmetros do modelo Log-Normal para  $\sigma$ .

Variável	Parâmetro	Estimativa	Erro Padrão	valor t	Pr(> t )
Intercepto	$\beta_0$	-1,456300	0,058924	-24,715	$< 2e - 16$
Irradiada	$\alpha_1$	0,157973	0,066640	2,371	0,01783
Itapetinig	$\alpha_2$	-0,027317	0,064929	-0,421	0,67399
Paloma	$\alpha_3$	-0,042103	0,065210	-0,646	0,51856
PeralAC	$\alpha_4$	-0,137313	0,065126	-2,108	0,03509
PeraMga	$\alpha_5$	-0,043674	0,064948	-0,672	0,50136
PeraOri	$\alpha_6$	0,055239	0,064952	0,850	0,39514
PrecOri	$\alpha_7$	-0,090547	0,064793	-1,397	0,16238
Puka	$\alpha_8$	-0,101169	0,064840	-1,560	0,11881
Sanguine	$\alpha_9$	-0,002653	0,065228	-0,041	0,96756
Valen	$\alpha_{10}$	0,194292	0,065330	2,974	0,00296
VALEN2	$\alpha_{11}$	0,191444	0,065468	2,924	0,00348
DAI 24	$\theta_1$	-0,376197	0,053513	-7,030	$2,59e - 12$
DAI 33	$\theta_2$	-0,238544	0,054427	-4,383	$1,21e - 05$
DAI 39	$\theta_3$	-0,426311	0,054232	-7,861	$5,40e - 15$
DAI 46	$\theta_4$	-0,501859	0,054474	-9,213	$< 2e - 16$
DAI 61	$\theta_5$	-0,276206	0,054568	-5,062	$4,42e - 07$
DAI 68	$\theta_6$	-0,137986	0,054312	-2,541	0,01112
DAI 75	$\theta_7$	-0,228896	0,054441	-4,205	$2,70e - 05$

Tabela 13 – Estimativas dos parâmetros do modelo Gama para  $\mu$ .

Variável	Parâmetro	Estimativa	Erro Padrão	valor t	Pr(> t )
Intercepto	$\beta_0$	0,47854	0,01652	28,973	$< 2e - 16$
Irradiada	$\alpha_1$	-0,29417	0,01781	-16,516	$< 2e - 16$
Itapetinig	$\alpha_2$	0,02708	0,01558	1,738	0,0823
Paloma	$\alpha_3$	-0,02697	0,01565	-1,723	0,0849
PeralAC	$\alpha_4$	-0,03412	0,01486	-2,296	0,0217
PeraMga	$\alpha_5$	-0,02178	0,01544	-1,411	0,1585
PeraOri	$\alpha_6$	0,06406	0,01638	3,909	$9,47e - 05$
PrecOri	$\alpha_7$	0,01035	0,01521	0,680	0,4964
Puka	$\alpha_8$	-0,06651	0,01510	-4,404	$1,10e - 05$
Sanguine	$\alpha_9$	-0,14227	0,01588	-8,957	$< 2e - 16$
Valen	$\alpha_{10}$	-0,34077	0,01757	-19,395	$< 2e - 16$
VALEN2	$\alpha_{11}$	-0,34491	0,01767	-19,521	$< 2e - 16$
DAI 24	$\theta_1$	0,11186	0,01486	7,528	$6,93e - 14$
DAI 33	$\theta_2$	0,29303	0,01599	18,328	$< 2e - 16$
DAI 39	$\theta_3$	0,25854	0,01500	17,233	$< 2e - 16$
DAI 46	$\theta_4$	0,26140	0,01464	17,856	$< 2e - 16$
DAI 61	$\theta_5$	0,42012	0,01597	26,313	$< 2e - 16$
DAI 68	$\theta_6$	0,53002	0,01671	31,719	$< 2e - 16$
DAI 75	$\theta_7$	0,47733	0,01618	29,501	$< 2e - 16$

Tabela 14 – Estimativas dos parâmetros do modelo Gama para  $\sigma$ .

Variável	Parâmetro	Estimativa	Erro Padrão	valor t	Pr(> t )
Intercepto	$\beta_0$	-1,474066	0,058462	-25,214	$< 2e - 16$
Irradiada	$\alpha_1$	0,174818	0,066320	2,636	0,008436
Itapetinig	$\alpha_2$	-0,025316	0,064570	-0,392	0,695039
Paloma	$\alpha_3$	-0,025836	0,0648410	-0,398	0,690324
PeralAC	$\alpha_4$	-0,129208	0,064761	-1,995	0,046123
PeraMga	$\alpha_5$	-0,046459	0,064612	-0,719	0,472173
PeraOri	$\alpha_6$	0,073857	0,064585	1,144	0,252898
PrecOri	$\alpha_7$	-0,073064	0,064462	-1,133	0,257125
Puka	$\alpha_8$	-0,089527	0,064516	-1,388	0,165346
Sanguine	$\alpha_9$	-0,003264	0,064890	0,050	0,959886
Valen	$\alpha_{10}$	0,185253	0,064888	2,855	0,004336
VALEN2	$\alpha_{11}$	0,187541	0,065034	2,884	0,003960
DAI 24	$\theta_1$	-0,376251	0,053106	-7,085	$1,75e - 12$
DAI 33	$\theta_2$	-0,233801	0,054155	-4,317	$1,63e - 05$
DAI 39	$\theta_3$	-0,408275	0,053988	-7,562	$5,34e - 14$
DAI 46	$\theta_4$	-0,484505	0,054222	-8,936	$< 2e - 16$
DAI 61	$\theta_5$	-0,261089	0,054297	-4,809	$1,60e - 06$
DAI 68	$\theta_6$	-0,133069	0,053980	-2,465	0,013755
DAI 75	$\theta_7$	-0,208514	0,054192	-3,848	0,000122

Os resultados apresentados nas Tabelas (11) e (13) mostram o quanto diâmetro das lesões causados pela *Xanthomonas* variam em média, em relação a variedade Hamlin e o quanto o diâmetro das lesões variam, em média em ao tempo comparado ao dia 18. Nota-se que as duas variedades de Valência e a Irradiada foram as que apresentaram o menor crescimento médio do diâmetro no decorrer do tempo, enquanto que Pera Ori foi a que apresentou o maior, seguido pela Itapetinig e Prec Ori. Quanto aos dias após a inoculação, nota-se que conforme o tempo passa o diâmetro tem uma tendência a aumentar de tamanho, havendo uma queda entre os dias 33 a 46, em que o dia 68 apresentou as maiores medidas do diâmetro. Tais resultados confirmaram o que a análise descritiva feita anteriormente havia mostrado.

Já os resultados apresentados nas tabelas (12) e (14) para o parâmetro  $\sigma$  das distribuições por si sós não possuem interpretações práticas de forma direta, porém uma boa estimativa feita para esse parâmetro consequentemente melhora as estimativas para o parâmetro  $\mu$  reduzindo seus erros-padrão.

Nas Tabelas (15) e (16), são apresentadas as estimativas das variâncias e covariâncias da parte aleatória do modelo log-normal, nas Tabelas (17) e (18), são apresentadas as estimativas das variâncias e covariâncias da parte aleatória do modelo gama, para os parâmetros  $\mu$  e  $\sigma$ , respectivamente, com seus desvios padrões e intervalos de confiança.

Tabela 15 – Estimativas das variâncias da parte aleatória do modelo log-normal para o parâmetro  $\mu$ .

Variável	Parâmetro	Desvio Padrão	Variância
Folha	$u_0$	0.03629049	0.001317
Residual	$e$	0.1880957	0.03538

Tabela 16 – Estimativas das variâncias da parte aleatória do modelo log-normal para o parâmetro  $\sigma$ .

Variável	Parâmetro	Desvio Padrão	Variância
Folha	$u_0$	0,04293544	0,001843452
Residual	$e$	0,9937678	0,9875744

Tabela 17 – Estimativas das variâncias da parte aleatória do modelo gama para o parâmetro  $\mu$ .

Variável	Parâmetro	Desvio Padrão	Variância
Folha	$u_0$	0.03579106	0.0012815
Residual	$e$	0.1897367	0.03600

Tabela 18 – Estimativas das variâncias da parte aleatória do modelo gama para o parâmetro  $\sigma$ .

Variável	Parâmetro	Desvio Padrão	Variância
Folha	$u_0$	0,04235986	0,001794358
Residual	$e$	1,0006682	1,001337

As estimativas de variância apresentadas na Tabela (15), representam o quanto da variância apresentada na medida do diâmetro é devido ao sujeito (o efeito aleatório). A variável Residual representa a variabilidade que não está associado as folhas, ou seja, é o parâmetro  $e$  do modelo, que representa os desvios aleatórios dos valores preditos não relacionados aos sujeitos. Isso reflete o fato de que toda medida possui alguns fatores que afetam o escopo dos dados, mas que estão fora do alcance do experimento.

As Tabelas (19) e (20) apresentam os valores dos  $R^2$  de Nagelkerke para os modelos log-normal e gama.

Tabela 19 – Valores de  $R^2$  Cox & Snell. e Tabela 20 – Valores de  $R^2$  Cragg & Uhler's.

Modelo	$R^2$
Log-Normal	0,6141053
Gama	0,6129558

Modelo	$R^2$
Log-Normal	0,7563171
Gama	0,7538048

Nestas Tabelas nota-se que os valores do  $R^2$  não foram tão altos quanto o desejado para um modelo relativamente bem ajustado, porém já era algo esperado visto que a análise descritiva feita anteriormente apresentou um comportamento para o diâmetro diferente do esperado para esse tipo experimento, em que o comportamento apresentado entre os dias 33 a 46, e 68 a 75 de decaimento está influenciando em tal medida e não foi explicado pelo pesquisador. Nota-se também que para ambos os modelos os valores do  $R^2$  são próximos.

A Figura (19) apresenta os gráficos de dispersão com linha de igualdade e o gráfico de Bland Altman para os modelos log-normal e gama.

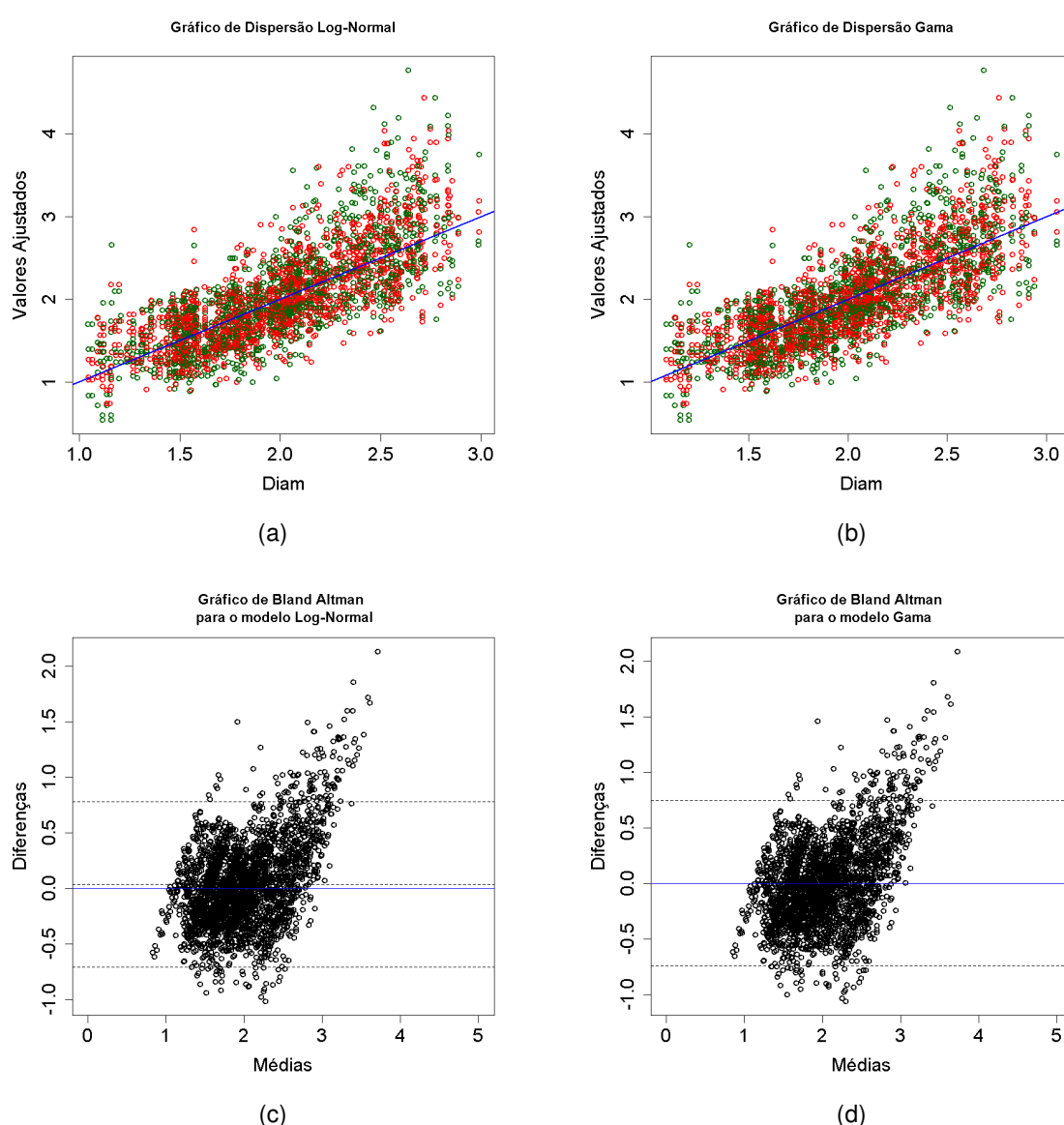


Figura 19 – Gráficos de dispersão com linha de igualdade e Gráficos de Bland Altman para os modelo log-normal e gama.

Nas Figuras (19 (a)) e (19 (b)) tem-se em verde os valores medidos para o diâme-

tro e em vermelho os valores ajustados pelo modelo, em tais gráficos espera-se que tais pontos apresentem medidas iguais, estando portanto se sobrepondo o máximo possível da linha em azul, o que de fato parece estar ocorrendo, ao menos até os valores em torno de 2,5cm. Nas Figuras (19 (c)) e (19(d)) tem-se no eixo  $x$  a média entre os valores medidos e ajustados pelo modelo para o diâmetro e no eixo  $y$  a diferença entre tais valores, em tais gráficos espera-se ver os pontos o mais próximo possível da linha horizontal em azul. As linhas tracejadas representam a média das diferenças e os intervalos superior e inferior para tal média, respectivamente. Portanto, espera-se também que a linha tracejada que representa a média das diferenças esteja próxima a linha em azul, o que ocorre em ambos os modelos.

---

## Capítulo 4

---

---

# Resultados 2<sup>o</sup> Ensaio

---

### 4.1 Análise Descritiva

As distribuições Log-Normal  $LN(\mu, \sigma)$  e Gama  $GA(\mu, \sigma)$  foram utilizadas novamente para analisar o segundo ensaio do banco de dados obtido por (FRANCO, 2016).

Tal ensaio contém 2880 observações do diâmetro médio das lesões (12 variedades  $\times$  8 Dias  $\times$  60 folhas  $\times$  6 Repetições), considerando todas as covariáveis e o efeito do tempo.

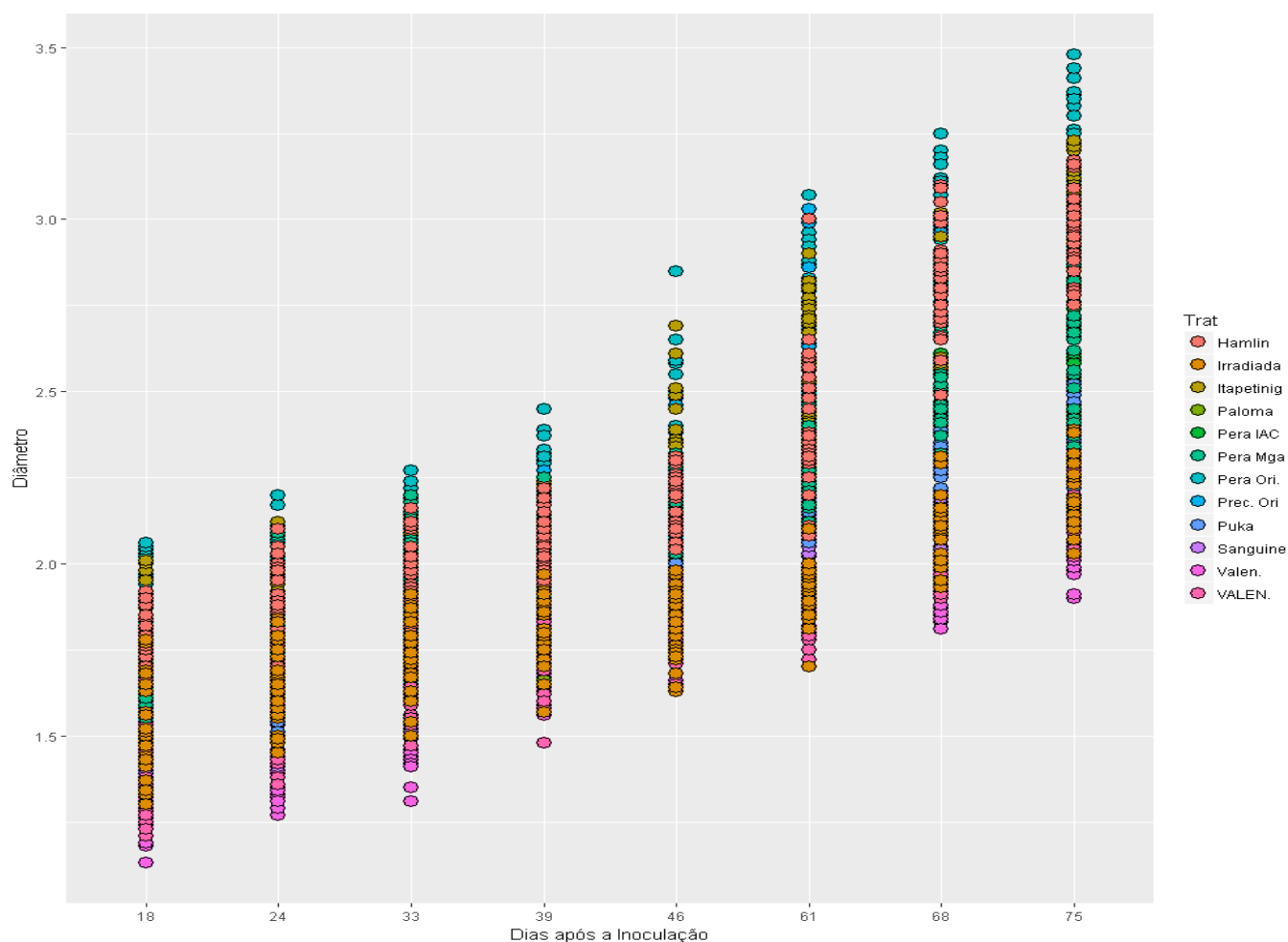


Figura 20 – Gráfico de dispersão dos dados

Na Figura (20) é apresentado o gráfico de dispersão dos dados com cada cor representando uma variedade diferente, onde pode ser notado como o diâmetro da lesão varia em relação a cada variedade e no decorrer do tempo. Aparentemente conforme o tempo passa há um aumento no tamanho do diâmetro, dessa vez sem nenhum decaimento crítico no meio do experimento.

Na Figura (21), nota-se que aparentemente as variedades PrecOri, PeraOri, Itapetinig e Hamlin possuem diâmetros de lesões maiores que as demais variedades, enquanto que as variedades VALEN, Valen e Irradiada possuem os menores. Além disso, se nota que conforme o tempo passa há um aumento no diâmetro em todas as variedades.

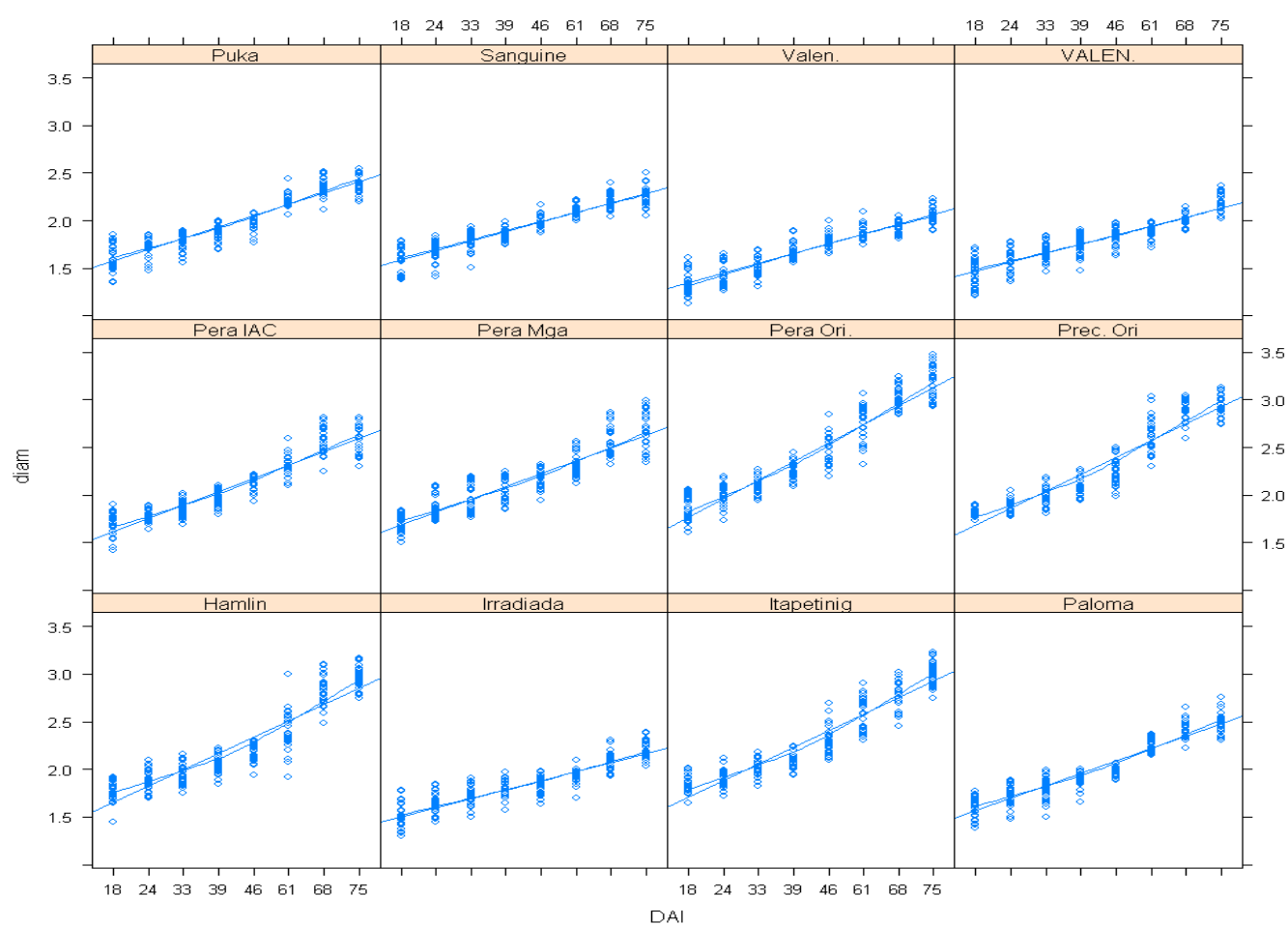


Figura 21 – Gráfico de dispersão por variedade

Na Figura (22) nota-se diferenças no comportamento do diâmetro das lesões em cada variedade, em que pode ser notado poucas variedades com comportamento simétrico.



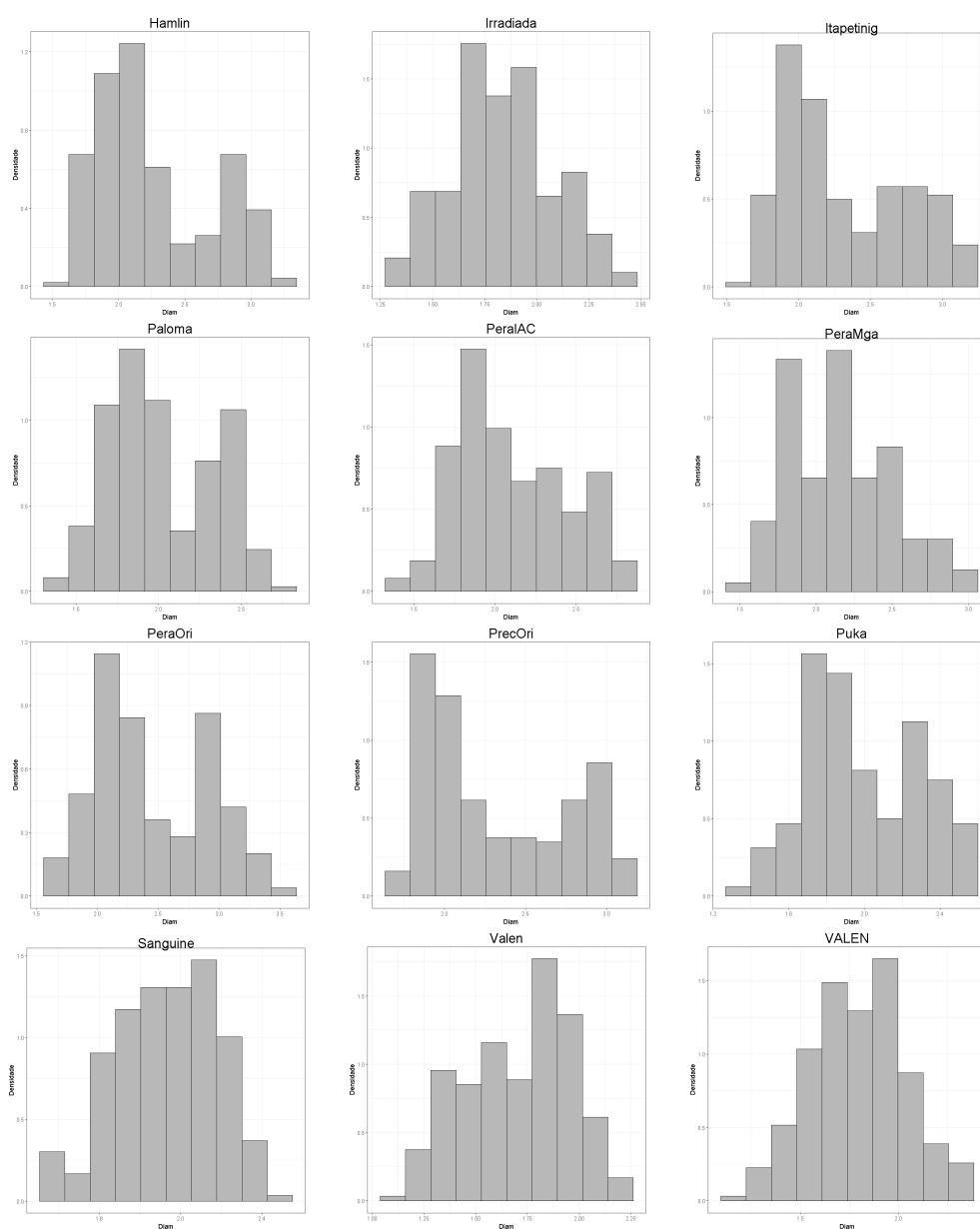


Figura 22 – Histograma da variável resposta diâmetro das lesões para as 12 variedades de laranja doce.

Na Tabela (21) tem-se um resumo da variável resposta diâmetro das lesões em relação a cada variedade:

Tabela 21 – Medidas Resumo para cada variedade.

	Variedades	n	Média	Desvio Padrão	Min.	Max.	Coefficiente de variação
1	Pera Ori.	240	2,45	0,47	1,61	3,48	0,19
2	Valen.	240	1,71	0,25	1,13	2,23	0,15
3	Prec. Ori	240	2,31	0,43	1,73	3,13	0,19
4	Sanguine	240	1,94	0,25	1,38	2,50	0,13
5	Paloma	240	2,02	0,32	1,38	2,76	0,16
6	Pera IAC	240	2,11	0,34	1,42	2,82	0,16
7	Puka	240	1,99	0,29	1,35	2,55	0,15
8	VALEN.	240	1,80	0,24	1,21	2,37	0,13
9	Itapetinig	240	2,32	0,43	1,65	3,23	0,18
10	Pera Mga	240	2,16	0,34	1,50	2,99	0,16
11	Hamlin	240	2,25	0,43	1,45	3,17	0,19
12	Irradiada	240	1,84	0,24	1,30	2,39	0,13

A partir da Tabela (21) nota-se que as médias das variedades variam relativamente bastante, indo de 1,71cm para a variedade Valen a 2,45cm para a variedade Pera Ori, com desvios-padrões iguais a 0,25cm e 0,47cm, respectivamente. Além disso, percebe-se uma grande variabilidade para os valores máximo e mínimo para cada variedade. Nota-se também que os valores dos coeficientes de variação variam de 0,13 a 0,19 representando o quanto, em média, os desvios atingem o valor da média amostral.

A Figura (23) apresenta o *box-plot* para o comportamento da variável resposta tamanho do diâmetro por tipo de variedade:

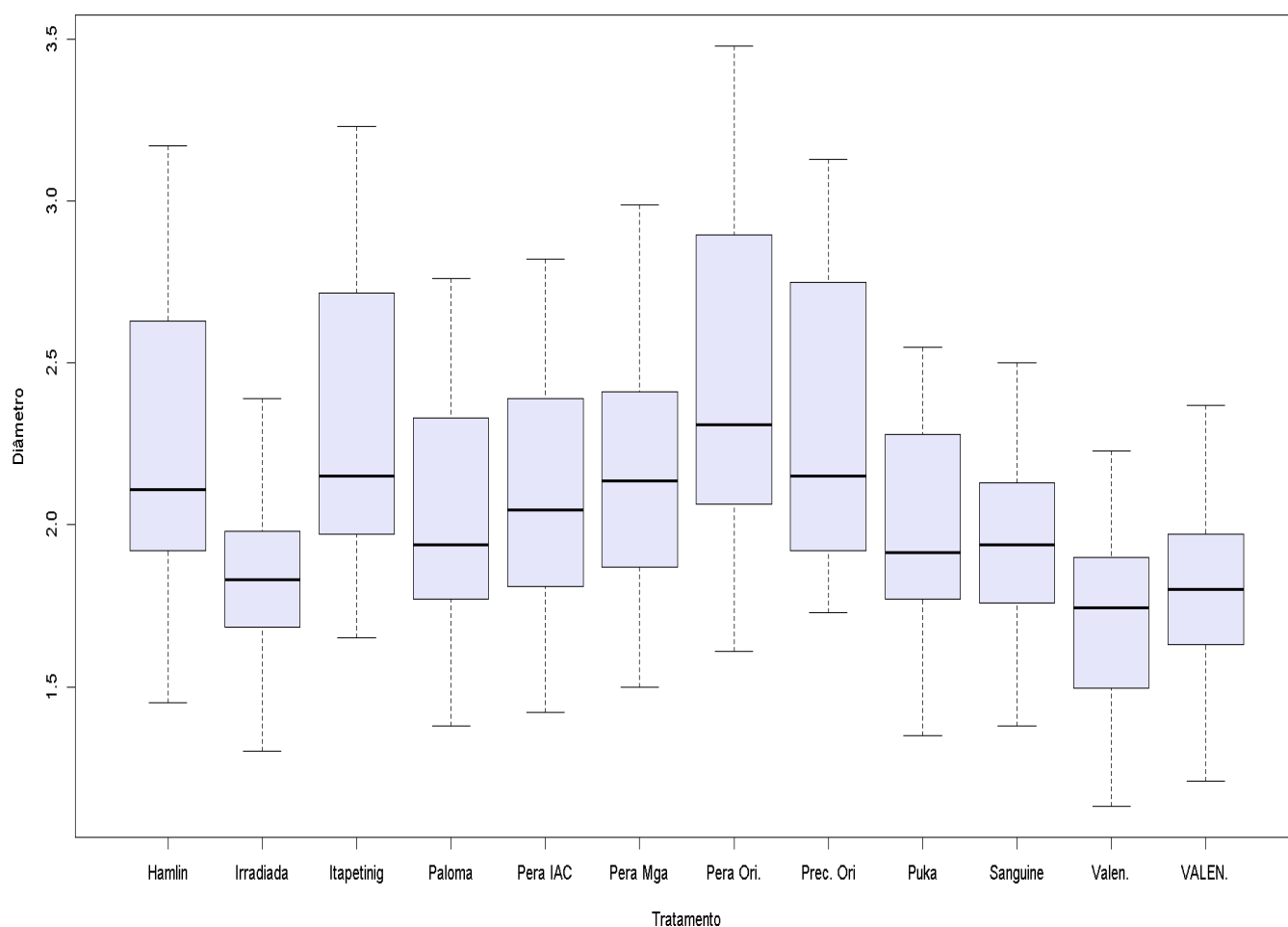


Figura 23 – *Box-plot* dos dados por variedade num total de 2880 observações.

Pela Figura (23) aparentemente a Irradiada e as duas espécies de Valência possuem níveis médios do tamanho do diâmetro menor em relação as outras variedades, enquanto que a Pera Ori. possui o maior nível médio. As outras variedades possuem níveis médios aparentemente próximos. Não se nota medidas candidatas a *outliers*.

A Figura (24) apresenta o *box-plot* do comportamento do diâmetro por dias após a inoculação:

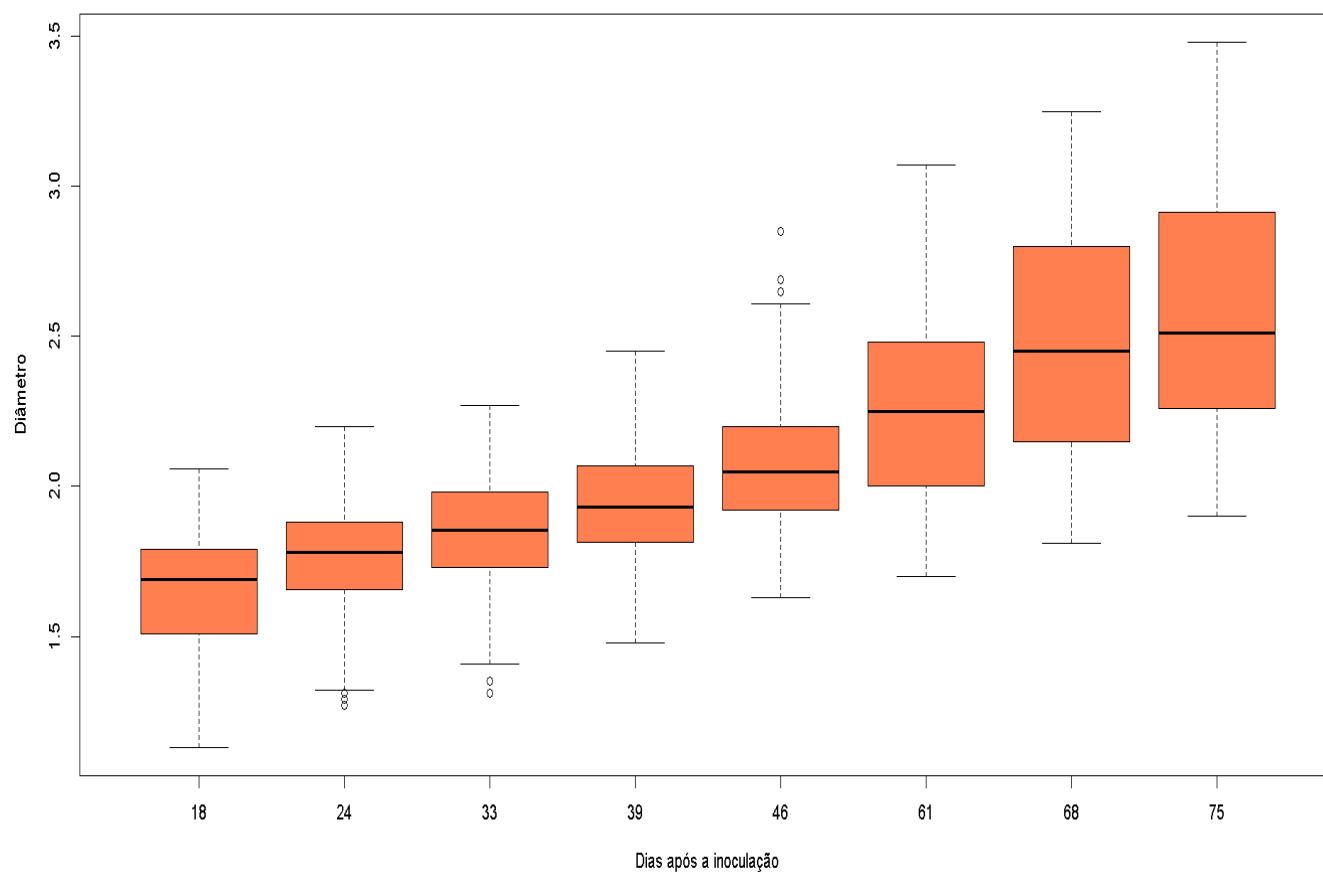


Figura 24 – *Box-plot* dos dados por dias após a inoculação para cada variedade.

Na Figura (24) nota-se um comportamento aparentemente crescente do diâmetro, conforme passa o tempo. Nota-se medidas candidatas a *outliers* apenas no dia 46.

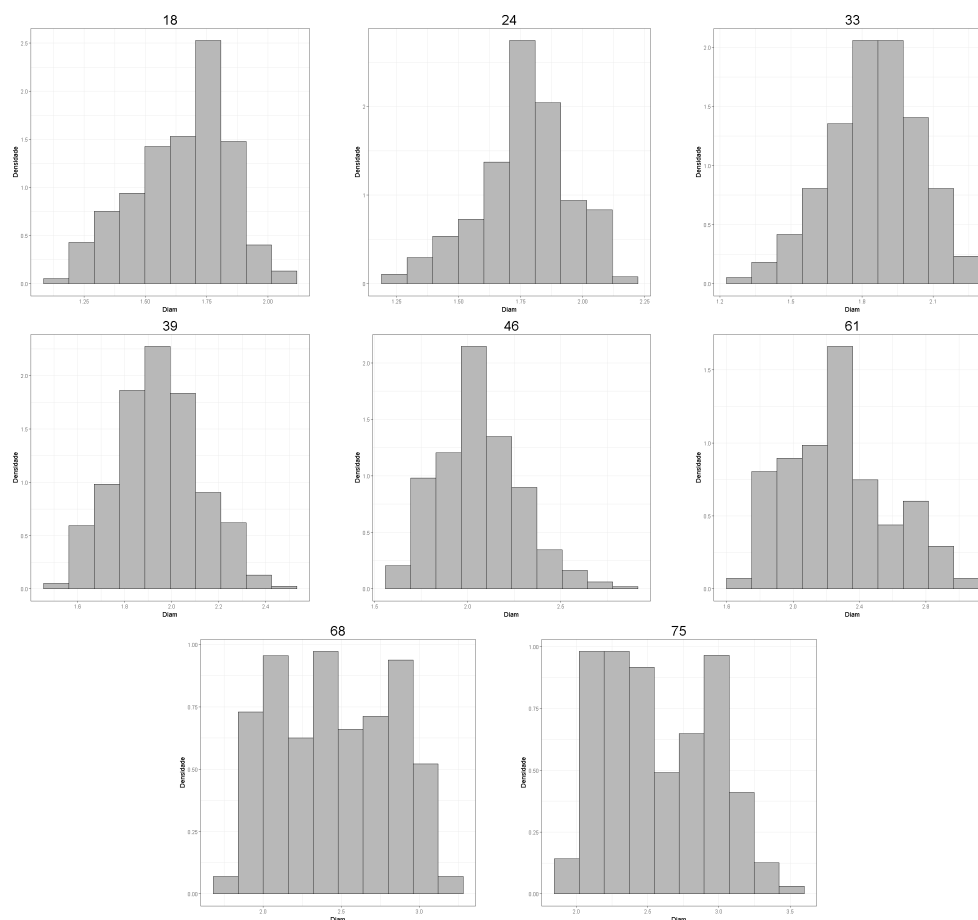


Figura 25 – Histograma da variável resposta diâmetro das lesões para os 8 dias após a inoculação.

Na Figura (25) é apresentado os histogramas para os dias após a inoculação. Podem ser notadas diferenças no comportamento do diâmetro das lesões em todos os dias, havendo assimetria a direita nos dias 46 e 61 e a esquerda nos dias 18, 24 e 33, o que pode indicar que de forma geral, em certos períodos o diâmetro da lesão cresceu menos que em outros.

A Figura (26) apresenta o do comportamento do diâmetro por folha.

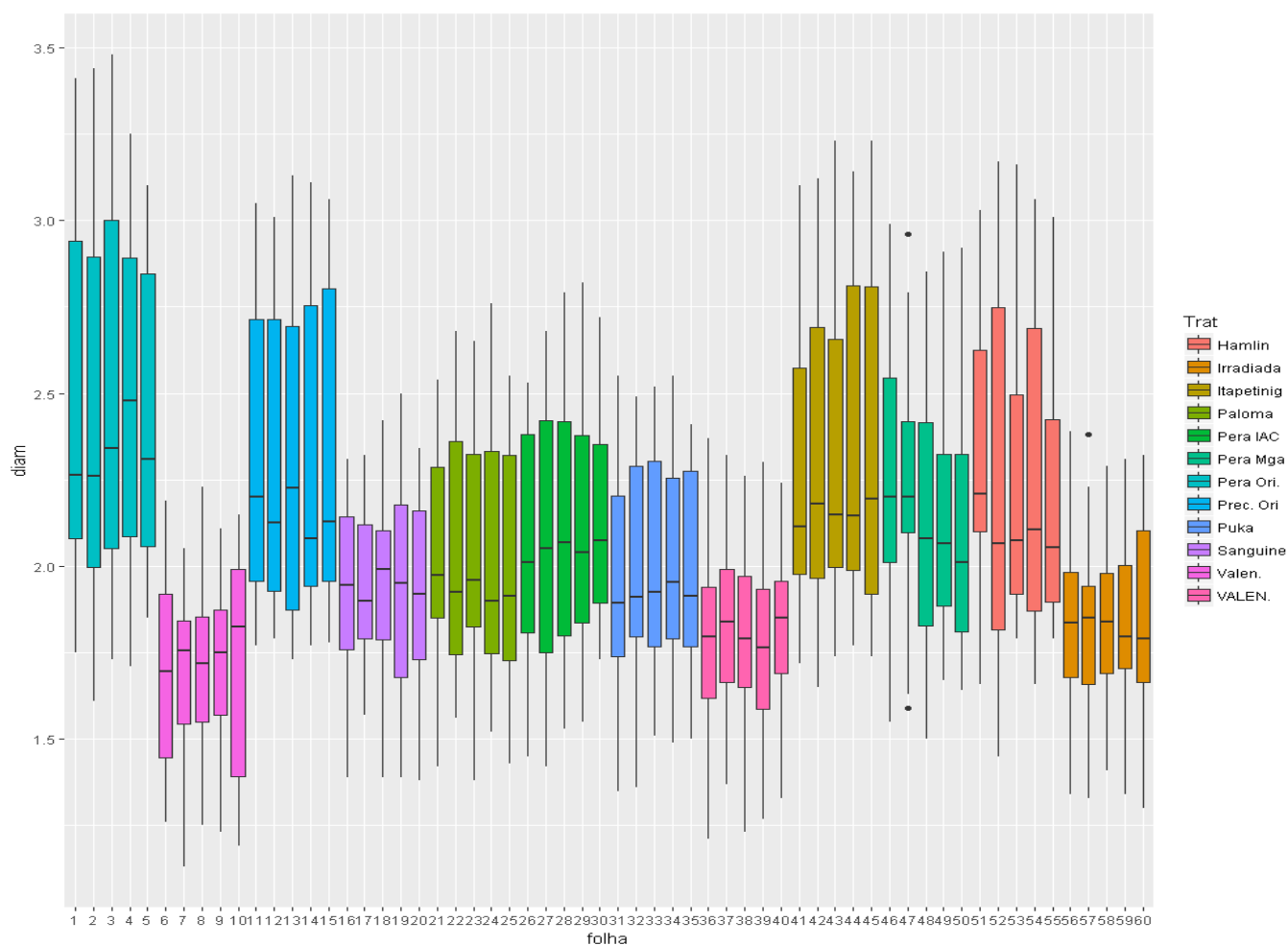


Figura 26 – *Box-plot* dos dados por folha.

Percebe-se um comportamento aparentemente diferente do diâmetro para todas as folhas, havendo variâncias e medianas diferentes de acordo com as variedades e as folhas pertencentes a elas. Nas folhas 47 e 57 das variedades Perâ Maringá e Irradiada, vê-se medidas candidatas a *outliers*. Como esperado pela Figura (23) as folhas das plantas das duas variedades de Valência e Irradiada, aparentemente possuem os menores diâmetros, enquanto que a Pera Ori e Itapetining possuem os maiores.

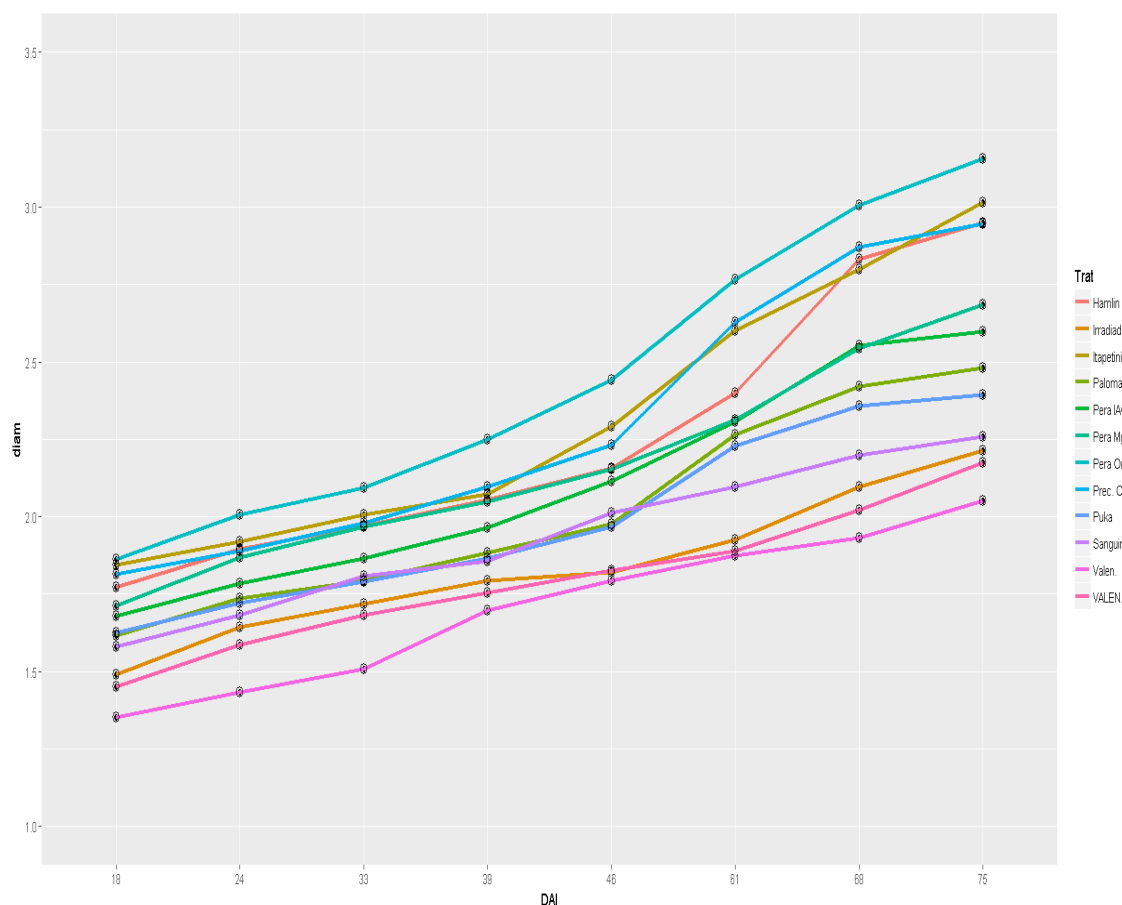


Figura 27 – Perfil Médio em cada tempo para cada variedade

A Figura (27) apresenta os perfis médios das variedades avaliadas. Nota-se que aparentemente, conforme o tempo passa o diâmetro da lesão aumenta em todas as variedades, além do que, nota-se novamente que as duas variedades de Valência e a Irradiada possuem as menores medidas para ele, e a Pera Ori a maior. Não se nota grandes decaimentos relevantes ao longo do tempo como era notado no primeiro ensaio.

## 4.2 Ajuste

No ajuste dos modelos buscou-se novamente analisar o comportamento do cancro cítrico nas 12 variedades de laranja doce, tendo em vista analisar a relação da variável resposta Diâmetro da Lesão aos possíveis fatores de influência, variedade e dias após a inoculação, com estes pertencentes a parte fixa do modelo e as folhas a parte aleatória.

Os modelos foram ajustados na variável resposta  $y_{ijk}$ ,  $i = 1, 2, \dots, 12$  (variedades),  $j = 1, 2, \dots, 8$  (dias após a inoculação),  $k = 1, 2, \dots, 60$  (folhas). A seleção do modelo foi baseada no Critério de Informação de Akaike (AIC), na Deviance Global (GD), no Critério bayesiano de Schwarz (BIC) e no teste de razão de Verossimilhança (LRT). Os modelos ajustados estão descritos nas Tabelas (22) e (23).

Tabela 22 – Modelos propostos para a distribuição Log-normal

	ID	Modelo
1	$LN_0$	$\eta = \beta_0$
2	$LN_1$	$\eta_i = \beta_0 + \alpha_i Variedade$
3	$LN_2$	$\eta_{ij} = \beta_0 + \alpha_i Variedade + \theta_j DAI$
4	$LN_3$	$\eta_{ijk} = \beta_0 + \alpha_i Variedade + \theta_j DAI + u_{0k}$

Tabela 23 – Modelos propostos para a distribuição Gama.

	ID	Modelo
1	$GA_0$	$\eta = \beta_0$
2	$GA_1$	$\eta_{ijk} = \beta_0 + \alpha_i Variedade$
3	$GA_2$	$\eta_{ijk} = \beta_0 + \alpha_i Variedade + \theta_j DAI$
4	$GA_3$	$\eta_{ijk} = \beta_0 + \alpha_i Variedade + \theta_j DAI + u_{0k}$
5	$GA_4$	$\eta_{ijk} = \beta_0 + \alpha_i Variedade + \theta_j DAI + u_{1k} Variedade$

Em tais modelos tem-se que  $\alpha_i$  representa os efeitos devido a  $i$ -ésima variedade,  $\theta_j$  os efeitos devido ao  $j$ -ésimo dia após a inoculação,  $u_k$  é o efeito da parte aleatória para a  $k$ -ésima folha.

Para ambas as Tabelas (22), (23) tem-se que o primeiro modelo é dado apenas pelo intercepto da parte fixa, no segundo é acrescentado o efeito das variedades a parte fixa, tendo como *baseline* a variedade Hamlin. No terceiro modelo é acrescentado o efeito dos dias após a inoculação a parte fixa. No quarto modelo é acrescentado o efeito das folhas por meio de um intercepto aleatório para cada uma delas. No quinto modelo é retirado o intercepto aleatório das folhas e adicionado um coeficiente angular aleatório para cada folha, em cada variedade. Finalmente, no sexto modelo é acrescentado novamente o intercepto aleatório para cada folha.

Os valores de todas as medidas de decisão de cada modelo são apresentados nas Tabelas (24) e (25):

Tabela 24 – Medidas do AIC, Deviance Global e BIC para os modelos Log-normal.

	Modelo	AIC	GD	BIC
1	$LN_0$	2758.685	2754.685	2770.616
2	$LN_1$	1751.555	1725.555	1829.107
3	$LN_2$	-3848.887	-3888.887	-3729.577
4	$LN_3$	-3917.803	-4036.911	-3562.531

Através de tais medidas de decisão e do algoritmo *step GAIC* (STASINOPOULOS;



RIGBY et al., 2007), foi escolhido o modelo  $LN_3$  como o melhor modelo para a distribuição Log-Normal.

Tabela 25 – Medidas do AIC, Deviance Global e BIC para os modelos Gama.

	Modelo	AIC	GD	BIC
1	$GA_0$	2820.178	2816.178	2832.109
2	$GA_1$	1768.072	1742.072	1845.624
3	$GA_2$	-3845.892	-3885.892	-3726.581
4	$GA_3$	-3913.806	-4032.41	-3560.038
5	$GA_4$	-3937.298	-4020.624	-3688.754

Através de tais medidas de decisão e do algoritmo, foi escolhido o modelo  $GA_3$  como o melhor modelo para a distribuição Gama, pois entre os que explicam melhor os dados ele é o mais parcimonioso.

Para os modelos escolhidos foram testadas diversas estruturas de covariância, porém nenhuma melhorou os modelos de forma significativa. Portanto, a matriz de variâncias e covariâncias permaneceu na forma Não-estruturada (2.25).

Os modelos escolhidos para cada distribuição ( $LN_3$ ,  $GA_3$ ) possuem a mesma função de ligação e formulação podendo ser escritos como segue:

$$\log(\mu) = \eta_{ijk\mu} = x_{ij\mu}^T \beta_{\mu} + Z_{k\mu}^T b_{k\mu} \quad (4.1)$$

$$\log(\sigma) = \eta_{ijk\sigma} = x_{ij\sigma}^T \beta_{\sigma^2} + Z_{k\sigma^2}^T b_{k\sigma^2}. \quad (4.2)$$

$$\begin{aligned} \log(\mu) = & \beta_0 + \alpha_1 \text{Irradiada} + \alpha_2 \text{Itapetinig} + \alpha_3 \text{Paloma} + \alpha_4 \text{PeralAC} + \\ & \alpha_5 \text{PeraMga} + \alpha_6 \text{PeraOri} + \alpha_7 \text{PrecOri} + \alpha_8 \text{Puka} + \alpha_9 \text{Sanguine} + \\ & \alpha_{10} \text{Valen} + \alpha_{11} \text{VALEN2} + \theta_1 \text{DAI 24} + \theta_2 \text{DAI 33} + \dots + \\ & \theta_8 \text{DAI 75} + u_0 \text{Folha 1} + \dots + u_0 \text{Folha 60}, \end{aligned} \quad (4.3)$$

e

$$\begin{aligned} \log(\sigma) = & \beta_0 + \alpha_1 \text{Irradiada} + \alpha_2 \text{Itapetinig} + \alpha_3 \text{Paloma} + \alpha_4 \text{PeralAC} + \\ & \alpha_5 \text{PeraMga} + \alpha_6 \text{PeraOri} + \alpha_7 \text{PrecOri} + \alpha_8 \text{Puka} + \alpha_9 \text{Sanguine} + \\ & \alpha_{10} \text{Valen} + \alpha_{11} \text{VALEN2} + \theta_1 \text{DAI 24} + \theta_2 \text{DAI 33} + \dots + \\ & \theta_8 \text{DAI 75} + u_0 \text{Folha 1} + \dots + u_0 \text{Folha 60}. \end{aligned} \quad (4.4)$$

A validade de tais modelos podem ser confirmadas verificando os pressupostos dos mesmos, os quais foram avaliados através da análise de resíduos. Tal análise foi baseada nos resíduos quantílicos aleatorizados e nos gráficos de diagnósticos apresentados nas Figuras (28) e (29) para os modelos Log-Normal e Gama, respectivamente.

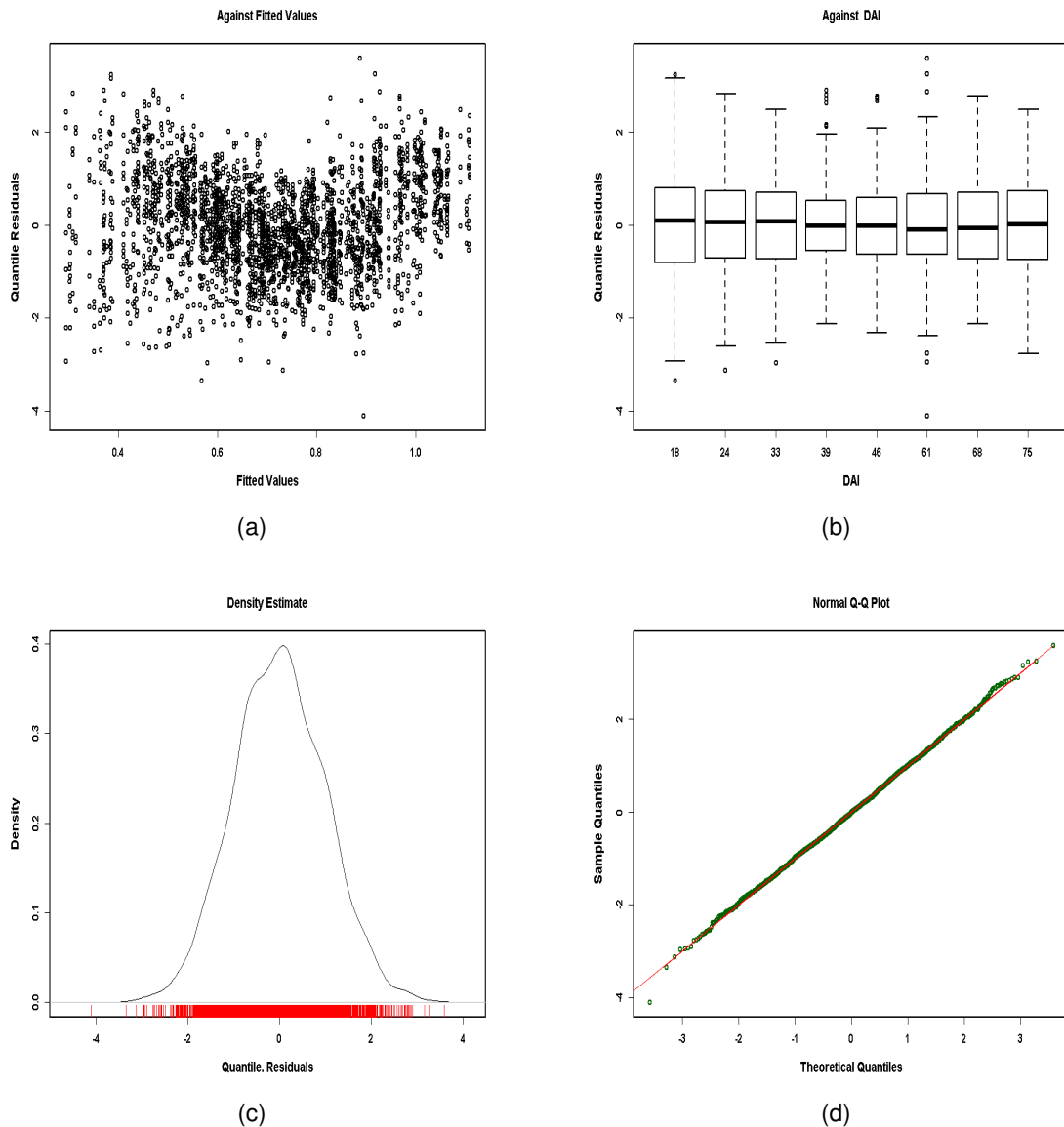


Figura 28 – (a) Valores Ajustados *vs* Quantis Residuais; (b) Dias Após a Inoculação *vs* Quantis Residuais; (c) Quantis Residuais *vs* Densidade Estimada; (d) Quantis Teóricos *vs* Quantis Amostrais.

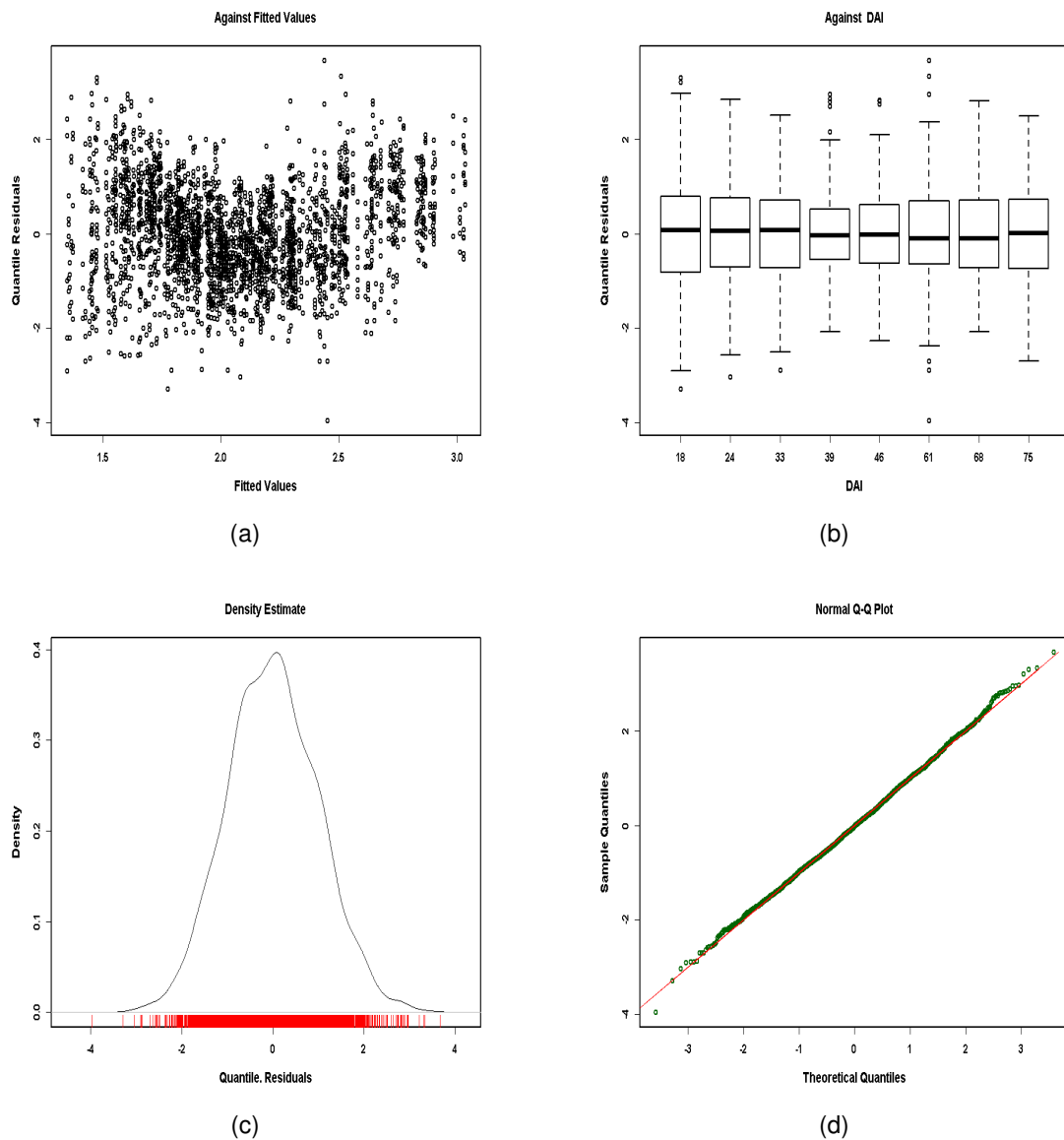


Figura 29 – (a) Valores Ajustados *vs* Quantis Residuais; (b) Dias Após a Inoculação *vs* Quantis Residuais; (c) Quantis Residuais *vs* Densidade Estimada; (d) Quantis Teóricos *vs* Quantis Amostrais.

Nas Figuras (28 (a)), (28 (b)), (29(a)) e (29(b)) dos valores ajustados *vs* quantis residuais e do índice *vs* quantis residuais, mostram que os resíduos estão distribuídos de forma aleatória em torno do zero, o que não revela nenhuma violação quanto a homoscedasticidade. Já os dois últimos quantis residuais *vs* densidade estimada e quantis teóricos *vs* quantis amostrais, indicam que os quantis residuais seguem uma distribuição Normal Padrão, que é requisito para um ajuste adequado.

Nas Tabelas (26) e (27) é apresentado as medidas descritivas dos resíduos quantílicos aleatorizadas para os modelos Log-Normal e Gama, respectivamente.

Tabela 26 – Medidas descritivas dos resíduos para o modelo log-normal.

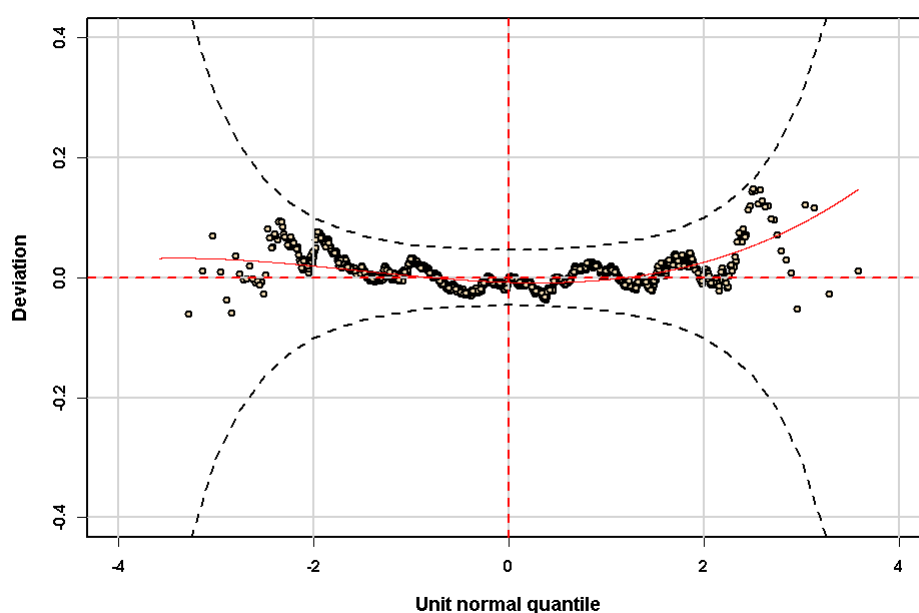
Variável	Parâmetro	$\Pr(> t )$
Média		0
Variância		1.000347
Coef. Assimetria		0.04441846
Coef. Curtose		3.037054

Tabela 27 – Medidas descritivas dos resíduos para o modelo gama.

Variável	Parâmetro	$\Pr(> t )$
Média		0
Variância		1.000347
Coef. Assimetria		0.08670851
Coef. Curtose		3.033534

Conforme já esperado após a análise gráfica, nota-se que as médias são próximas de 0, a variância de 1, assim como a assimetria e a curtose próximas de 0 e 3, respectivamente, conforme esperado para uma variável aleatória com distribuição Normal Padrão, o que indica um ajuste adequado.

As Figuras (30) e (31) apresentam o *Worm plot* dos resíduos quantílicos aleatorizados para os modelos Log-Normal e Gama, respectivamente. Em tais gráficos espera-se ver os pontos o mais próximo possível da linha horizontal em torno do zero e que 95% deles estejam contidos dentro dos intervalos de confiança elípticos.

Figura 30 – *Worm plot* dos resíduos quantílicos aleatorizados para o modelo Log-Normal.

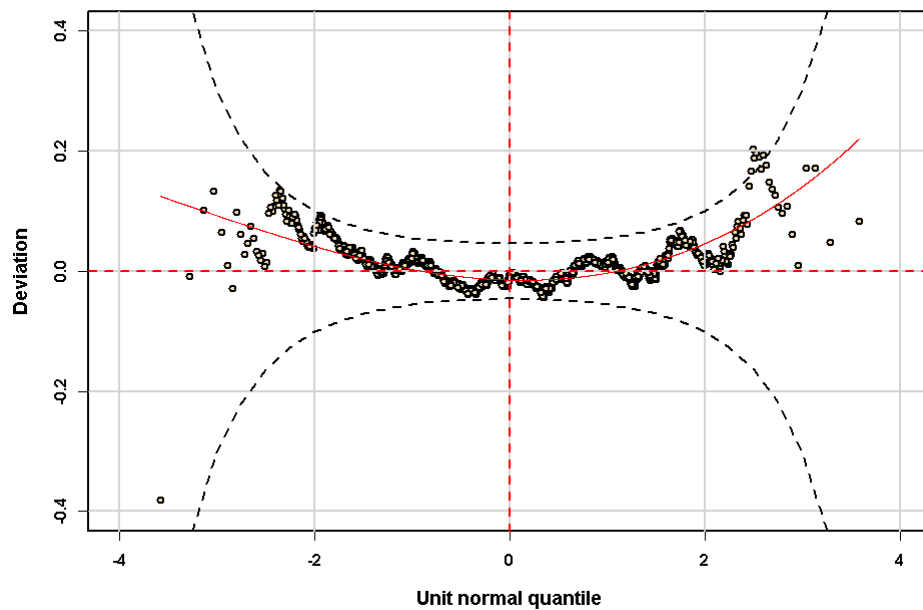


Figura 31 – *Worm plot* dos resíduos quantílicos aleatorizados para o modelo Gama.

Considerando tais Figuras (30) e (31) nota-se que aparentemente os resíduos dos dois modelos são muito parecidos, tendo o modelo Log-Normal, aparentemente, resíduos um pouco melhores que o Gama.

As Figuras (32) e (33) apresentam o gráfico *detrended transformed Owen's* (OWEN, 1995), o qual também é uma ferramenta de diagnóstico para checar se os resíduos quantílicos vem de uma distribuição normal, que é verdadeiro caso a linha horizontal esteja entre os intervalos de confiança, consequentemente a distribuição assumida para a variável resposta é razoável.

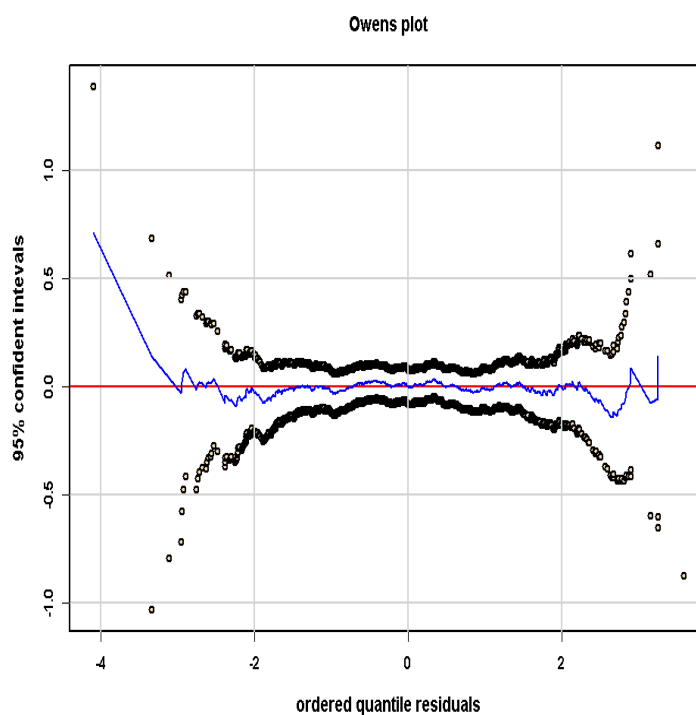


Figura 32 – Gráfico *detrended transformed Owen's* dos resíduos quantílicos aleatorizados para o modelo Log-Normal.

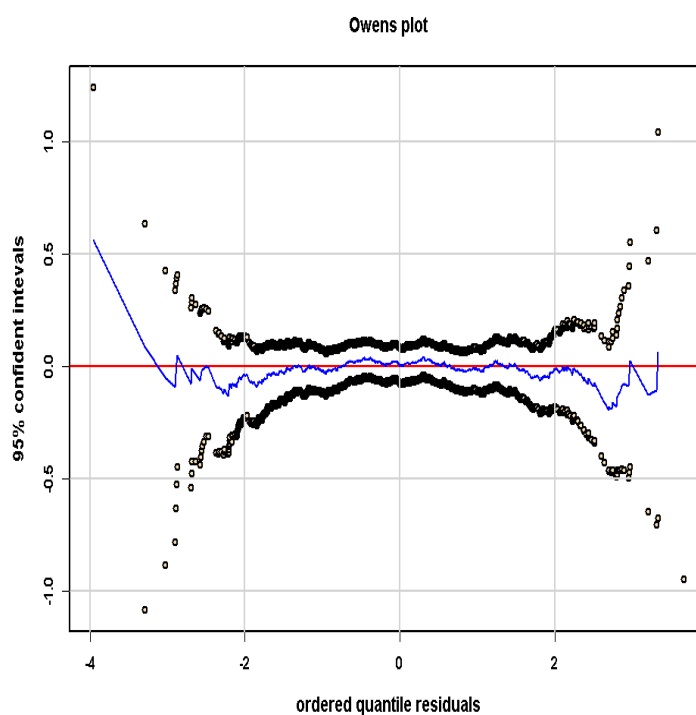


Figura 33 – Gráfico *detrended transformed Owen's* dos resíduos quantílicos aleatorizados para o modelo Gama.

Em tais Figuras (32) e (33) nota-se novamente a proximidade dos resíduos para

os dois modelos, tendo o modelo Log-Normal, aparentemente, a linha de tendência mais próxima da linha horizontal em torno do 0, indicando um melhor ajuste.

No ajuste de tal modelo foi utilizado a função logarítmica como função de ligação para o modelo Log-Normal, assim como para o modelo Gama. Nas Tabelas (28) e (29), são apresentadas as estimativas dos parâmetros fixos do modelo Log-Normal, nas Tabelas (30) e (31) são apresentadas as estimativas dos parâmetros fixos do modelo Gama, para os parâmetros  $\mu$  e  $\sigma$  respectivamente.

Tabela 28 – Estimativas dos parâmetros do modelo Log-Normal para  $\mu$ .

Variável	Parâmetro	Estimativa	Erro Padrão	Valor t	Pr(> t )
Intercepto	$\beta_0$	0.579951	0.005003	115.920	< 2e-16
Irradiada	$\alpha_1$	-0.193554	0.005494	-35.232	< 2e-16
Itapetinig	$\alpha_2$	0.026848	0.005101	5.263	1.52e-07
Paloma	$\alpha_3$	-0.103815	0.004540	-22.867	< 2e-16
PeralAC	$\alpha_4$	-0.061018	0.004500	-13.559	< 2e-16
PeraMga	$\alpha_5$	-0.036685	0.004820	-7.611	3.72e-14
PeraOri	$\alpha_6$	0.084303	0.004976	16.940	< 2e-16
PrecOri	$\alpha_7$	0.022291	0.005081	4.387	1.19e-05
Puka	$\alpha_8$	-0.116129	0.004729	-24.558	< 2e-16
Sanguine	$\alpha_9$	-0.139137	0.005466	-25.456	< 2e-16
Valen	$\alpha_{10}$	-0.266006	0.005804	-45.835	< 2e-16
VALEN2	$\alpha_{11}$	-0.214217	0.005878	-36.443	< 2e-16
DAI 24	$\theta_1$	0.065926	0.004658	14.154	< 2e-16
DAI 33	$\theta_2$	0.111554	0.004628	24.102	< 2e-16
DAI 39	$\theta_3$	0.160661	0.004393	36.570	< 2e-16
DAI 46	$\theta_4$	0.219708	0.004350	50.508	< 2e-16
DAI 61	$\theta_5$	0.319499	0.004647	68.757	< 2e-16
DAI 68	$\theta_6$	0.399900	0.004761	83.994	< 2e-16
DAI 75	$\theta_7$	0.438694	0.004869	90.106	< 2e-16

Tabela 29 – Estimativas dos parâmetros do modelo Log-Normal para  $\sigma$ .

Variável	Parâmetro	Estimativa	Erro Padrão	Valor t	Pr(> t )
Intercepto	$\beta_0$	-2.614833	0.060831	-42.985	$< 2e - 16$
Irradiada	$\alpha_1$	0.079943	0.068198	1.172	0.241211
Itapetinig	$\alpha_2$	0.009492	0.065573	0.145	0.884912
Paloma	$\alpha_3$	-0.247318	0.066694	-3.708	0.000213
PeralAC	$\alpha_4$	-0.258731	0.065844	-3.929	8.72e-05
PeraMga	$\alpha_5$	-0.114171	0.066029	-1.729	0.083903
PeraOri	$\alpha_6$	-0.005291	0.065812	-0.080	0.935933
PrecOri	$\alpha_7$	0.005477	0.065561	0.084	0.933430
Puka	$\alpha_8$	-0.181309	0.067759	-2.676	0.007499
Sanguine	$\alpha_9$	0.037909	0.069156	0.548	0.583620
Valen	$\alpha_{10}$	0.167822	0.068813	2.439	0.014798
VALEN2	$\alpha_{11}$	0.179633	0.067981	2.642	0.008278
DAI 24	$\theta_1$	-0.217608	0.053568	-4.062	5.00e-05
DAI 33	$\theta_2$	-0.239005	0.053694	-4.451	8.88e-06
DAI 39	$\theta_3$	-0.402810	0.054647	-7.371	2.23e-13
DAI 46	$\theta_4$	-0.423484	0.054514	-7.768	1.11e-14
DAI 61	$\theta_5$	-0.262757	0.055851	-4.705	2.67e-06
DAI 68	$\theta_6$	-0.235866	0.055906	-4.219	2.53e-05
DAI 75	$\theta_7$	-0.189956	0.055631	-3.415	0.000648

Tabela 30 – Estimativas dos parâmetros do modelo Gama para  $\mu$ .

Variável	Parâmetro	Estimativa	Erro Padrão	valor t	Pr(> t )
Intercepto	$\beta_0$	0.582542	0.004991	116.719	$< 2e-16$
Irradiada	$\alpha_1$	-0.193053	0.005499	-35.110	$< 2e-16$
Itapetinig	$\alpha_2$	0.026744	0.005089	5.255	1.59e-07
Paloma	$\alpha_3$	-0.104481	0.004521	-23.110	$< 2e-16$
PeralAC	$\alpha_4$	-0.061646	0.004485	-13.745	$< 2e-16$
PeraMga	$\alpha_5$	-0.036918	0.004808	-7.678	2.23e-14
PeraOri	$\alpha_6$	0.084228	0.004963	16.972	$< 2e-16$
PrecOri	$\alpha_7$	0.022138	0.005070	4.367	1.31e-05
Puka	$\alpha_8$	-0.116609	0.004715	-24.731	$< 2e-16$
Sanguine	$\alpha_9$	-0.138832	0.005453	-25.461	$< 2e-16$
Valen	$\alpha_{10}$	-0.265186	0.005814	-45.608	$< 2e-16$
VALEN2	$\alpha_{11}$	-0.213330	0.005880	-36.281	$< 2e-16$
DAI 24	$\theta_1$	0.065074	0.004643	14.017	$< 2e-16$
DAI 33	$\theta_2$	0.110618	0.004614	23.974	$< 2e-16$
DAI 39	$\theta_3$	0.159266	0.004380	36.366	$< 2e-16$
DAI 46	$\theta_4$	0.218289	0.004337	50.329	$< 2e-16$
DAI 61	$\theta_5$	0.318542	0.004649	68.525	$< 2e-16$
DAI 68	$\theta_6$	0.399090	0.004765	83.746	$< 2e-16$
DAI 75	$\theta_7$	0.437979	0.004872	89.889	$< 2e-16$



Tabela 31 – Estimativas dos parâmetros do modelo Gama para  $\sigma$ .

Variável	Parâmetro	Estimativa	Erro Padrão	valor t	Pr(> t )
Intercepto	$\beta_0$	-2.620166	0.060829	-43.074	< 2e-16
Irradiada	$\alpha_1$	0.086521	0.068194	1.269	0.204637
Itapetinig	$\alpha_2$	0.013013	0.065540	0.199	0.842627
Paloma	$\alpha_3$	-0.248261	0.066674	-3.724	0.000200
PeralAC	$\alpha_4$	-0.257490	0.065841	-3.911	9.42e-05
PeraMga	$\alpha_5$	-0.111148	0.066014	-1.684	0.092353
PeraOri	$\alpha_6$	-0.002354	0.065778	-0.036	0.971452
PrecOri	$\alpha_7$	0.009363	0.065521	0.143	0.886384
Puka	$\alpha_8$	-0.179735	0.067777	-2.652	0.008051
Sanguine	$\alpha_9$	0.040322	0.069126	0.583	0.559730
Valen	$\alpha_{10}$	0.176022	0.068772	2.559	0.010536
VALEN2	$\alpha_{11}$	0.185087	0.067956	2.724	0.006497
DAI 24	$\theta_1$	-0.218742	0.053526	-4.087	4.50e-05
DAI 33	$\theta_2$	-0.239531	0.053645	-4.465	8.33e-06
DAI 39	$\theta_3$	-0.403206	0.054595	-7.385	2.00e-13
DAI 46	$\theta_4$	-0.423789	0.054468	-7.780	1.01e-14
DAI 61	$\theta_5$	-0.257065	0.055862	-4.602	4.38e-06
DAI 68	$\theta_6$	-0.229946	0.055887	-4.114	3.99e-05
DAI 75	$\theta_7$	-0.185779	0.055633	-3.339	0.000851

Os resultados apresentados nas tabelas (28) e (30) mostram o quanto diâmetro das lesões causados pela *Xanthomonas* variam em média em relação a variedade Hamlin e o quanto o diâmetro das lesões variam, em média em relação ao tempo comparado ao dia 18. Nota-se que as duas variedades de Valência e a Irradiada foram as que apresentaram o menor crescimento médio do diâmetro no decorrer do tempo, enquanto que Pera Ori foi a que apresentou o maior, seguido pela Itapetinig e Prec Ori. Quanto aos dias após a inoculação nota-se que conforme o tempo passa o diâmetro tem uma tendência a aumentar de tamanho, em que o dia 75 apresentou as maiores medidas do diâmetro. Tais resultados confirmaram o que a análise descritiva feita anteriormente havia mostrado.

Nas Tabelas (32) e (33), são apresentadas as estimativas das variâncias e covariâncias da parte aleatória do modelo Log-Normal, nas Tabelas (34) e (35), são apresentadas as estimativas das variâncias e covariâncias da parte aleatória do modelo Gama, para os parâmetros  $\mu$  e  $\sigma$ , respectivamente.

Tabela 32 – Estimativas das variâncias da parte aleatória do modelo Log-Normal para o parâmetro  $\mu$ .

Variável	Parâmetro	Desvio Padrão	Variância
Folha	$u_0$	0.01367	0.0001869
Residual	$e$	0.05946	0.0035355

Tabela 33 – Estimativas das variâncias da parte aleatória do modelo Log-Normal para o parâmetro  $\sigma$ .

Variável	Parâmetro	Desvio Padrão	Variância
Folha	$u_0$	0.1558522	0.02428991
Residual	$e$	0.922315	0.850665

Tabela 34 – Estimativas das variâncias da parte aleatória do modelo Gama para o parâmetro  $\mu$ .

Variável	Parâmetro	Desvio Padrão	Variância
Folha	$u_0$	0.01356466	0.000184
Residual	$e$	0.05957348	0.003549

Tabela 35 – Estimativas das variâncias da parte aleatória do modelo Gama para o parâmetro  $\sigma$ .

Variável	Parâmetro	Desvio Padrão	Variância
Folha	$u_0$	0.1560059	0.02433784
Residual	$e$	0.9204926	0.8473066

As estimativas de variância apresentadas nas Tabelas (32) e (34), representam o quanto da variância apresentada na medida do diâmetro é devido as folhas e o quanto da variância não pode ser explicada pelo modelo.

As Tabelas (36) e (37) apresentam os valores dos  $R^2$  de Nagelkerke para os modelos Log-Normal e Gama.

Tabela 36 – Valores de  $R^2$  Cox & Tabela 37 – Valores de  $R^2$  Cragg & Uhler's.

Modelo	$R^2$	Modelo	$R^2$
Log-Normal	0.9157485	Log-Normal	1.487184
Gama	0.9175203	Gama	1.470675

Em tais Tabelas nota-se que os valores do  $R^2$  para os dois modelos estão em torno de 0,9, ou seja, tais modelos estão explicando pouco mais de 90% da variabilidade do diâmetro da lesão.

A Figura (34) apresenta os gráficos de dispersão com linha de igualdade e o gráfico de Bland Altman para os modelos log-normal e gama.

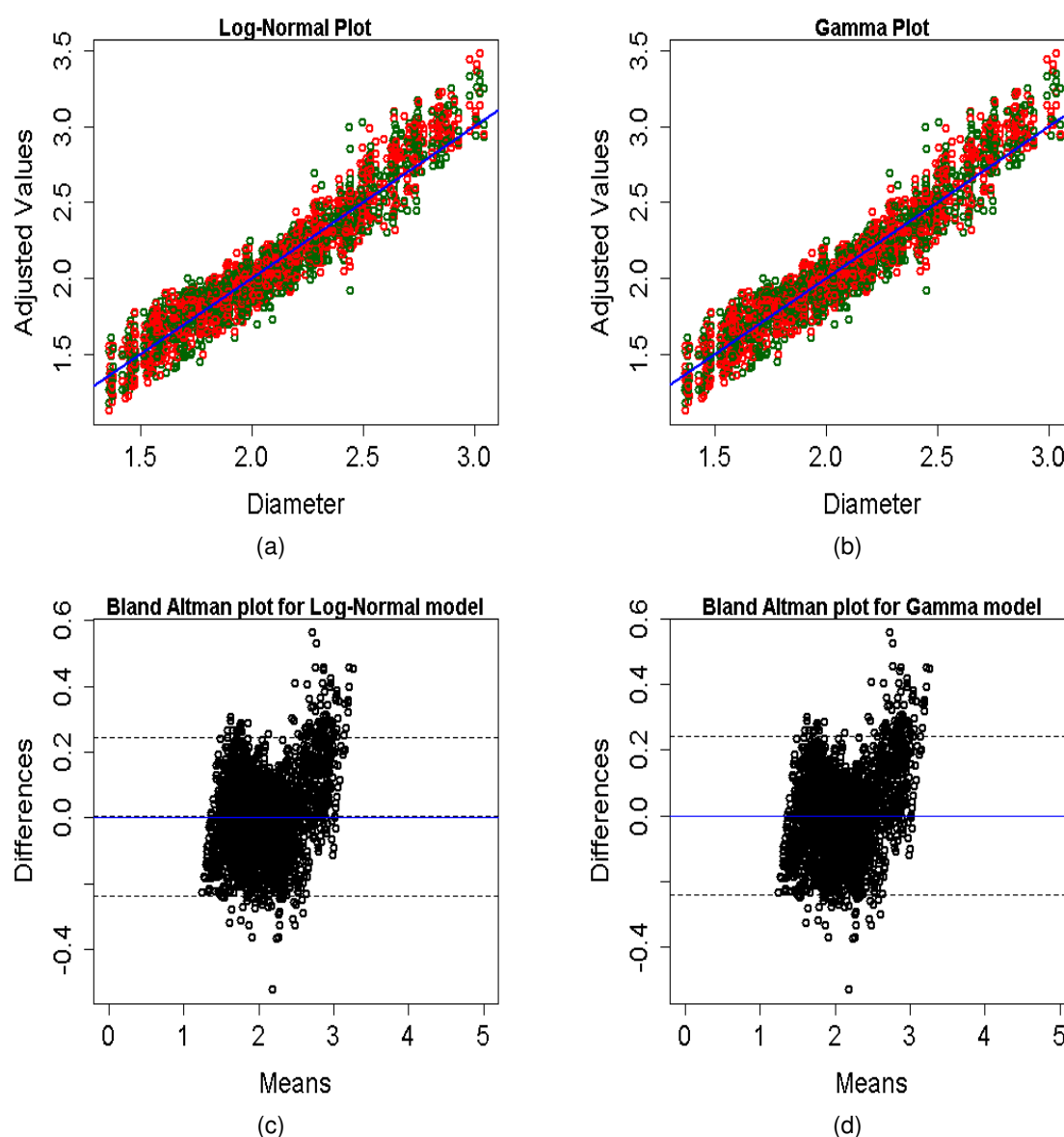


Figura 34 – Gráficos de dispersão com linha de igualdade e Gráficos de Bland Altman para os modelo Log-Normal e Gama.

Nas Figuras (34 (a)) e (34 (b)) tem-se em verde os valores medidos para o diâmetro e em vermelho os valores ajustados pelos modelos, em tais gráficos espera-se que tais pontos apresentem medidas iguais, estando portanto se sobrepondo o máximo possível da linha em azul, o que de fato está ocorrendo. Nas Figuras (19 (c)) e (19(d)) nota-se novamente um bom ajuste, estando a linha da média das diferenças bem próxima de 0 e tendo 95% das diferenças dentro dos intervalos de confiança. Além do que as escalas para as diferenças são próximas de 0.

---

## Capítulo 5

---

---

# Conclusão

---

Neste estudo, foi discutida a metodologia de modelos lineares generalizados mistos para variáveis aleatórias com distribuição Log-Normal e Gama. Esses modelos foram aplicados em dados cujas respostas mostraram adesão a tais distribuições. A ligação entre os preditores e a média da variável de resposta foi feita pela função logarítmica. Os modelos generalizados mistos permitem introduzir uma fonte de dependência que, nos modelos de regressão usuais, não são levados em consideração. Esta dependência é quantificada pela correlação entre as observações que ocorrem nos dados observados ao longo do tempo. A introdução de fatores de efeitos aleatórios no modelo, permite que esta correlação seja modelada. Desta forma, o modelo apresentará melhor ajuste, uma vez que será capaz de capturar a variabilidade dentro dos sujeitos, cujas medidas foram realizadas ao longo do tempo. Portanto, esses modelos foram adequados para descrever o comportamento do cancro cítrico avaliado em variedades de laranja doce.

No experimento que deu origem aos dados analisados, o diâmetro da lesão provocada nas folhas foi medido ao longo do tempo em 12 variedades de laranja doce. O ajuste dos dados observados ao modelo proposto proporcionou resultados adequados, uma vez que os pressupostos necessários para a validade do modelo foram satisfeitos. As duas variedades de Valência apresentaram maior resistência ao cancro cítrico, seguido da variedade Pera Irradiada. Por outro lado, a variedade Pera Oriçanga mostrou-se menos resistente à doença, seguida das variedades Pera Itapetininga e Precoce Oriçanga. Portanto, pode-se dizer que esta metodologia foi adequada para modelar os dados, uma vez que permitiu a detecção das variedades mais suscetíveis ao cancro cítrico, que em experimentos desse tipo, é o objetivo do pesquisador.

---

## Referências

---

- BARBOSA, M. *Uma abordagem para análise de dados com medidas repetidas utilizando modelos lineares mistos*. Tese (Doutorado) — Universidade de São Paulo, 2009. Citado 2 vezes nas páginas 12 e 19.
- BRUGNARA, E. C.; THEODORO, G. d. F.; NESI, C. N.; VERONA, L. A. F.; MARINGONI, A. C. Damage caused by citrus canker on valencia sweet orange yield. *IDESIA (Chile) Enero-Abril, 2012*, v. 30, n. 1, 2012. Citado na página 11.
- BUUREN, S. v.; FREDRIKS, M. Worm plot: a simple diagnostic device for modelling growth reference curves. *Statistics in medicine*, Wiley Online Library, v. 20, n. 8, p. 1259–1277, 2001. Citado na página 25.
- COLE, T. J.; GREEN, P. J. Smoothing reference centile curves: the lms method and penalized likelihood. *Statistics in medicine*, Wiley Online Library, v. 11, n. 10, p. 1305–1319, 1992. Citado na página 23.
- DIGGLE, P. *Analysis of longitudinal data*. [S.l.]: Oxford University Press, 2002. Citado 3 vezes nas páginas 7, 22 e 28.
- DUNN, P. K.; SMYTH, G. K. Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, Taylor & Francis, v. 5, n. 3, p. 236–244, 1996. Citado na página 24.
- FRANCO, J. G. Avaliação da resistência de genótipos de laranja doce (*citrus sinensis*) ao cancro cítrico em condições de casa de vegetação. 2016. Citado 7 vezes nas páginas 8, 11, 14, 15, 16, 41 e 65.
- GINOS, B. F. Parameter estimation for the lognormal distribution. Brigham Young University-Provo, 2009. Citado 2 vezes nas páginas 27 e 28.
- GONÇALVES-ZULIANI, A. M. O.; CARDOSO, K. A. K.; JUNIOR, J. B.; ZANUTTO, C. A.; HASHIGUTI, H. T.; BOCK, C. H.; NAKAMURA, C. V.; NUNES, W. M. d. C. Reaction of detached leaves from different varieties of sweet orange to inoculation with *xanthomonas citri* subsp. *citri*. *Summa Phytopathologica*, SciELO Brasil, v. 42, n. 2, p. 125–133, 2016. Citado na página 11.
- GOTO, M. Citrus canker. in: Plant diseases of international importance. *Englewood Cliffs, NJ: Prentice Hall*, p. 250–269, 1992. Citado na página 10.

- GOTTWALD, T.; GRAHAM, J.; CIVEROLO, E.; BARRETT, H.; HEARN, C. Differential host range reaction of citrus and citrus relatives to citrus canker and citrus bacterial spot determined by leaf mesophyll susceptibility. *Plant Disease*, v. 77, n. 10, p. 1004–1009, 1993. Citado na página 10.
- GOTTWALD, T. R.; GRAHAM, J. H. A device for precise and nondisruptive stomatal inoculation of leaf tissue with bacterial pathogens. *Phytopathology*, v. 82, n. 9, p. 930–935, 1992. Citado na página 10.
- GRAHAM, J.; GOTTWALD, T. et al. Variation in aggressiveness of *Xanthomonas campestris* pv. *citrumelo* associated with citrus bacterial spot in florida citrus nurseries. *Phytopathology*, v. 80, n. 2, p. 190–196, 1990. Citado na página 11.
- GRAHAM, J.; GOTTWALD, T.; RILEY, T.; ACHOR, D. Penetration through leaf stomata and growth of strains of *Xanthomonas campestris* in citrus cultivars varying in susceptibility to bacterial diseases. *Phytopathology*, v. 82, n. 11, p. 1319–1325, 1992. Citado 2 vezes nas páginas 10 e 11.
- HARVEY, G. Multilevel models in educational and social research. Charles Griffin & Company LTD, 1987. Citado na página 7.
- JOHNSON SAMUEL KOTZ, N. B. N. L. *Continuous univariate distributions*. 2. ed. Wiley-Interscience, 1995. (Wiley Series in Probability and Statistics). ISBN 0471584940, 9780471584940. Disponível em: <<http://gen.lib.rus.ec/book/index.php?md5=7126200E2ED74B4084AAE095E8C43A4D>>. Citado na página 33.
- JR, J. B.; FILHO, A. B. Estratégias de controle do cancro cítrico. *Summa Phytopathologica*, v. 32, p. S143–S148, 2006. Citado 2 vezes nas páginas 10 e 11.
- KIMATI, H.; AMORIM, L.; FILHO, A. B.; CAMARGO, L.; REZENDE, J. *Manual de fitopatologia: Volume 2: doenças das plantas cultivadas*. [S.l.]: Agronômica Ceres, 1997. Citado na página 8.
- LAIRD, N. M.; WARE, J. H. Random-effects models for longitudinal data. *Biometrics*, JSTOR, p. 963–974, 1982. Citado na página 17.
- LARANJEIRA, F. F.; AMORIM, L.; FILHO, A. B.; AGUILAR-VILDOSO, C. I.; FILHO, H. D. C.; JUNIOR, D. M.; NEGRI, J. D.; PIO, R.; JUNIOR, J. P. Fungos, procariotos e doenças abióticas. MATTOS JUNIOR, D.; DE NEGRI, JD; PIO, RM, p. 511–566, 2005. Citado 2 vezes nas páginas 8 e 10.
- LEE, Y.; NELDER, J. A. Hierarchical generalized linear models. *Journal of the Royal Statistical Society. Series B (Methodological)*, JSTOR, p. 619–678, 1996. Citado na página 22.
- LIMPERT, E.; STAHEL, W. A.; ABBT, M. Log-normal distributions across the sciences: Keys and clues on the charms of statistics, and how mechanical models resembling gambling machines offer a link to a handy way to characterize log-normal distributions, which can provide deeper insight into variability and probability—normal or log-normal: That is the question. *BioScience*, Oxford University Press, v. 51, n. 5, p. 341–352, 2001. Citado na página 26.

- LINDSEY, J. K. Models for repeated measurements. *Oxford: Clarendon Press*, 1993. Citado na página 7.
- LITTELL, R. C.; PENDERGAST, J.; NATARAJAN, R. Tutorial in biostatistics: modelling covariance structure in the analysis of repeated measures data. *Statistics in medicine*, v. 19, n. 13, p. 1793–1819, 2000. Citado 3 vezes nas páginas 12, 22 e 23.
- LONGFORD, N. T. Random coefficient models. In: *Handbook of statistical modeling for the social and behavioral sciences*. [S.l.]: Springer, 1995. p. 519–570. Citado na página 8.
- MANCO, O. C. U. *Modelos de regressão beta com efeitos aleatórios normais e não normais para dados longitudinais*. Tese (Doutorado) — Universidade de São Paulo, 2013. Citado na página 20.
- NAGELKERKE, N. J. A note on a general definition of the coefficient of determination. *Biometrika*, Oxford University Press, v. 78, n. 3, p. 691–692, 1991. Citado na página 25.
- NOCITI, L. A.; CAMARGO, M.; NETO, J. R.; FRANCISCHINI, F. J.; JR, J. B. Agressividade de linhagens de *xanthomonas axonopodis* pv. *aurantifolii* tipo c em lima ácida 'galego'. *Fitopatologia Brasileira*, SciELO Brasil, v. 31, p. 140–146, 2006. Citado na página 11.
- OWEN, A. B. Nonparametric likelihood confidence bands for a distribution function. *Journal of the American Statistical Association*, Taylor & Francis Group, v. 90, n. 430, p. 516–521, 1995. Citado 2 vezes nas páginas 57 e 80.
- PAWITAN, Y. *In all likelihood: statistical modelling and inference using likelihood*. [S.l.]: Oxford University Press, 2001. Citado na página 22.
- PINHEIRO, J. C. *Topics in mixed effects models*. Tese (Doutorado) — UNIVERSITY OF WISCONSIN–MADISON, 1994. Citado na página 11.
- R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2017. Disponível em: <<https://www.R-project.org/>>. Citado na página 51.
- RAUDENBUSH, A. S. B. S. W. *Hierarchical Linear Models: Applications and Data Analysis Methods*. 2nd. ed. [S.l.]: SAGE Publications, 2002. (Advanced Quantitative Techniques in the Social Sciences 1). ISBN 076191904X, 9780761919049. Citado na página 7.
- RIGBY, R.; STASINOPOULOS, D. A flexible regression approach using gamlss in r. *University of Lancaster: Lancaster, UK*, 2010. Citado na página 24.
- RIGBY, R. A.; STASINOPOULOS, D. M. Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, Wiley Online Library, v. 54, n. 3, p. 507–554, 2005. Citado 2 vezes nas páginas 21 e 51.
- RIGBY, R. A.; STASINOPOULOS, M. D. Mean and dispersion additive models. In: *Statistical theory and computational aspects of smoothing*. [S.l.]: Springer, 1996. p. 215–230. Citado na página 24.
- SCHAAD, N. W.; POSTNIKOVA, E.; LACY, G.; SECHLER, A.; AGARKOVA, I. V.; STROMBERG, P. E.; STROMBERG, V. K.; VIDAVER, A. M. Emended classification of *xanthomonad* pathogens on citrus. *Papers in Plant Pathology*, p. 96, 2006. Citado 2 vezes nas páginas 8 e 14.

- SILVA, C.; GOIS, G.; SILVA, I.; SILVA, L.; ALMEIDA, R.; ABUD, A. Avaliação do potencial de uso do bagaço de laranja visando a produção de etanol 2g. *Blucher Chemical Engineering Proceedings*, v. 1, n. 2, p. 1139–1146, 2015. Citado na página 8.
- SINGER, J. M.; ANDRADE, D. d. Análise de dados longitudinais. *Simpósio Nacional de Probabilidade e Estatística*, Embrapa São Paulo, v. 7, 1986. Citado na página 49.
- STASINOPOULOS, D. M.; RIGBY, R. A. et al. Generalized additive models for location scale and shape (gamlss) in r. *Journal of Statistical Software*, v. 23, n. 7, p. 1–46, 2007. Citado 2 vezes nas páginas 52 e 76.
- STASINOPOULOS, M.; RIGBY, B.; AKANTZILIOTOU, C. Instructions on how to use the gamlss package in r second edition. 2008. Citado 2 vezes nas páginas 26 e 51.
- VILORIA, Z.; DROUILLARD, D.; GRAHAM, J.; GROSSER, J. Screening triploid hybrids of ‘lakeland’limequat for resistance to citrus canker. *Plant Disease*, Am Phytopath Society, v. 88, n. 10, p. 1056–1060, 2004. Citado na página 13.
- WALLACE, D.; GREEN, S. B. Analysis of repeated measures designs with linear mixed models. Lawrence Erlbaum Associates Publishers, 2002. Citado 6 vezes nas páginas 7, 11, 12, 19, 31 e 39.
- WARE, J. H. Linear models for the analysis of longitudinal studies. *The American Statistician*, Taylor & Francis Group, v. 39, n. 2, p. 95–101, 1985. Citado na página 7.
- WINTER, B. A very basic tutorial for performing linear mixed effects analyses. *arXiv preprint arXiv:1308.5499*, 2013. Citado na página 12.