

**MODELOS DE REGRESSÃO PARA A DESCRIÇÃO
DO CRESCIMENTO DO CAFEIRO IRRIGADO E
NÃO IRRIGADO APÓS RECEPA**

JAIR WYZYKOWSKI

2009

JAIR WYZYKOWSKI

**MODELOS DE REGRESSÃO PARA A DESCRIÇÃO DO CRESCIMENTO
DO CAFEIRO IRRIGADO E NÃO IRRIGADO APÓS RECEPA**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-graduação em Estatística e Experimentação Agropecuária, para obtenção do título de "Mestre".

Orientador

Prof. Dr. Augusto Ramalho de Moraes

LAVRAS
MINAS GERAIS -BRASIL
2009

**Ficha Catalográfica Preparada pela Divisão de Processos Técnicos da
Biblioteca Central da UFLA**

Wyzykowski, Jair.

Modelos de regressão para a descrição do crescimento do
cafeeiro irrigado e não irrigado após recepa / Jair Wyzykowski. -
Lavras : UFLA, 2009.

76 p. : il.

Dissertação (Mestrado) - Universidade Federal de Lavras, 2009.

Orientador: Augusto Ramalho de Moraes.

Bibliografia.

1. Regressão polinomial. 2. Regressão não linear. 3. Modelo
misto. 4. Diâmetro de copa. 5. Ajuste de curvas. I. Universidade
Federal de Lavras. II. Título.

CDD-519.536

JAIR WYZYKOWSKI

**MODELOS DE REGRESSÃO PARA A DESCRIÇÃO DO CRESCIMENTO
DO CAFEIEIRO IRRIGADO E NÃO IRRIGADO APÓS RECEPA**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-graduação em Estatística e Experimentação Agropecuária, para obtenção do título de "Mestre".

APROVADA em 04 de dezembro de 2009

Prof. Dr. Fabyano Fonseca e Silva

UFV

Prof. Dr. Natalino Calegário

UFLA/DCF

Prof^a. Dr^a. Taciana Villela Savian

UFLA/PRODOC/CAPES

Prof. Dr. Augusto Ramalho de Moraes

UFLA

(Orientador)

LAVRAS
MINAS GERAIS - BRASIL

AGRADECIMENTOS

Ao professor Augusto Ramalho de Moraes, pelo precioso auxílio, amizade, paciência e compreensão.

Ao professor Paulo Bola, pela oportunidade de trabalho que me foi dada como Tutor no Curso de Pós-Graduação em Matemática e Estatística da Ufla, viabilizando minha estadia em Lavras e conseqüentemente a conclusão deste trabalho.

Aos demais professores do Departamento de Ciências Exatas (DEX) da Ufla por todo ensinamento, troca de experiências e amizade.

Aos amigos Altemir (Bibi), Edcarlos e Ana Patricia, pela recepção e companheirismo durante a minha estadia em Lavras.

Ao amigo Manoel, pelo apoio e auxílio no início do curso.

À Universidade Federal de Lavras e ao Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, bem como a todos os funcionários do DEX-UFLA.

À Josi, sempre disposta a ajudar, desempenhando seu papel de maneira eficiente com grande competência e dedicação.

A todos os colegas da pós-graduação do DEX da UFLA. Especialmente, a Deyse, Walmes, Fábio, Rose e Simone.

Aos amigos baianos Waldete, Vitória, Cleílton, Elma, Tania e Jucilaine.

À Euzi, pelo carinho, apoio e compreensão...

À toda minha família pelo apoio, principalmente à minha mãe pelo incentivo e auxílio em todos os momentos que foram necessários.

SUMÁRIO

LISTA DE TABELAS	i
LISTA DE FIGURAS	iii
RESUMO	vi
ABSTRACT	vii
1 INTRODUÇÃO	1
2 REFERENCIAL TEÓRICO	3
2.1 Importância do cafeeiro	3
2.2 A irrigação do cafeeiro	4
2.3 A modelagem do crescimento de plantas	5
2.4 Modelos de regressão linear	6
2.4.1 O método dos mínimos quadrados e pressuposições para a estimação de β e σ^2	9
2.4.2 O método da máxima verossimilhança (ML)	15
2.4.3 Método da máxima verossimilhança restrita (REML)	17
2.4.4 Análise dos resíduos	18
2.5 Modelo de regressão não linear	20
2.6 O Modelo linear misto	22
2.7 O modelo não linear misto	25
2.8 Análise dos resíduos: o qq-plot	27
2.9 Análise do ajuste da curva pelo R^2	27
3 MATERIAL E MÉTODOS	29
3.1 Realização do experimento, delineamento experimental e tratamentos	29
3.2 Os modelos testados	30
3.2.1 Modelo linear polinomial	30

3.2.2 Modelo linear com efeito misto	31
3.2.3 Modelo não linear	32
3.2.4 Modelo não linear com efeito misto	33
4 RESULTADOS E DISCUSSÃO	35
4.1 Ajuste do modelo linear polinomial	36
4.2 O ajuste do modelo linear com efeitos mistos	40
4.3 O ajuste do modelo não linear	43
4.4 O ajuste dos modelos não linear com efeitos mistos	51
4.4.1 O modelo Logístico	51
4.4.2 O modelo de Gompertz	56
4.5 Considerações finais	62
5 CONCLUSÃO	64
REFERÊNCIAS	65
ANEXOS	69

LISTA DE TABELAS

1	Estimativas dos coeficientes de regressão referentes ao ajuste do modelo linear polinomial a dados do diâmetro de copa de cafeeiro para os cinco tratamentos.	37
2	Estimativas dos parâmetros para o ajuste do modelo não linear logístico ao diâmetro de copa para os cinco tratamentos.	45
3	Estimativas dos parâmetros para o ajuste do modelo não linear Gompertz ao diâmetro de copa para os cinco tratamentos.	49
4	Comparação entre o modelo com modelagem da variância (CM) e sem modelagem de variância (SM).	51
5	Resumo da análise da variância do modelo CM para o crescimento do diâmetro de copa do cafeeiro ao longo do tempo para os tratamentos (Logístico).	51
6	Valores dos contrastes que foram significativos entre tratamentos ao nível de 5% para o modelo logístico.	53
7	Estimativas dos parâmetros assíntota (A), ponto de inflexão (I), escala (S) e seus respectivos erros padrão e intervalos de confiança (95%), referentes ao ajuste do modelo logístico com efeitos mistos a dados do diâmetro de copa do cafeeiro.	56
8	Resumo da análise da variância para o crescimento do diâmetro de copa do cafeeiro ao longo do tempo para os tratamentos (Gompertz).	57
9	Valores dos contrastes que foram significativos entre tratamentos em nível de 1% para o modelo Gompertz.	58

10	Estimativas dos parâmetros (A), (I), (S) e seus respectivos erros padrão e intervalos de confiança (95%), referentes ao ajuste do modelo gompertz a dados do diâmetro de copa do cafeeiro significativas a 1%.	62
----	--	----

LISTA DE FIGURAS

1	Representação geométrica do método dos mínimos quadrados mostrando o vetor de observações y , o espaço gerado ou explicado pelo delineamento $X\hat{\beta}$ e o vetor dos erros $y - X\hat{\beta}$	13
2	Representação gráfica dos valores observados do diâmetro da copa com relação ao tempo de avaliação para os cinco tratamentos . . .	36
3	Representação gráfica da relação entre os quantis observados e os quantis teóricos esperados relativos aos resíduos do ajuste do modelo linear polinomial para os cinco tratamentos; histograma dos resíduos; estatística do teste de normalidade de Shapiro-Wilk. . . .	38
4	Representação gráfica do modelo linear polinomial, intervalos de confiança e estimativas do coeficiente de determinação (R^2) ajustado para os cinco tratamentos.	39
5	Representação gráfica da relação entre os quantis observados e os quantis teóricos esperados relativos aos resíduos do ajuste do modelo linear misto (a) e resíduos padronizados em função do tempo (b).	41
6	Representação gráfica dos valores ajustados utilizando o modelo linear misto para cada tratamento.	41
7	Representação gráfica dos valores ajustados para cada indivíduo (ou planta) utilizando o modelo linear de efeitos mistos.	42
8	Representação gráfica do modelo não linear logístico para os cinco tratamentos.	44

9	Representação gráfica dos intervalos de confiança das estimativas dos parâmetros escala (S), de inflexão (I) e assíntota (A) para o ajuste do modelo não linear logístico ao diâmetro de copa para os cinco tratamentos.	45
10	Representação gráfica da relação entre os quantis observados e os quantis teóricos esperados relativos aos resíduos do ajuste do modelo não linear logístico e estatística do teste de normalidade de Shapiro-Wilk para os cinco tratamentos.	46
11	Representação gráfica do modelo não linear de Gompertz para os cinco tratamentos.	47
12	Representação gráfica dos intervalos de confiança das estimativas dos parâmetros assíntota (A), parâmetro numérico relacionado ao valor da função quando o tempo tende a zero $t = 0$ (B) e, parâmetro numérico para a escala no eixo do tempo t (C) para o ajuste do modelo não linear Gompertz ao diâmetro de copa para os cinco tratamentos.	48
13	Representação gráfica da relação entre os quantis observados e os quantis teóricos esperados relativos aos resíduos do ajuste do modelo não linear de Gompertz e estatística do teste de normalidade de Shapiro-Wilk.	50
14	Representação gráfica do intervalo de confiança relativo aos parâmetros assíntota (A), Inflexão (I) e Escala (S) do modelo logístico, para os cinco tratamentos.	52

15	Representação gráfica dos quantis observados padronizados em função dos quantis esperados para a distribuição normal padrão para cada termo aleatório do modelo considerado (a) e diagrama de dispersão dos resíduos em função do tempo (b).	54
16	Representação gráfica dos valores preditos para o crescimento do diâmetro da copa do cafeeiro em função do tempo com os efeitos fixos do modelo, para cada manejo de irrigação.	54
17	Representação gráfica dos valores preditos para o crescimento do diâmetro da copa de nove plantas de cafeeiro em função do tempo, utilizando o modelo logístico com efeitos mistos.	55
18	Representação gráfica dos intervalos de confiança relativos ao parâmetro assíntota (A), parâmetro numérico relacionado ao valor da função quando o tempo tende a zero $t = 0$ (B) e, parâmetro numérico para a escala no eixo do tempo t (C).	57
19	Representação gráfica dos quantis observados padronizados em função dos quantis esperados para a distribuição normal padrão para cada termo aleatório do modelo considerado (a) e diagrama de dispersão dos resíduos em função dos valores ajustados (b). . .	59
20	Representação gráfica dos valores preditos para o crescimento do diâmetro da copa do cafeeiro em função do tempo com os efeitos fixos do modelo, para cada manejo de irrigação.	60
21	Representação gráfica dos valores preditos para o crescimento do diâmetro da copa de nove plantas de cafeeiro em função do tempo, utilizando o modelo de Gompertz com efeitos mistos.	61

RESUMO

WYZYKOWSKI, Jair. **Modelos de regressão para descrição do crescimento do cafeeiro irrigado e não irrigado após recepa**. 2009. 76 p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras.*

O cultivo do café irrigado tem se expandido em todo território nacional, principalmente em locais onde a produção é mais significativa como Minas Gerais, Espírito Santo e São Paulo. Assim, a irrigação tem sido uma grande aliada do produtor, visto que, se bem conduzida, tem reduzido perdas e aumentado a rentabilidade. O objetivo deste estudo foi testar modelos lineares e não lineares com efeitos fixos e aleatórios como uma alternativa para avaliação do crescimento do diâmetro de copa do cafeeiro sob diferentes manejos de irrigação. Foram utilizados dados provenientes de um experimento realizado na área do Departamento de Engenharia da UFLA. O delineamento foi o de blocos casualizados. Avaliou-se periodicamente ao longo de 1100 dias o crescimento da copa das plantas. Observou-se que os modelos não lineares logístico e de Gompertz são adequados para a descrição do desenvolvimento do diâmetro de copa do cafeeiro e que o uso de modelos de efeito misto é viável, tanto para ajustes lineares quanto para não lineares. Os modelos não linear são preferidos dado que seus parâmetros tem interpretação biológica.

Palavras-chave: Regressão polinomial; regressão não linear; modelo misto; diâmetro de copa; ajuste de curvas.

***Orientador:** Augusto Ramalho de Moraes - UFLA.

ABSTRACT

WYZYKOWSKI, Jair. **Regression models for growth description of the irrigated and unirrigated pruned coffee tree.** 2009. 76 p. Dissertation (Master of Statistics and Agricultural Experimentation) Federal University of Lavras, Lavras.*

The irrigated coffee crop grown has expanded throughout the national territory, especially where production is more significant as Minas Gerais, Espírito Santo and São Paulo. Thus, irrigation has been a great producer ally, because, if well conducted, has reduced losses and increased profitability. The aim of this study was to test the linear and nonlinear mixed-effects models as an alternative for assessing the canopy diameter growth of coffee under different irrigation management schemes. We used data from an experiment conducted in the UFLA Engineering Department. The design was a randomized block. The canopy plant growth was measured periodically over 1100 days. It was observed that nonlinear logistic and Gompertz models are suitable for describing the canopy diameter growth of coffee and the use of mixed effect models is feasible for linear and nonlinear fitted models. The nonlinear models are preferred because its parameters have biological meaning.

Key-words: Polynomial regression; nonlinear regression; mixed model; plant canopy growth; curve fitting.

***Guidance:** Augusto Ramalho de Morais - UFLA (Major Professor).

1 INTRODUÇÃO

As inovações tecnológicas tem contribuído para um aumento substancial na produtividade agrícola. Entre as culturas beneficiadas com estas inovações está o café que tem recebido mais recentemente investimentos em irrigação.

O cultivo do café irrigado tem se expandido em todo o Brasil, principalmente em locais em que a produção é significativa como Minas Gerais, Espírito Santo e São Paulo. Recentemente, outros Estados ingressaram nesta lista, como a Bahia, Goiás e Mato Grosso. A irrigação, utilizada de forma correta, pode ser uma grande aliada do produtor, com a redução de perdas e aumento da rentabilidade. No entanto, para que a técnica da irrigação seja eficiente é necessário conhecer a forma como ocorre o desenvolvimento das plantas.

Assim, identificar, descrever e prever as relações entre eventos que envolvem o desenvolvimento do cafeeiro sob diferentes formas de irrigação é de fundamental importância. Uma das formas é encontrar modelos que relacionem variáveis que descrevam a realidade, avaliando a possível relação entre uma variável dependente com uma ou mais variáveis independentes. Isto pode ser conseguido por meio da utilização de modelos de regressão lineares ou não lineares.

A decisão de qual modelo utilizar cabe ao pesquisador que, inicialmente, de forma intuitiva, com base no conhecimento do fenômeno em estudo e na análise estatística, decide qual modelo pode melhor descrever o desenvolvimento da planta. Entre os modelos destacam-se os de regressão linear e não linear. O método de regressão linear já é bem consolidado e existem inúmeros modelos já testados tanto na avaliação do crescimento de cafeeiro como de outras plantas, entretanto, muitos destes não são capazes de refletir a realidade do comportamento do objeto em estudo, resultando em apenas uma equação matemática sem sentido prático.

Ademais, apenas o ajuste não garante a validade sendo necessária a observação de uma série de pressupostos referentes a autocorrelação e homoscedasticidade dos resíduos e a multicolinearidade nos dados.

Os modelos são não lineares quando a derivada da função com relação aos parâmetros depende de um ou mais parâmetros. Alguns modelos não lineares podem ser linearizados através de transformações, entretanto, nem sempre as transformações são possíveis e mesmo quando elas são possíveis as interpretações podem não levar a resultados satisfatórios para as situações práticas.

As respostas medidas para cada tratamento, quando nos referimos ao crescimento são geralmente não lineares e constituem medidas repetidas no tempo dentro da mesma unidade experimental. Para esse tipo de análise, frequentemente consideram-se assumir os tratamentos e blocos como sendo de efeitos fixos e as plantas (indivíduos) medidas dentro das unidades experimentais como sendo em efeito aleatório (PEEK et al., 2002), caracterizando assim um modelo misto. Assim, para a avaliação do crescimento de plantas, foram testados modelos lineares e não lineares com efeitos fixos e aleatórios. Os modelos não lineares ajustados foram o Logístico e o Gompertz.

O objetivo deste estudo foi utilizar modelos com efeitos mistos lineares e não lineares como alternativa para avaliação do crescimento do diâmetro de copa do cafeeiro sob diferentes manejos de irrigação.

2 REFERENCIAL TEÓRICO

2.1 Importância do cafeeiro

O Brasil foi em 2007 o maior produtor de café em grão do mundo com uma produção de aproximadamente 2.249.010 toneladas, o que equivale a 28,86% da produção mundial. Em seguida vem o Vietnã com 12,33% seguido da Colômbia com 8,95% (Food and Agriculture Organization of the United Nations - FAO, 2009). O Brasil também liderou as exportações em 2006 participando com 25,6% das exportações mundiais seguido do Vietnã com 22,9% e Colômbia com 18,2% (FAO, 2009).

O café foi introduzido no Brasil por Francisco de Mello Palheta a partir de sementes trazidas da Guiana Francesa para Belém do Pará (Custódio, 2009), e em 1825 chegou a Minas Gerais. Em 2007, Minas Gerais foi o líder nacional absoluto na produção, produzindo cerca de 43,9% do total produzido, seguido do Espírito Santo com 27,5% e São Paulo com 10,4% (Instituto Brasileiro de Geografia e Estatística - IBGE, 2009).

O cafeeiro é uma planta de porte arbustivo ou arbóreo, caule lenhoso e lignificado, apresentando ramo vertical (tronco) denominado de ortotrópico e ramos laterais (produtivos) chamados de plagiotrópicos. Pertencente à família Rubiaceae, a seção *Eucoffea* é considerada como a de maior importância econômica do gênero *Coffea*, compreendendo as espécies que possuem mais cafeína e são as mais cultivadas, *Coffea arabica* L. e *Coffea canephora* Pierre (Matiello et al., 2002).

As linhagens de Acaiá são oriundas de seleções do cruzamento entre Sumatra e Bourbon Vermelho. Suas características principais são o alto vigor, boa produtividade, frutos de coloração vermelha, porte alto com arquitetura cônica e menor diâmetro de copa (Matiello et al., 2002).

No ano de 1995 foi lançado em Minas Gerais pelo sistema de pesquisa, uma cultivar denominada de Acaia Cerrado derivada da Acaia (LEP 474-19). A progênie mais promissora foi a MG-1474 (Mendes & Guimarães, 1996). Os investimentos em pesquisa crescem a cada dia como um indicativo de que, realmente esta cultura tem destaque dentre as mais importantes para o Brasil.

2.2 A irrigação do cafeeiro

As mudanças climáticas e o avanço do plantio de café em regiões onde tradicionalmente não existia o cultivo como o Cerrado, Triângulo Mineiro e Oeste Baiano fazem com que cada vez mais seja utilizada a irrigação, com a qual tem-se maior segurança de colheita, produtividades elevadas que compensam os investimentos e conseqüentemente um produto de melhor qualidade. Assim, o suprimento de água ao cafeeiro é fundamental para seu pleno desenvolvimento evitando-se o comprometimento da produção (Matiello et al., 2002).

Em regiões tradicionalmente produtoras, como o sul de Minas Gerais, a ocorrência de veranicos nas fases críticas do desenvolvimento da cultura tem causado, em muitas lavouras, uma redução significativa na produção. As deficiências hídricas na fase de frutificação ou expansão afeta o desenvolvimento dos grãos e, se ocorrerem na fase de granação, os mesmos poderão ficar chochos ou mal granados (Camargo, 1989).

Além da formação dos grãos, a deficiência hídrica também pode comprometer o crescimento do cafeeiro. Este fenômeno foi observado por Araújo (1982) que avaliou o crescimento do diâmetro de caule e da copa comparando tratamentos com e sem irrigação. Matiello & Dantas (1987) verificaram que o crescimento do diâmetro de copa e altura da planta foram muito superiores no cafeeiro com a irrigação em relação ao cafeeiro sem irrigação.

Com relação as formas de se fazer a irrigação, se destaca a via gotejamento ou localizada, a qual evita o molhamento total da planta, diminuindo o gasto com o controle de doenças e o consumo de água.

Finalmente, quando se fala em café, o interesse do produtor é principalmente saber se os custos serão cobertos pela produção. A irrigação fornece uma maior segurança quanto a precisão destes cálculos.

2.3 A modelagem do crescimento de plantas

No planejamento de experimentos o pesquisador está preocupado em estimar, com a maior precisão possível, o efeito dos tratamentos sobre o material a ser investigado.

O estudo das curvas de crescimento é usado largamente em biologia e agricultura, entre outras áreas. Assim, muitas funções e formas de análise foram propostas para modelar o crescimento. Os mais utilizados são os modelos lineares, entretanto, cresce cada vez mais o interesse por modelos não lineares como os sigmoidais Gompertz, Logístico e Richards, descritos por Ratkowsky (1983), entre outros. E ainda se destaca atualmente a inclusão de modelos de efeitos mistos.

Os modelos lineares foram, e ainda são, os mais utilizados, talvez, devido à maior facilidade de cálculos até então. Os modelos de regressão linear, como o polinomial, já bem conhecido e estudado e seus princípios básicos estão descritos em vários textos como Drapper & Smith (1998) e Graybill & Iyer (2004).

Com o enfoque de desenvolvimento de um modelo linear de efeito misto na estimativa de crescimento e da produção em área basal, para povoamento de eucalipto, Calegario et al. (2004) verificaram que o modelo linear de efeito misto gerou estimativas precisas tanto para o efeito fixo (idade), quanto para o efeito aleatório (unidade amostral). As estimativas individuais obtidas com o modelo

linear misto tenderam a ser mais informativas pois representaram os efeitos individuais e, as estimativas referentes aos efeitos fixos, proporcionaram robustez dos resultados em relação ao fenômeno estudado.

Os modelos não lineares são os que, pelo formato que a curva descreve, explicariam de forma adequada o crescimento de plantas, como, por exemplo, o diâmetro da copa do cafeeiro. Isso pode ser observado plotando-se a variável resposta em relação ao tempo e analisando-se até visualmente, como é o comportamento da curva característica de cada modelo.

Calegario et al. (2005) usaram o modelo logístico não linear para descrever o crescimento em altura de plantas de eucalipto justificando que a mesma tem relação direta com a produtividade.

Avanços na área de computação e o surgimento de novos softwares, tem possibilitado o uso da regressão não linear na descrição de curvas resposta.

Não existe informação a respeito da aplicação da teoria de modelos mistos na descrição do desenvolvimento do cafeeiro irrigado. Entretanto foram realizados muitos trabalhos que estudam o crescimento do cafeeiro com o uso de modelos polinomiais, entre eles Alves et al. (2000), Oliveira (2003), Santana et al. (2004), Martins et al. (2006) e Figueiredo et al. (2006).

2.4 Modelos de regressão linear

Um modelo estatístico consiste de no mínimo dos componentes: resposta, erro, parâmetros e a parte sistemática. Segundo Schabenberger & Pierce (2002) a maioria dos modelos pode ser representada pela decomposição:

$$Resposta = estrutura + erro \quad (2.1)$$

em que “Resposta” é a medida de interesse a ser mensurada, geralmente represen-

tada por “y”; estrutura é a parte sistemática que contém os parâmetros do modelo, chamada de função média; o erro é a diferença entre a função média e as observações.

De maneira simples, pode-se representar a equação básica de um modelo estatístico como:

$$y_{ki} = f(x_{0i}, x_{1i}, x_{2i}, \dots, x_{ki}, \theta_0, \theta_1, \theta_2, \dots, \theta_p) + e_{ki} \quad (2.2)$$

em que $x_{0i}, x_{1i}, x_{2i}, \dots, x_{ki}$ são as “variáveis medidas”, $\theta_0, \theta_1, \theta_2, \dots, \theta_p$ são os parâmetros representando a parte sistemática, y_{ki} é a variável resposta e e_{ki} é o erro. Assume-se que os erros tem média zero, ou seja, $E[e_{ki}] = 0$.

Em regressão múltipla o objetivo é predizer a variável dependente ou variável resposta y com base na suposição de que a mesma assume uma relação linear com as variáveis independentes ou preditoras x_1, x_2, \dots, x_k . Esta relação linear pode ser expressa para um modelo com n observações, por:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + e_i, i = 1, 2, \dots, n. \quad (2.3)$$

em que os β'_s são os parâmetros e e_i representa os fatores não controlados, ou não incluídos no modelo.

Escrevendo (2.3) para cada uma das n observações tem-se:

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \dots + \beta_k x_{1k} + e_1 \\ y_2 &= \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \dots + \beta_k x_{2k} + e_2 \\ &\vdots \\ y_n &= \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \dots + \beta_k x_{nk} + e_n \end{aligned}$$

Estas n equações podem ser escritas na forma matricial da seguinte forma:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & x_{1k} \\ 1 & x_{21} & x_{22} & x_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & x_{nk} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}$$

ou

$$y = X\beta + e. \quad (2.4)$$

em que y é um vetor de realizações de variáveis aleatórias, da dimensão $n \times 1$, X é a matriz conhecida com os valores x'_s do planejamento ou do delineamento, da dimensão $n \times (k + 1)$; β é o vetor dos parâmetros, de dimensão $(k + 1) \times 1$; e é o vetor dos erros $(k + 1) \times 1$.

De acordo com Rencher & Schaalje (2008) as pressuposições básicas para os erros e_i são de (2.3) são:

1. $E(e_i) = 0$ para $i = 1, 2, \dots, n$ ou $E(y_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik}$.
2. $var(e_i) = \sigma^2$ para $i = 1, 2, \dots, n$ ou $var(y_i) = \sigma^2$.
3. $cov(e_i, e_j) = 0$ para todo $i \neq j$ ou $cov(y_i, y_j) = 0$.

ou para o modelo (2.4) podem ser expressas como:

1. $E(e) = 0$ ou $E(y) = X\beta$
2. $cov(e) = \sigma^2 I$ ou $cov(y) = \sigma^2 I$

Nota-se que $cov(e) = \sigma^2 I$ inclui as pressuposições de que $var(e_i) = \sigma^2$ e $cov(e_i, e_j) = 0$.

2.4.1 O método dos mínimos quadrados e pressuposições para a estimação de β e σ^2

O método dos mínimos quadrados (MQ) tem por objetivo minimizar a soma dos quadrados dos desvios entre os valores observados y_s e os valores preditos \hat{y}_s com a finalidade de estimar os parâmetros $\beta_0, \beta_1, \dots, \beta_k$. Não é necessária nenhuma pressuposição a respeito de y (Schabenberger & Pierce, 2002; Faraway, 2004; Rencher & Schaalje, 2008).

Com base nas equações

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2.5)$$

$$= \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \dots - \hat{\beta}_k x_{ik})^2 \quad (2.6)$$

segundo Drapper & Smith (1998) e Rencher & Schaalje (2008), entre outros, para encontrar $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$ que minimiza (2.6) deriva-se $\sum_i \hat{e}_i^2$ com relação a cada $\hat{\beta}_j$ (com $j = 1, 2, \dots, k$) igualando o resultado a zero, para produzir $(k + 1)$ equações que podem ser resolvidas simultaneamente. Pode-se representar (2.6) como

$$\hat{e}'\hat{e} = \sum_{i=1}^n (y_i - x'_i \hat{\beta})^2 = (y - X\hat{\beta})'(y - X\hat{\beta}) \quad (2.7)$$

em que $x'_i = (1, x'_{i1}, x'_{i2}, \dots, x'_{ik})$ são as n linhas de X . Resolvendo a multiplicação de 2.7 tem-se

$$\hat{e}'\hat{e} = y'y - 2y'X\hat{\beta} + \hat{\beta}'X'X\hat{\beta} \quad (2.8)$$

Derivando-se 2.8 com relação a $\hat{\beta}$ e igualando-se a zero obtém-se

$$\frac{\partial \hat{e}'\hat{e}}{\partial \hat{\beta}} = 0 - 2X'y + 2X'X\hat{\beta} = 0 \quad (2.9)$$

o que nos leva ao sistema de equações normal

$$X'X\hat{\beta} = X'y \quad (2.10)$$

Como a matriz X tem posto coluna completo e igual a $r(x) = k + 1$, a matriz $X'X$ é positiva definida, de posto completo; logo existe a matriz inversa de $X'X$. Assim, a solução única do sistema de (2.10) é

$$\hat{\beta} = (X'X)^{-1}X'y \quad (2.11)$$

O vetor $\hat{\beta}$, assim determinado, minimiza a soma de quadrados dos erros em (2.6), pois a segunda derivada de (2.9) em relação a $\hat{\beta}$ é

$$\frac{\partial^2 \hat{e}'\hat{e}}{\partial \hat{\beta}^2} = 2X'X \quad (2.12)$$

que é uma função positiva, implicando que o modelo de (2.6) tem mínimo. Assim, $\hat{\beta}$ é chamado de estimador de mínimos quadrados (Rencher & Schaalje, 2008). Se a matriz X for singular, então uma alternativa é a utilização da inversa generalizada (Schabenberguer & Pierce, 2002).

O estimador dos mínimos quadrados tem algumas propriedades importantes (Weisberg, 2005; Rencher & Schaalje, 2008), como:

1. Se $E(y) = X\beta$, então $\hat{\beta}$ é um estimador não viesado para β . Aplicando-se

o operador da esperança matemática, tem-se

$$\begin{aligned}
 E(\hat{\beta}) &= E[(X'X)^{-1}X'y] \\
 &= (X'X)^{-1}X'E(y) \\
 &= (X'X)^{-1}X'X\beta \\
 &= I\beta \\
 &= \beta
 \end{aligned}$$

2. Se $cov(y) = \sigma^2 I$, a matriz de covariância para $\hat{\beta}$ é dado por $\sigma^2(X'X)^{-1}$.
 Por definição de covariância, como em Rencher & Schaalje (2008), tem-se

$$\begin{aligned}
 cov(\hat{\beta}) &= cov[(X'X)^{-1}X'y] \\
 &= (X'X)^{-1}X'cov(y)[(X'X)^{-1}X']' \\
 &= (X'X)^{-1}X'(\sigma^2 I)X(X'X)^{-1} \\
 &= \sigma^2(X'X)^{-1}X'X(X'X)^{-1} \\
 &= \sigma^2(X'X)^{-1}
 \end{aligned}$$

Se $E(y) = X\beta$ e $cov(y) = \sigma^2 I$, os estimadores de mínimos quadrados (MQ) $\hat{\beta}_j$, $j = 0, 1, \dots, k$ tem variância mínima entre todos os estimadores lineares não viesados (Rencher & Schaalje, 2008).

O método dos MQ não fornece uma função dos valores y e x que podem ser minimizados para obtenção de σ^2 . Entretanto, pode-se obter um estimador não viesado para σ^2 com base no estimador $\hat{\beta}$ dos MQ. Assumindo-se que σ^2 é a mesma para cada y_i , $i = 1, 2, \dots, n$ e que σ^2 é definida por $\sigma^2 = E[y_i - E(y_i)]^2$, tem-se que:

$$E(y_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} = x'_i \beta$$

em que x'_i são as n linha de X . Então σ^2 será

$$\sigma^2 = E[y_i - x'_i \beta]^2$$

Um estimador não viesado para σ^2 é dado pela expressão

$$\hat{\sigma}^2 = \frac{1}{n - k - 1} \sum_{i=1}^n (y_i - x'_i \hat{\beta})^2 \quad (2.13)$$

em que n é o tamanho da amostra e k é o número de x_s . A expressão do estimador de 2.13 pode ser escrita na forma matricial como:

$$\hat{\sigma}^2 = \frac{1}{n - k - 1} (y - X\hat{\beta})'(y - X\hat{\beta}) \quad (2.14)$$

$$= \frac{y'y - \hat{\beta}'X'y}{n - k - 1} = \frac{SSE}{n - k - 1} \quad (2.15)$$

em que $SSE = (y - X\hat{\beta})'(y - X\hat{\beta}) = y'y - \hat{\beta}'X'y$ é a soma dos quadrados dos erros (Rencher & Schaalje, 2008).

O método dos mínimos quadrados se baseia no princípio geométrico conforme ilustrado na Figura 1, que representa o conhecido Teorema de Pitágoras. Os parâmetros β são estimados por valores de $\hat{\beta}$ que minimizam a soma de quadrados dos desvios $(y - X\hat{\beta})'(y - X\hat{\beta})$. Na sua estimação não se exige que os termos erro tenham distribuição normal.

O método dos mínimos quadrados ordinários (MQO) de acordo com Schabenberger & Pierce (2002), leva ao melhor estimador linear não viesado e com variância mínima se $e \sim N(0, \sigma^2 I)$



FIGURA 1 Representação geométrica do método dos mínimos quadrados mostrando o vetor de observações y , o espaço gerado ou explicado pelo delineamento $X\hat{\beta}$ e o vetor dos erros $y - X\hat{\beta}$.

Na utilização de algum modelo de regressão quer seja com aplicação dos MQO, o diagnóstico de problemas com o modelo é um passo importante para que a inferência seja feita de forma correta. Com esse enfoque, na utilização do modelo de regressão linear (Gaussiano), segundo Gujarati (2006), as seguintes premissas devem ser consideradas:

- a) O modelo é linear nos parâmetros.
- b) Os valores da variável X são fixos em amostras repetidas, ou seja, considera-se que os valores assumidos pelo regressor X foram fixados em amostras repetidas.
- c) O valor médio dos erros é nulo, $E(u_i | X_i) = 0$, indicando que os fatores não incluídos explicitamente no modelo não afetam o valor médio de y_i .
- d) Exige-se homoscedasticidade dos erros u_i de modo que dado o valor de X_i ,

a variância de u_i deve ser a mesma para todas as observações:

$$\begin{aligned}Var(u_i|X_i) &= E[u_i - E(u_i|X_i)]^2 \\ &= E(u_i^2 | X_i) \\ &= \sigma^2\end{aligned}$$

e) Ausência de correlação entre os erros; para dois diferentes erros u_i e u_j , a covariância entre eles é:

$$\begin{aligned}Cov(u_i u_j | X_i X_j) &= E\{[u_i - E(u_i|X_i)]\{[u_j - E(u_j|X_j)]\} \\ &= E(u_i|X_i)(u_j|X_j) \\ &= 0\end{aligned}$$

f) Ausência de covariância entre o erro u_i e a variável regressora X_i ,

$$\begin{aligned}Cov(u_i X_i) &= E[u_i - E(u_i)][X_i - E(X_i)] \\ &= E[u_i(X_i - E(X_i))] \\ &= E[u_i(X_i) - E(u_i)E(X_i)] \\ &= E[u_i(X_i)] \\ &= 0\end{aligned}$$

g) O número de observações deve ser maior que o número de variáveis explanatórias ou regressoras.

h) Os valores de X de uma dada amostra não devem ser os mesmos.

i) A especificação do modelo deve ser feita de forma correta. Não deve haver

viés ou erro de especificação no modelo empregado na análise.

- j) Ausência de multicolinearidade perfeita entre as variáveis regressoras, de modo que não deve haver relações lineares perfeitas entre as variáveis explanatórias.

Essas exigências, nem sempre são facilmente diagnosticadas, havendo-se a necessidade de se lançar mão de alguns tipos de análise auxiliares. A seguir são apresentados alguns outros métodos de estimação e diagnóstico em regressão.

2.4.2 O método da máxima verossimilhança (ML)

Até aqui não fez-se nenhuma pressuposição de que a variável aleatória y_1, y_2, \dots, y_n tenha distribuição normal. A partir daqui vamos assumir que

$$y \sim N(X\beta, \sigma^2 I) \text{ ou } e \sim N(0, \sigma^2 I)$$

Sob normalidade, $\sigma_{ij} = 0$ implica que as variáveis y (ou e) são independentes e não correlacionadas (Rencher & Schaalje, 2008).

Assumindo-se normalidade pode-se obter os estimadores pelo método da máxima verossimilhança. A função de verossimilhança é a função densidade conjunta das observações y_i $i = 1, 2, \dots, n$, representada por $L(\beta, \sigma^2)$. Assim, para o modelo $y = X\beta + e$, a função de verossimilhança tem por finalidade encontrar valores para β e σ^2 que maximiza $L(\beta, \sigma^2)$.

Para uma amostra de tamanho n e considerando a independência entre as observações y_i , tem-se, segundo Jorgensen (1993) e Rencher & Schaalje (2008)

$$L(\beta, \sigma^2) = \prod_{i=1}^n f(y_i; x_i' \beta, \sigma^2) \quad (2.16)$$

$$L(\beta, \sigma^2) = f(y; \beta, \sigma^2) \quad (2.17)$$

$$= \frac{1}{(2\pi)^{n/2} |\sigma^2 I|^{1/2}} e^{-(y-X\beta)'(\sigma^2 I)^{-1}(y-X\beta)/2} \quad (2.18)$$

$$= \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-(y-X\beta)'(y-X\beta)/2\sigma^2} \quad (2.19)$$

Aplicando logaritmo na função de verossimilhança de (2.19), obtém-se

$$\ln L(\beta, \sigma^2) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} (y - X\beta)'(y - X\beta) \quad (2.20)$$

a qual é mais adequada para tratamento algébrico.

Tomando-se as derivadas parciais de 2.20 com relação a β e σ^2 e igualando-se os resultados a zero obtém-se os estimadores de máxima verossimilhança de β e σ^2 dados por

$$\hat{\beta} = (X'X)^{-1} X'y \quad (2.21)$$

$$\hat{\sigma}^2 = \frac{1}{n} (y - X\hat{\beta})'(y - X\hat{\beta}) \quad (2.22)$$

Os estimadores de máxima verossimilhança de $\hat{\beta}$ em (2.21) são os mesmos estimadores de $\hat{\beta}$ dos mínimos quadrados apresentados em (2.11). O estimador $\hat{\sigma}^2$ em (2.22) é viesado, pois seu denominador é n ao invés de $n - k - 1$. Frequentemente, utiliza-se o estimador não viesado de (2.15).

2.4.3 Método da máxima verossimilhança restrita (REML)

O método da máxima verossimilhança restrita foi introduzido como uma técnica para estimar componentes da variância em modelo linear de efeitos aleatórios ou de efeitos mistos (Ravishanker, 2001). Foi utilizado inicialmente por Patterson & Thompson (1971) no estudo da recuperação da informação intrabloco para delineamentos em blocos incompletos. É uma variante do processo de estimação da máxima verossimilhança para modelos mistos.

Enquanto o método de máxima verossimilhança assume que não existe erro na estimativa da média ao estimar a variância, subestimando este termo, o método da máxima verossimilhança restrita produz estimativas não viesadas da variância, dado que, remove o viés existente na estimação da média (Diggle et. al. 1994).

A obtenção dos estimadores REML se dá maximizando a parte da função de verossimilhança que em termos do modelo $y = X\beta + Z\gamma + e$ é invariante para $X\beta$, em que $X\beta$ é a parte fixa do modelo e $Z\gamma$ é a parte aleatória.

Neste método, as observações são divididas em uma parte referente a efeitos fixos, outra a efeitos aleatórios, sendo a função densidade de probabilidade das observações dada pela soma das funções densidade de probabilidade de cada parte. Maximizando-se a função densidade de probabilidade da parte referente aos efeitos aleatórios, em relação aos componentes da variância, elimina-se o viés que resulta da perda de graus de liberdade na estimação dos efeitos fixos do modelo.

Se os dados são balanceados, as equações do REML são iguais aos estimadores ANOVA, não viesados e com variância mínima. O estimador REML considera os graus de liberdade envolvidos nas estimativas dos efeitos fixos, o que não acontece com o ML. Se os dados são desbalanceados, segundo Searle (1987), os estimadores ML e REML serão viesados.

Segundo Venables & Ripley (2002), o método de estimação *default* do soft-

ware *R* (R Development Core Team, 2009) para modelos mistos é o REML, no qual os parâmetros na estrutura de variância são estimados pela maximização da verossimilhança marginal dos resíduos obtidos no ajuste pelo método dos mínimos quadrados e a partir daí os efeitos fixos são estimados pela máxima verossimilhança assumindo que a estrutura de variância é conhecida.

Finalmente, é importante lembrar que os estimadores dos componentes da variância para ML e REML não são formas explícitas, ou seja, o estimador de cada componente só pode ser encontrado por métodos numéricos por estar em função de outros componentes.

2.4.4 Análise dos resíduos

O vetor de erros não pode ser estimado sem o conhecimento dos parâmetros β_s e, é dado por

$$\hat{e} = y - X\hat{\beta} = y - \hat{y} \quad (2.23)$$

Os resíduos dados em 2.23 são usados em muitos procedimentos para verificar a validade ou a adequabilidade do modelo, de acordo com seus pressupostos.

Segundo Schabenberger & Pierce (2002), a verificação dos resíduos brutos da equação 2.23 não são a melhor forma de análise, dado que, os mesmos podem ser correlacionados e heteroscedásticos. Os resíduos studentizados são adequadamente padronizados e minimizam o problema da heterogeneidade (Schabenberger & Pierce, 2002).

Desde que os resíduos ou erros do modelo são variáveis aleatórias não observáveis parece natural, utilizar os resíduos obtidos com relação aos valores ajustados. Assim, os vetores de erros $\hat{e} = y - \hat{y} = y - X\hat{\beta}$, em que $\hat{\beta}$ é um estimador não viesado de β , tem a propriedade de que na média seja nula, logo $E[\hat{e}] = E[e] = 0$; mas os resíduos estimados podem ser correlacionados e hete-

roscedásticos. Os resíduos podem ser definidos pela expressão $\hat{e} = y - \hat{y} = y - Hy$ em que $H = X(X'X)^{-1}X'$ é a matriz chamada “hat” dado que ela produz os valores de \hat{y} quando multiplicada por y e, representa o projetor ortogonal das observações sobre o espaço paramétrico da matriz de delineamento X . A matriz H possui algumas propriedades importantes, entre elas, a H é simétrica ($H' = H$) e idempotente ($HH = H$), o que implica também que $(I - H)(I - H) = (I - H)$ (Schabenberger & Pierce, 2002).

A variância dos erros é dada por:

$$\begin{aligned} Var[\hat{e}] &= Var[y - Hy] = Var[y(I - H)] \\ &= \sigma^2 I(I - H)(I - H) = \sigma^2(I - H) \end{aligned}$$

H não é uma matriz diagonal e os elementos na sua diagonal nem sempre são iguais.

Devido a sua natureza heteroscedástica que pode haver entre os erros, não recomenda-se usar os resíduos brutos para a verificação das pressuposições do modelo. Os resíduos brutos devem ser primeiro apropriadamente transformados. O i -ésimo valor da diagonal de H é chamado leverage do i -ésimo dado. Considerando que h_{ii} é o elemento da diagonal da matriz H , a variância de um erro é

$$Var[\hat{e}_i] = \sigma^2(1 - h_{ii}) \quad (2.24)$$

E, os resíduos padronizados com média zero e variância 1 são obtidos a por

$$\hat{e}_i^* = \frac{\hat{e}_i}{(\sigma\sqrt{1 - h_{ii}})} \quad (2.25)$$

Como σ^2 é desconhecida, σ é substituído por $\hat{\sigma}$, raiz quadrada do quadrado médio do erro do modelo. O resíduo

$$r_i = \frac{\hat{e}_i}{\hat{\sigma}\sqrt{1 - h_{ii}}} \quad (2.26)$$

é chamado de resíduo padronizado e é mais apropriado para as medidas de diagnóstico.

Para verificar os pressupostos de normalidade, pode-se utilizar métodos gráficos como o que relaciona os quantis teóricos com os quantis esperados ou testes de normalidade tradicionais como o de Shapiro Wilk e Kolmogorov-Smirnov cujas aplicações com mais detalhes podem ser vistas em Shapiro & Wilk (1965) e Stephens (1974).

2.5 Modelo de regressão não linear

Segundo Ratkowsky (1983), Draper & Smith (1998), Bates & Watts (1988), entre outros, a definição para que um modelo seja designado não linear é de que pelo menos uma das derivadas parciais da variável dependente com relação a algum parâmetro presente no modelo, dependa de algum parâmetro. Os procedimentos para análise de modelos não lineares são descritos por Gallant (1975), Bates & Watts (1978), Ratkowski (1983), Seber & Wild (1989) e Schabenberguer & Pierce (2002).

As vantagens no uso da modelagem não linear superam as desvantagens conforme salienta Schabenberguer & Pierce (2002) que destacam os seguintes aspectos:

- a) São modelos mais parcimoniosos que os lineares. Para construir uma curva com inflexão utilizando-se um modelo polinomial, é necessário uma função

cúbica com quatro parâmetros na função média. Já, uma função não linear pode apresentar o ponto de inflexão com menos parâmetros.

- b) Muitas medidas evoluem para um nível máximo (platô) ou mínimo, assim permanecendo. Esse fenômeno pode ser difícil de ser representado por um modelo linear. Muitos modelos não lineares tem sido estudados e construídos para exibir o comportamento de limite assintótico, a inflexão e simetria da curva.
- c) Muitos modelos não lineares derivam de princípios biológicos, físicos ou químicos. Por exemplo, se y é o tamanho de uma planta no tempo t e a é o seu tamanho máximo, assumindo que a taxa de crescimento $\partial y/\partial t$ é proporcional ao crescimento restante $(a - y)$, isso leva a equação diferencial $\partial y/\partial t = (a - y)$. Integrando-se a mesma, chega-se ao modelo linear de crescimento com três parâmetros $y(t) = \alpha + (\gamma - \alpha) \exp(-\beta t)$. O parâmetro γ representa o tamanho inicial $y(0)$.
- d) Os parâmetros dos modelos não lineares tipicamente significam quantidades e tem interpretação direta com relação ao fenômeno estudado. No exemplo do ítem anterior, α é o tamanho final, γ é o tamanho inicial e β rege a taxa de mudança que determina a rapidez em que a planta cresce de γ para α .

Os modelos não lineares podem ser modelos de regressão intrinsecamente lineares e não lineares (Draper & Smith, 1998). Alguns modelos são não lineares nos parâmetros, mas por transformações, podem se tornar modelos de regressão linear nos parâmetros (Gujarati, 2006). Estes são ditos intrinsecamente lineares.

Um modelo não-linear pode ser linearizado para a obtenção das estimativas dos parâmetros (Bates & Watts, 1988). Entretanto, existe um inconveniente de que além do parâmetro perder sua interpretação verdadeira, pode-se alterar a estrutura

e distribuição do erro e assim, as suposições de normalidade, independência e homogeneidade da variância, podem não ser satisfeitas.

Um modelo de regressão que não é linear e nem intrinsecamente linear nos parâmetros é chamado de intrinsecamente não-linear nos parâmetros (Draper & Smith, 1998).

Alguns tipos de funções de regressões não lineares têm sido utilizadas para descrever variáveis físicas e sistemas biológicos, e além do mais, sendo que uma das vantagens é que em tais modelos os parâmetros têm interpretações práticas (Graybill & Iyer, 2004).

Um crescente número de pesquisadores na área de ciências biológicas acredita que as relações entre as variáveis biológicas são melhores descritas por funções não lineares (Schabenberger & Pierce, 2002). Assim, é interessante levar em consideração esta possibilidade de análise.

O uso da análise de regressão não linear requer também algumas pressuposições e alguns cuidados, sendo popularizada com o avanço da informática fornecendo muitas vezes resultados mais aplicáveis do que a própria regressão linear. A regressão não linear é preferível porque os coeficientes da curva ajustada tem um significado biológico (Calegario et al., 2005).

Poucos são os trabalhos que utilizaram modelos não lineares para estudos com o cafeeiro. Dentre estes, pode-se citar Silva & Volpe (2005) que estudaram o crescimento do fruto do cafeeiro utilizando o modelo logístico de três parâmetros.

2.6 O Modelo linear misto

Muitos modelos estatísticos podem ser expressos como modelos lineares que incorporam efeitos fixos e aleatórios. Os *efeitos fixos* são associados a população ou a certos níveis de fatores experimentais que podem ser repetidos. Os *efeitos*

aleatórios associados as unidades experimentais e são retirados aleatoriamente da população. Um modelo que envolve efeitos fixos e efeitos aleatórios, além do erro experimental e da constante μ é chamado modelo misto (Pinheiro & Bates, 2000).

Uma das justificativas para se utilizar modelo misto, é que o mesmo possibilita a predição de efeitos aleatórios na presença de efeitos fixos.

O modelo linear misto suporta a análise de um grande número de problemas estatísticos (é robusto) e pode ser representado matricialmente de forma geral, segundo Pinheiro & Bates (2000), por:

$$y = X\tau + Z\gamma + e \quad (2.27)$$

em que y é um vetor de variáveis observáveis de dimensões $n \times 1$, X é a matriz de delineamento dos efeitos fixos de dimensão $n \times p$ e Z é uma matriz $n \times q$ de incidência de efeitos aleatórios (covariáveis), τ é o vetor dos parâmetros de efeitos fixos de dimensão $p \times 1$, γ é o vetor de efeitos aleatórios e e é o vetor n -dimensional de erros aleatórios.

Considerando-se que os parâmetros de efeitos aleatórios e os erros aleatórios seguem uma distribuição normal, são não correlacionados e com média zero, com matrizes de variâncias e covariâncias dadas por

$$Var(\gamma) = E(\gamma\gamma') = D$$

e

$$Var(e) = E(ee') = R$$

Tem-se que

$$V = Var(y) = Var(X\tau + Z\gamma + e) = ZDZ' + R$$

Assumindo-se que V é uma matriz não singular e

$$E(y) = E(X\tau + Z\gamma + e) = X\tau,$$

então,

$$y \sim N(X\tau, ZDZ' + R)$$

A equação dos modelos mistos pode ser encontrada minimizando-se a soma dos quadrados dos resíduos (mínimos quadrados) ou maximizando-se a função densidade de probabilidade conjunta de y e γ . Assim, maximizando-se a função densidade de probabilidade conjunta, as equações de modelos misto que permitem soluções para os parâmetros dos efeitos fixos (τ^0) e predições para efeitos aleatórios ($\hat{\gamma}$), de acordo com Pinheiro & Bates (2000) são dadas por:

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ -X'R^{-1}X & X'R^{-1}Z\hat{\gamma} + D^{-1} \end{bmatrix} \begin{bmatrix} \beta^0 \\ \hat{\gamma} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix}$$

A estimação dos componentes da variância e a solução das equações de modelos mistos (MME) depende da matriz de variâncias e covariâncias V cujos compo-

nentes não são conhecidos, embora a estrutura seja. Assim, é necessário substituí-los por suas estimativas.

O método da análise da variância se presta bem para obtenção de estimativas nos casos em que se tem dados balanceados, produzindo equações lineares dos componentes da variância e sendo suas soluções tomadas como estimadores dos componentes da variância. Para dados desbalanceados, as estimativas são baseadas na máxima verossimilhança, como por exemplo, com a utilização dos métodos de ML e REML (Pinheiro & Bates, 2000).

Os modelos mistos permitem a observação da variação dos coeficientes de regressão entre os indivíduos. Esses modelos têm um componente que descreve a variação entre indivíduos. O modelo de efeitos aleatórios permite tanto descrever a tendência temporal, lembrando que nem sempre a variável independente é o tempo, levando-se em conta a correlação que existe entre medidas sucessivas como, também, estimar a variação medida ao longo do tempo. As medidas dos indivíduos não necessitam ser igualmente espaçadas e balanceadas e as análises podem ser conduzidas com os dados de indivíduos que apresentam ausência de informação em algum momento. O uso de modelos de efeitos aleatórios é também adequado para dados em que a variabilidade entre os indivíduos é maior do que a variabilidade dentro do indivíduo (Fausto et al., 2009). O modelo assume que a variação na resposta para todos os indivíduos seguem um mesmo modelo ou tendência, porém os indivíduos podem se comportar de formas diferentes. Assim, cada indivíduo terá a sua própria curva de crescimento (Fausto et al., 2009).

2.7 O modelo não linear misto

O modelo não linear misto, conforme proposto por Lindstron & Bates (1990) pode ser visto como um modelo hierárquico. Para um único nível as j_s observações

nos i_s grupos, o modelo é apresentado como

$$y_{ij} = f(\phi_{ij}, v_{ij}) + e_{ij} \quad i = 1, 2, \dots, M \text{ e } j = 1, 2, \dots, n_i \quad (2.28)$$

em que y_{ij} é a variável resposta para o i -ésimo grupo e a j -ésima observação, M é o número de grupos, n_i é o número de observações no i -ésimo grupo, f é uma função diferenciável de um vetor de parâmetros específico ϕ_{ij} , v_{ij} é o vetor de covariáveis e e_{ij} é o termo erro aleatório entre os grupos, normalmente distribuídos.

A função f é não linear em pelo menos um componente do vetor de parâmetros ϕ_{ij} de um grupo específico, o qual é modelado conforme sugerido por Pinheiro & Bates (2000) por:

$$\phi_{ij} = A_{ij}\beta + B_{ij}b_i, \quad b_i \sim N(0, \Psi) \quad (2.29)$$

em que β é um vetor p -dimensional de efeitos fixos e b_i é um q -dimensional vetor de efeitos aleatórios associado ao i -ésimo grupo (não varia com j) com matriz de variância e covariância Ψ . As matrizes A_{ij} e B_{ij} tem as dimensões apropriadas e dependem do grupo e dos valores de algumas covariáveis na j -ésima observação.

Devido a f poder ser qualquer função não linear de ϕ_{ij} , a representação de um grupo específico de coeficientes pode ser escolhida. Então A_{ij} e B_{ij} são as vezes simplesmente matrizes de incidência (Pinheiro & Bates, 2000). A forma de estimação dos parâmetros pode ser vista em Schabenberger & Pierce (2002).

2.8 Análise dos resíduos: o qq-plot

O qq-plot ou gráfico quantil-quantil é um gráfico que testa a conformidade entre a distribuição empírica e uma dada distribuição teórica.

Esta representação gráfica surgiu em meados dos anos 1960 e a partir daí tornou-se uma metodologia muito utilizada dada a riqueza de suas propriedades e facilidade de interpretação (Dodge, 2008).

A construção dos qq-plots foi feita colocando-se os resíduos em ordem crescente e associando cada dado com o quantil da distribuição normal padrão. Como regra de decisão observa-se se a representação gráfica entre os quantis da variável aparecem alinhados com os quantis da distribuição. Se isso ocorre é porque os dados são normalmente distribuídos.

Outros métodos para visualização da análise dos resíduos aqui utilizados foram o histograma com a plotagem da densidade e os testes de Shapiro-Wilk e Kolmogorov - Smirnov conforme sugestão de Draper & Smith (1998).

2.9 Análise do ajuste da curva pelo R^2

Assim como no ajuste das curvas tem-se a preocupação quanto a verificação dos pressupostos de análise da variância, deve-se estar atento, também, à qualidade deste ajuste, de modo a ver o quanto “bem” uma linha de regressão está ajustada aos dados (Gujarati, 2006). Para duas variáveis o R^2 é uma medida que mostra o quanto uma linha de regressão se ajusta aos dados.

Gujarati (2006) destaca duas propriedades do R^2 como sendo:

1. Um número não negativo (No modelo linear simples é a razão entre duas somas de quadrados);
2. Seus valores limites estão entre $0 \leq R^2 \leq 1$. Se o valor estiver mais próximo de 1, melhor é o seu ajuste.

Uma das formas de se expressar o nível de previsão do ajuste é por meio do coeficiente de determinação R^2 que é dado pela razão entre a soma de quadrados da regressão e a soma de quadrados total. O R^2 foi utilizado como medida de ajuste das curvas.

3 MATERIAL E MÉTODOS

3.1 Realização do experimento, delineamento experimental e tratamentos

Os dados para a realização deste trabalho foram provenientes de um experimento realizado na área experimental do Departamento de Engenharia da Universidade Federal de Lavras. A cultivar em estudo foi a Acaiá MG -1474 implantada em março de 1997 no espaçamento 3,00x0,60m, sendo recepada em outubro de 2004 e conduzida sob irrigação. O delineamento experimental utilizado foi o de blocos casualizados, com quatro repetições e cinco tratamentos. Cada parcela foi constituída por 3 linhas de plantio com 10 plantas (indivíduos) em cada uma; como área útil foi considerada cinco plantas úteis. As plantas dessas linhas foram utilizadas para avaliação do crescimento do diâmetro da copa, o qual foi medido com barra graduada na altura do terço médio da planta e perpendicular a linha de café. Os tratamentos utilizados constaram de manejos de irrigação:

- SI = Sem irrigação (testemunha);
- A-75 = Irrigação o ano todo sempre que a planta consumir 75% da água disponível na camada de 0-40 cm;
- A-25 = Irrigação o ano todo sempre que a planta consumir 25% da água disponível na camada de 0-40 cm;
- A-7525 = Irrigação o ano todo, porém: nos meses de jan/fev/mar/jul/out/nov/dez só foi irrigado quando a água disponível de armazenamento na camada de 0-40 cm atingisse valores inferiores a 75%; nos demais meses, ou seja, abr/mai/jun/ago/set só foi irrigado se a água disponível de armazenamento na camada de 0-40 cm atingisse valores inferiores a 25% ;

- PS-25 = Irrigação somente nos meses abr/maio /jun/ago/set sempre que a planta consumir 25% da água disponível na camada de 0-40 cm.

O tipo de irrigação utilizado foi o localizado por gotejamento. As características vegetativas foram avaliadas no período de Julho/2005 a Junho/2008.

As medidas foram irregularmente espaçadas no tempo e os dados desbalanceados. Foram feitas 20 medições irregularmente espaçadas ao longo do mesmo.

3.2 Os modelos testados

3.2.1 Modelo linear polinomial

O modelo para o experimento em blocos ao acaso pode ser representado pela expressão:

$$y_{ij} = m + t_i + b_j + e_{ij} \quad (3.1)$$

em que: y_{ij} são os valores observados; m é uma constante inerente a todas as observações (média geral); t_i é o efeito do i -ésimo tratamento ($i = 1, 2, \dots, I$); b_j é o efeito do j -ésimo bloco ($j = 1, 2, \dots, J$); e_{ij} é o erro aleatório.

Para cada tratamento realizou-se a plotagem dos dados observados em função do tempo e com base em uma visualização prévia, decidiu-se o ajuste utilizando uma polinomial dado por uma função de terceiro grau para cada tratamento, ou seja:

$$y_{ijk} = \mu_i + b_j + \beta_{10} + \beta_{11}t_k + \beta_{12}t_k^2 + \beta_{13}t_k^3 + e_{ijk} \quad (3.2)$$

em que y_{ijk} é a resposta obtida no i -ésimo tratamento para a j -ésima repetição no k -ésimo tempo; μ_i é uma constante inerente a todas as observações; b_j é o j -ésimo bloco ($j = 1, 2, 3, 4$); $\beta_0, \beta_1, \beta_2$ e β_3 são os parâmetros a serem estimados, t_k é a variável tempo ($k = 1, 2, \dots, 20$) considerando-se a primeira observação como

sendo no tempo 1 ($t = 1$) e as demais em número de dias após a primeira avaliação ($t_2 = 31$, medida feita 30 dias após a primeira), e e_{ijk} é o erro aleatório associado à resposta obtida para o i -ésimo tratamento na j -ésima repetição e k -ésimo tempo. A pressuposição básica é de que o $e_{ijk} \sim N(0, \sigma^2)$.

As análises foram realizadas considerando-se os blocos como controle da heterogeneidade local e sendo verificados os pressupostos de normalidade utilizando-se o teste de Shapiro-Wilk e homogeneidade da variância dos resíduos padronizados e plotagem dos resultados, conforme sugestão de Drapper & Smith (1998) e Schabenberger & Pierce (2002).

Utilizou-se de instrumento computacional a função *lm* na versão 2.9.2 do programa R (R development core team, 2008), conforme rotina própria apresentada no Anexo A.

3.2.2 Modelo linear com efeito misto

Para o modelo de (3.2.1) considerou-se que os parâmetros referentes a tratamentos (t_i) e blocos (b_j) são de efeitos fixos e, o erro experimental (e_{ijk}) e a variação dos indivíduos dentro da parcela, como de efeito aleatórios, conforme o modelo

$$y_{ijkl} = \mu_i + b_j + \beta_{10} + \beta_{11}t_k + \beta_{12}t_k^2 + \beta_{13}t_k^3 + e_{ijkl} \quad (3.3)$$

em que y_{ijkl} é a resposta obtida no i -ésimo tratamento ($i = 1, 2, \dots, 5$) para a j -ésima repetição ($j = 1, 2, 3, 4$) no k -ésimo tempo para o l -ésimo indivíduo; μ_i é uma constante inerente a todas as observações; b_j é o j -ésimo bloco ($j = 1, 2, 3, 4$); $\beta_0, \beta_1, \beta_2$ e β_3 são os parâmetros a serem estimados, t_k é a variável tempo ($k = 1, 2, \dots, 20$) considerando-se a primeira observação como sendo no tempo 1 ($t = 1$) e as demais em número de dias após a primeira avaliação ($t_2 = 31$, medida feita

30 dias após a primeira), e e_{ijkl} é o erro aleatório associado à resposta obtida para o i -ésimo tratamento na j -ésima repetição, k -ésimo tempo e l -ésimo indivíduo. A pressuposição básica é de que o $e_{ijkl} \sim N(0, \sigma^2)$.

Após o ajuste do modelo polinomial de terceiro grau para cada tratamento foram verificados os pressupostos da análise da variância, conforme Pinheiro & Bates (2000).

Utilizou-se o pacote a função *lmer* do pacote *lme4* do R e a rotina apresentada no Anexo B.

3.2.3 Modelo não linear

Os modelos não lineares utilizados foram o Logístico e o de Gompertz (Ratkowsky, 1983), os quais, pelo formato das curvas que descrevem, são muito utilizados na descrição do crescimento de plantas, adaptando-se muito bem para esse fim. Todos os efeitos de tratamentos e blocos foram considerados fixos.

O modelo logístico foi representado por

$$y_k = \frac{A}{1 + e^{\left(\frac{I - t_k}{S}\right)}} + u_k$$

em que y_k é a resposta obtida no k -ésimo tempo, A representa a assíntota de crescimento do diâmetro da copa, tendendo para o diâmetro máximo quando o tempo tende para o infinito, I representa o ponto de inflexão da curva, S é o parâmetro de escala e u_k é o termo erro aleatório.

O modelo de Gompertz foi representado por

$$y_k = Ae^{(-BC^{t_k})} + u_k$$

em que y_k é a resposta obtida no k -ésimo tempo, A representa a assíntota de cresci-

mento do diâmetro da copa quando o tempo tende para o infinito, B é o parâmetro numérico relacionado ao valor da função quando o tempo tende a zero $t = 0$, C é o parâmetro numérico para a escala no eixo do tempo t e u_k é o termo erro aleatório. Utilizou-se a rotina apropriada desenvolvida no programa R apresentada no Anexo C.

Outras parametrizações para o modelo logístico e de Gompertz podem ser encontradas em Ratkowsky (1983) e Seber & Wild (1989).

3.2.4 Modelo não linear com efeito misto

Os modelos não lineares utilizados foram o Logístico e o de Gompertz por serem os mais conhecidos e usados para descrever o crescimento de plantas, animais e de tecidos. Os efeitos de tratamentos e blocos foram considerados fixos e, os efeitos do erro experimental e de plantas dentro de cada parcela como aleatórios.

O modelo logístico (Ratkowsky, 1983; Seber & Wild, 1989) foi representado por

$$y_k = \frac{A}{1 + e^{\left(\frac{I - t_k}{S}\right)}} + u_k$$

em que y_k é a resposta obtida no k -ésimo tempo, A representa a assíntota de crescimento do diâmetro da copa, tendendo para o diâmetro máximo quando o tempo tende para o infinito, I representa o ponto de inflexão da curva, S é o parâmetro de escala e u_k é o termo erro aleatório.

O modelo de Gompertz (Ratkowsky, 1983; Seber & Wild, 1989) foi representado por

$$y_k = Ae^{(-BC^{t_k})} + u_k$$

em que y_k é a resposta obtida no k -ésimo tempo, A representa a assíntota de crescimento do diâmetro da copa quando o tempo tende para o infinito, B é o

parâmetro numérico relacionado ao valor da função quando o tempo tende a zero $t = 0$, C é o parâmetro numérico para a escala no eixo do tempo t e u_k é o termo erro aleatório.

As pressuposições básicas para ambos os modelos são de que os efeitos aleatórios são normalmente distribuídos para os diferentes indivíduos e os erros do modelo são normais, independentes, identicamente distribuídos e independentes dos efeitos aleatórios. Para melhoria das estimativas no modelo logístico foi realizada a modelagem da heterogeneidade da variância fazendo-se:

$$Var(e) = \sigma^2 |Tempo|^{2\delta},$$

em que δ é o parâmetro que modela casos nos quais a variância aumenta ou diminui de acordo com a ação da covariável, ou seja, é o coeficiente da função de modelagem da variância. A modelagem da heterogeneidade da variância foi realizada com uso da função *varPower* do pacote *nlme* do R e os modelos sem e com modelagem foram comparados por meio do teste da razão de verossimilhança.

A estimação dos parâmetros foi feita por meio do método da máxima verossimilhança restrita (REML), utilizando-se a função *nlme* (Pinheiro & Bates, 2000) do software R (R Development Core Team, 2008). Utilizou-se a rotina apropriada desenvolvida no programa R apresentada no Anexo D.

4 RESULTADOS E DISCUSSÃO

O crescimento do cafeeiro evidenciado pelo comportamento do diâmetro da copa, conforme se pode observar na Figura 2, mostra que até aproximadamente 450 dias houve uma tendência de crescimento linear e, após esse período o desenvolvimento da copa tende a ser lento ou até estável.

No entanto, como as medições foram realizadas em intervalos não igualmente espaçados, isto poderá afetar o ajuste de algum modelo. Como por exemplo, entre os tempos de 200 e 400 dias não houve nenhuma avaliação.

Observando o comportamento ilustrado na Figura 2, este sugere que o desenvolvimento do diâmetro de copa é aproximadamente quadrático ou cúbico para os cinco tratamentos.

Alguns autores como Santana et. al. (2004), estudando o crescimento do cafeeiro adensado influenciado por diferentes níveis de irrigação localizada, utilizaram a plotagem simples de uma curva passando pelo ponto médio das observações, como na Figura 2 e, a partir daí fizeram inferência sobre o comportamento do fenômeno em estudo, concluindo sobre as diferenças, significativas ou não, entre os tratamentos. Aqui, os resultados apresentados na Figura 2 servem apenas de base para se ter uma idéia do comportamento do crescimento do diâmetro da copa do cafeeiro sob diferentes manejos de irrigação. Entretanto, pode-se verificar que, visualmente, o comportamento das plantas conduzidas sem irrigação diferem das dos demais tratamentos.

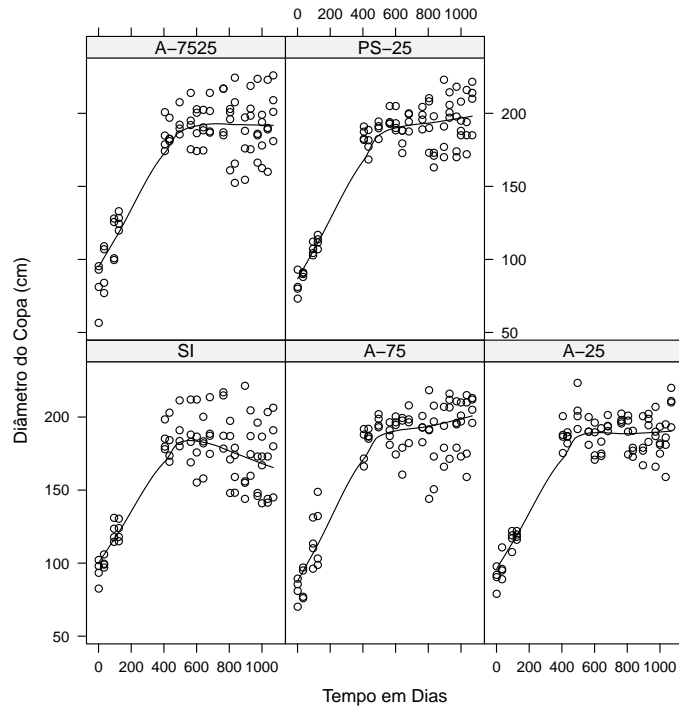


FIGURA 2 Representação gráfica dos valores observados do diâmetro da copa com relação ao tempo de avaliação para os cinco tratamentos

4.1 Ajuste do modelo linear polinomial

Na Tabela 1 são apresentados os coeficientes de regressão ajustados pelo método dos mínimos quadrados ordinários. Todos os coeficientes foram estatisticamente significativos pelo teste t ($\alpha = 0,01$).

Nota-se que as estimativas para o intercepto (β_0) nos tratamentos SI, A-75, A-25 e A-7525 foram semelhantes, com pouca variação entre elas, mas, bastante diferentes da estimativa para o tratamento PS-25, o qual apresentou valor menor do que nos irrigados, como um indicativo de que estas plantas apresentaram, em média, tamanho menor no momento em que se iniciou as avaliações. Já as estima-

tivas do efeito linear (β_1) não se diferenciaram muito umas das outras. Observa-se que a estimativa do efeito linear para o tratamento SI indica uma menor inclinação da curva, o que pode representar um crescimento mais lento das plantas.

TABELA 1 Estimativas dos coeficientes de regressão referentes ao ajuste do modelo linear polinomial a dados do diâmetro de copa de cafeeiro para os cinco tratamentos.

Tratamento	β_0	β_1	β_2	β_3
SI	83,1122	0,4316	-0,0006	2,35E-07
A-75	83,0701	0,4989	-0,0007	3,23E-07
A-25	81,6638	0,4794	-0,0007	3,07E-07
A-7525	83,2296	0,4813	-0,0007	2,87E-07
PS-25	73,8039	0,4586	-0,0006	2,56E-07

Todos os coeficientes estimados são significativos a 1% pelo teste t.

Percebe-se que todas as estimativas dos efeitos quadráticos (β_2) e cúbico (β_3) tiveram valores bem pequenos, entretanto todos tiveram um alto grau de significância. Assim, optou-se por não desconsiderá-los, o que poderia ser pensado principalmente com o coeficiente relacionado ao termo cúbico.

A Figura 3 mostra os quantis observados, em função dos quantis esperados para os resíduos do ajuste polinomial referente a cada tratamento. Pelo teste Shapiro-Wilk observa-se que apenas para o tratamento A-75 o pressuposto de normalidade da distribuição dos resíduos padronizados não foi satisfeita.

A Figura 4 mostra os ajustes do modelo linear polinomial para cada tratamento, e os intervalos de confiança para a média. Percebe-se pelo que todas as curvas tiveram um bom ajuste, verificadas pelos valores de R^2 acima de 0,85; sugerindo que os modelos explicam acima de 85% das variações ocorridas no diâmetro de copa devido as variações no tempo.

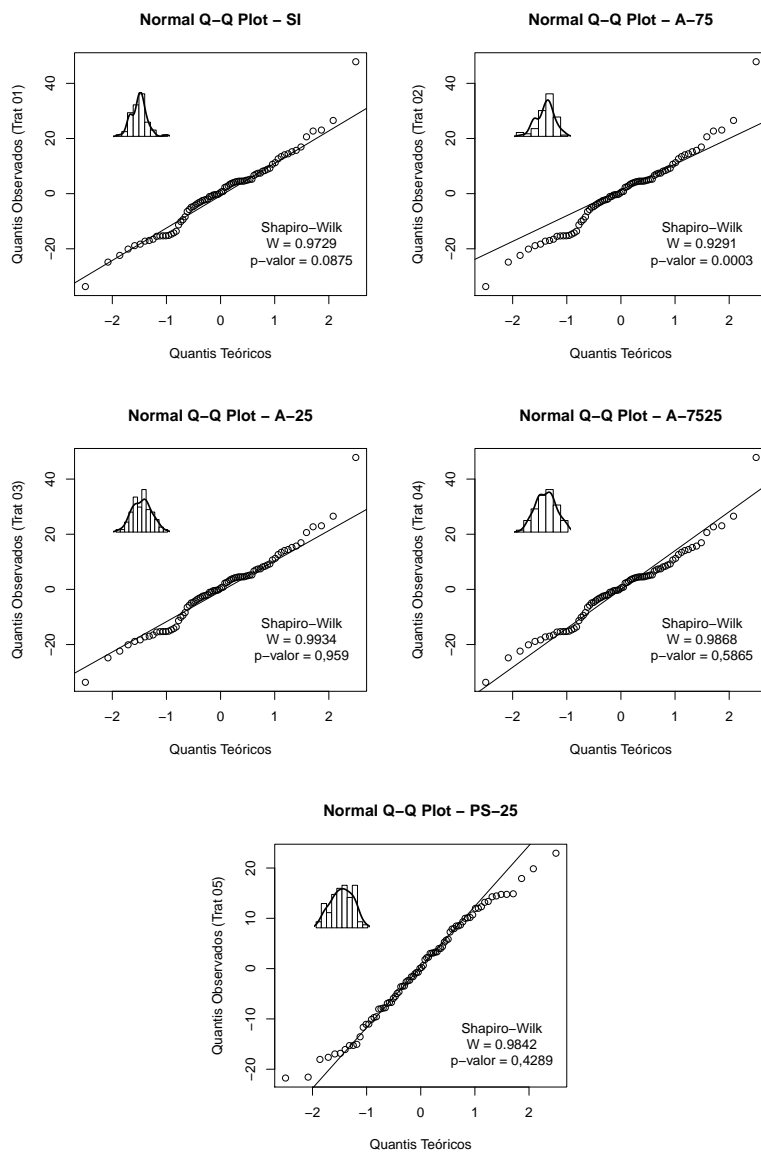


FIGURA 3 Representação gráfica da relação entre os quantis observados e os quantis teóricos esperados relativos aos resíduos do ajuste do modelo linear polinomial para os cinco tratamentos; histograma dos resíduos; estatística do teste de normalidade de Shapiro-Wilk.

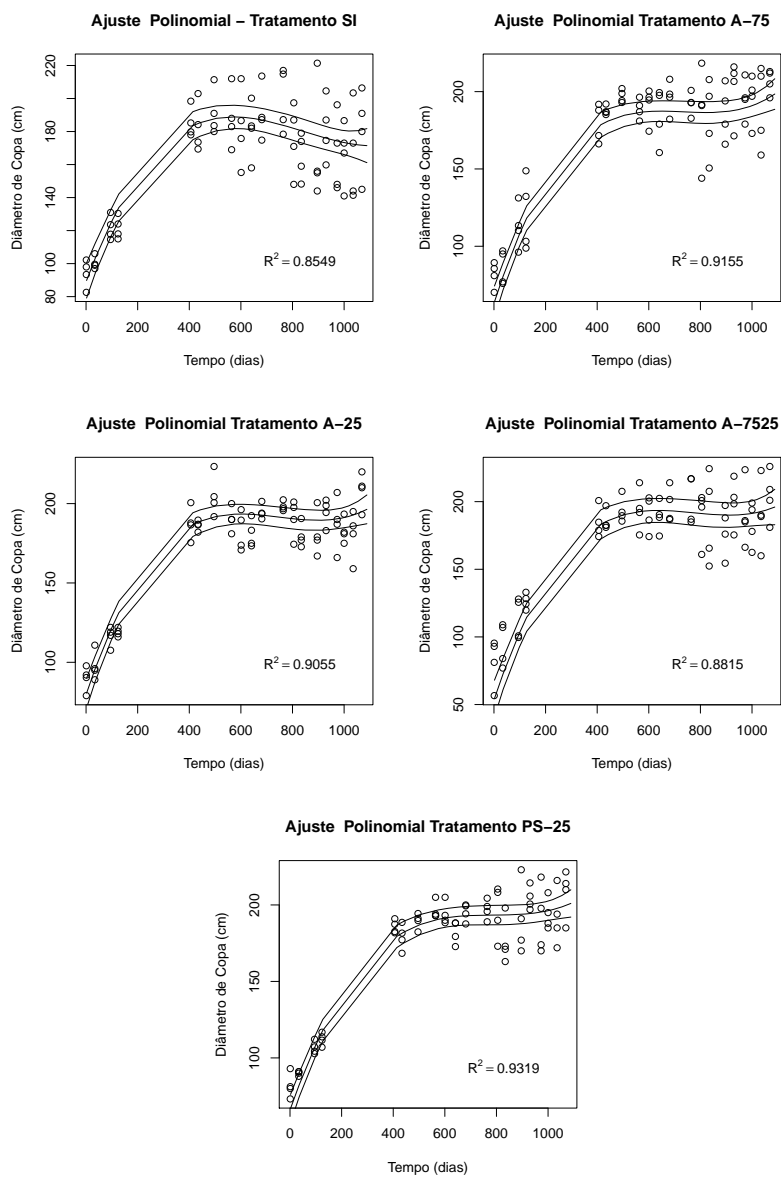


FIGURA 4 Representação gráfica do modelo linear polinomial, intervalos de confiança e estimativas do coeficiente de determinação (R^2) ajustado para os cinco tratamentos.

Alves et al. (2000) avaliando os efeitos de diferentes lâminas de irrigação

e parcelamentos para a aplicação de fertilizantes via água de irrigação no crescimento do cafeeiro, utilizando um modelo polinomial linear para avaliar o diâmetro da copa, entre outras variáveis, encontraram o valor de 0,2697 para o coeficiente linear, o qual indica o crescimento médio esperado do diâmetro da copa. Esse valor difere em muito dos valores aqui encontrados, que foram acima de 0,43.

4.2 O ajuste do modelo linear com efeitos mistos

O modelo linear com efeitos mistos, considerando-se tratamento e blocos fixos e indivíduo como aleatório, da mesma forma como o ajuste do modelo polinomial sem efeitos aleatórios, apresentou bom ajuste quando se utilizou o logaritmo do tempo como variável independente. Quando utilizou-se o tempo na forma original, os pressupostos da análise da variância, normalidade e homogeneidade da variância dos resíduos, não foram satisfeitos e não foi possível a modelagem da mesma para contornar o problema. Na Figura 5 são apresentados a relação entre os quantis observados e os quantis teóricos esperados relativos aos resíduos do ajuste do modelo linear misto quando se utilizou o logaritmo do tempo, em (a), e resíduos padronizados em função do tempo em (b). Pelo teste Shapiro-Wilk observa-se que o pressuposto de normalidade da distribuição dos resíduos padronizados foi satisfeita. Também em (b), observa-se que os resíduos apresentam uma distribuição regular ao longo do tempo não sendo indicativo de heterogeneidade de variância.

Na Figura 6 tem-se a representação gráfica dos valores ajustados utilizando o modelo linear misto para cada tratamento. Nota-se que os resultados são bastante parecidos com aqueles obtidos quando se utilizou o modelo linear polinomial; o tratamento sem irrigação segue uma tendência diferente dos tratamentos com irrigação, os quais apresentaram o mesmo comportamento.

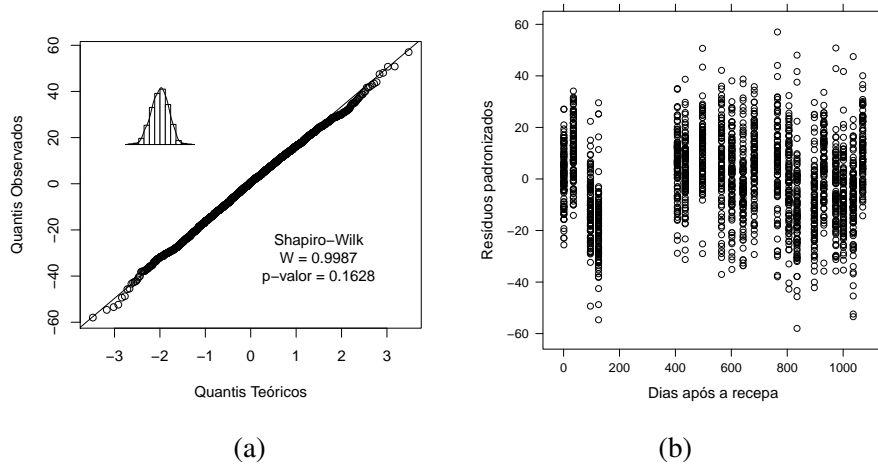


FIGURA 5 Representação gráfica da relação entre os quantis observados e os quantis teóricos esperados relativos aos resíduos do ajuste do modelo linear misto (a) e resíduos padronizados em função do tempo (b).

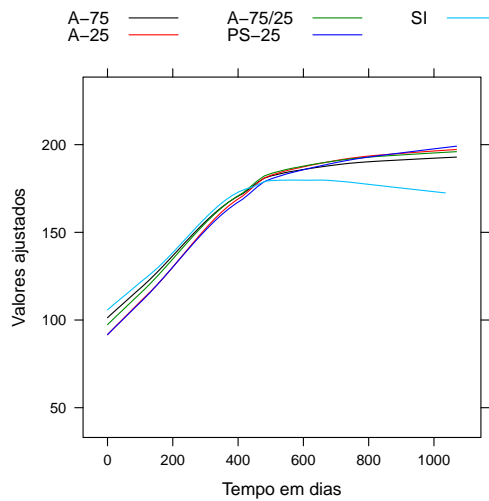


FIGURA 6 Representação gráfica dos valores ajustados utilizando o modelo linear misto para cada tratamento.

Na Figura 7 estão ilustrados valores ajustados para nove indivíduos utilizando

o modelo linear de efeitos mistos. Cabe aqui lembrar que estes resultados referem-se a um modelo que não satisfaz os pressupostos da análise da variância, ou seja, as estimativas foram feitas em função do tempo e não em função de seu logaritmo, utilizado-se como ferramenta para o ajuste do modelo linear misto, a função *lme* da biblioteca *nlme* que permite esse tipo de apresentação utilizando-se a função *augpred*. Para as estimativas anteriores foi utilizada a biblioteca *lmer4* que não oferece este tipo de representação. O propósito aqui é então colocar apenas como forma ilustrativa de uma opção de apresentação. Entretanto, visualmente, pode-se notar que o ajuste foi bom para todos os indivíduos.

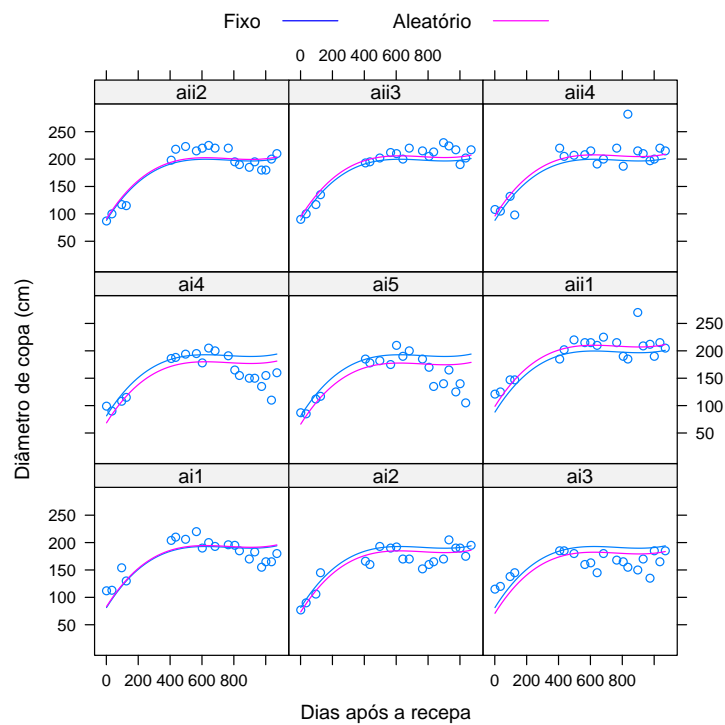


FIGURA 7 Representação gráfica dos valores ajustados para cada indivíduo (ou planta) utilizando o modelo linear de efeitos mistos.

4.3 O ajuste do modelo não linear

Na Figura 8 está representado o ajuste do modelo não linear logístico para os cinco tratamentos. As curvas tiveram um bom ajuste, verificadas pelos valores R^2 acima de 0,82 para os tratamentos com irrigação e 0,67 para o sem irrigação. Percebe-se que existe diferença entre os tratamentos com irrigação e o sem irrigação que atinge a assíntota em um valor inferior aos demais tratamentos, indicando que há um crescimento menor quando se não se faz irrigação. Os demais tratamentos apresentam comportamento semelhante. A diferença fica clara, quando apresenta-se graficamente os coeficientes estimados, conforme a Figura 9. Na tabela 2 são apresentadas as estimativas dos parâmetros para o ajuste do modelo não linear logístico ao diâmetro de copa para os cinco tratamentos.

Na Figura 10 está a representação gráfica da relação entre os quantis observados e os quantis teóricos esperados relativos aos resíduos do ajuste do modelo não linear logístico e estatística do teste de normalidade de Shapiro-Wilk. Apenas o tratamento A-75 não apresentou resíduos com distribuição normal após o ajuste da curva.

Na Figura 11 está representado o ajuste do modelo não linear Gompertz para os cinco tratamentos. Percebe-se que todas as curvas tiveram um bom ajuste, verificadas pelos valores R^2 acima de 0,82 para os tratamentos com irrigação e 0,67 para o sem irrigação; sugerindo que os modelos explicam acima de 67% das variações ocorridas no diâmetro de copa devido às variações no tempo. Percebe-se que, como no modelo logístico anterior, o tratamento sem irrigação atinge a assíntota em um valor inferior aos demais tratamentos, indicando que há um crescimento menor quando não se faz irrigação. Os demais tratamentos apresentam comportamento semelhante. A diferença fica clara, quando apresenta-se graficamente os coeficientes estimados, conforme a Figura 12.

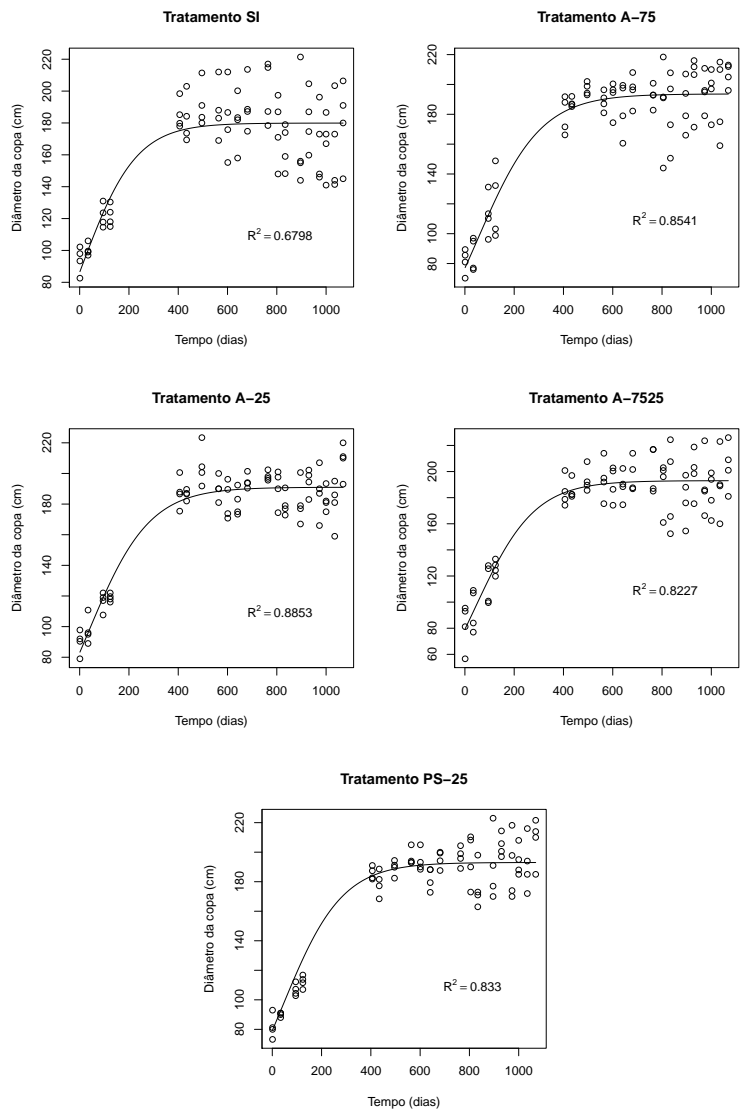


FIGURA 8 Representação gráfica do modelo não linear logístico para os cinco tratamentos.

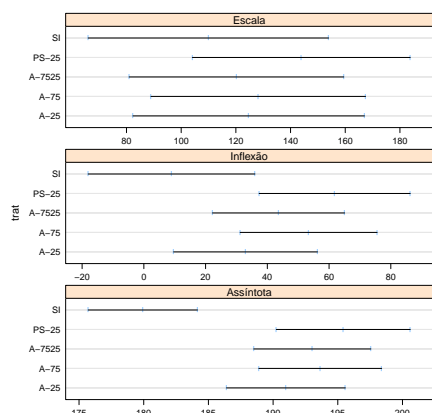


FIGURA 9 Representação gráfica dos intervalos de confiança das estimativas dos parâmetros escala (S), de inflexão (I) e assíntota (A) para o ajuste do modelo não linear logístico ao diâmetro de copa para os cinco tratamentos.

TABELA 2 Estimativas dos parâmetros para o ajuste do modelo não linear logístico ao diâmetro de copa para os cinco tratamentos.

Tratamento	Coeficientes estimados		
	Assíntota	Inflexão	Escala
SI	179,9308	8,9738	109,9529
A-75	193,6441	53,3837	128,1875
A-25	190,9920	32,9499	124,6927
A-7525	193,0303	43,6350	120,2205
PS-25	195,4257	61,8242	143,9367

Todos os coeficientes estimados são significativos a 1% pelo teste t.

Na tabela 3 são apresentadas as estimativas dos parâmetros para o ajuste do modelo não linear Gompertz relativas ao diâmetro de copa para os cinco tratamentos.

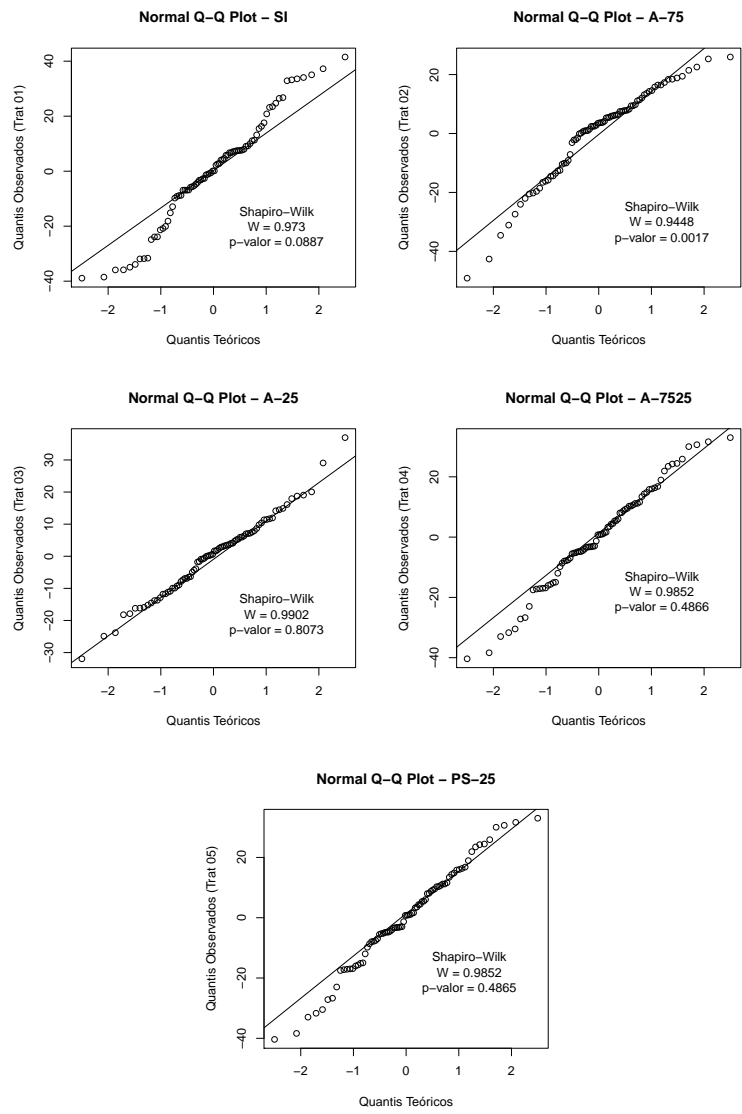


FIGURA 10 Representação gráfica da relação entre os quantis observados e os quantis teóricos esperados relativos aos resíduos do ajuste do modelo não linear logístico e estatística do teste de normalidade de Shapiro-Wilk para os cinco tratamentos.

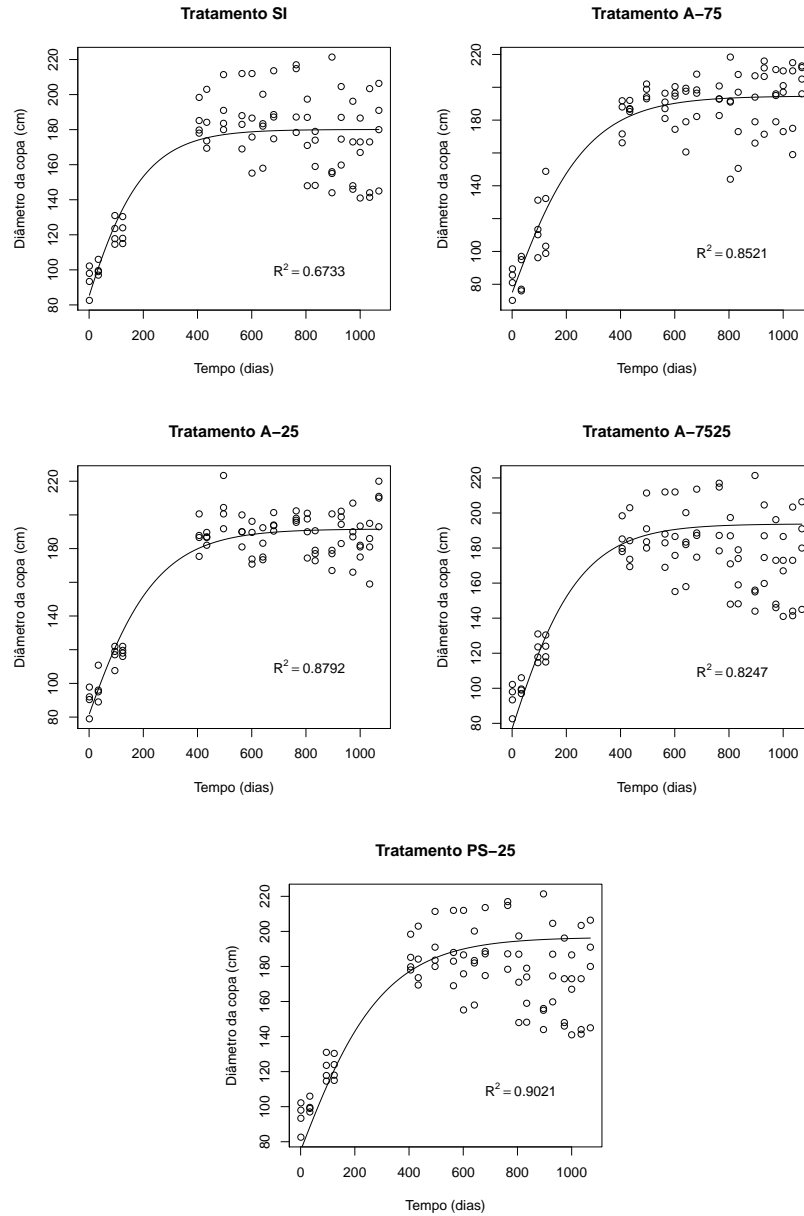


FIGURA 11 Representação gráfica do modelo não linear de Gompertz para os cinco tratamentos.

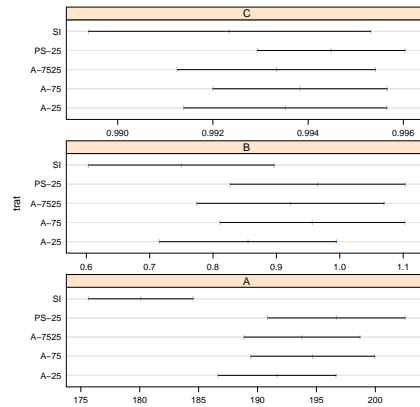


FIGURA 12 Representação gráfica dos intervalos de confiança das estimativas dos parâmetros assíntota (A), parâmetro numérico relacionado ao valor da função quando o tempo tende a zero $t = 0$ (B) e, parâmetro numérico para a escala no eixo do tempo t (C) para o ajuste do modelo não linear Gompertz ao diâmetro de copa para os cinco tratamentos.

Na Figura 13 está a representação gráfica da relação entre os quantis observados e os quantis teóricos esperados relativos aos resíduos do ajuste do modelo não linear Gompertz e estatística do teste de normalidade de Shapiro-Wilk. Apenas o tratamento A-75 não apresentou resíduos com distribuição normal após o ajuste da curva.

TABELA 3 Estimativas dos parâmetros para o ajuste do modelo não linear Gompertz ao diâmetro de copa para os cinco tratamentos.

Tratamento	Coeficientes estimados		
	A	B	C
SI	180,0955	0,7501	0,9924
A-75	194,6957	0,9570	0,9938
A-25	191,6607	0,8552	0,9935
A-7525	193,7807	0,9224	0,9933
PS-25	196,6974	0,9652	0,9945

Todos os coeficientes estimados são significativos a 1% pelo teste t.

Os resultados encontrados por Carvalho et al. (2006), evidenciam que o diâmetro de copa tem um comportamento assintótico, sendo esta tendência descrita por um modelo logístico de três parâmetros. Os resultados aqui obtidos corroboram os de Carvalho et al. (2006), os quais também haviam detectado que os tratamentos com irrigação apresentam uma assíntota significativamente maior que os sem irrigação.

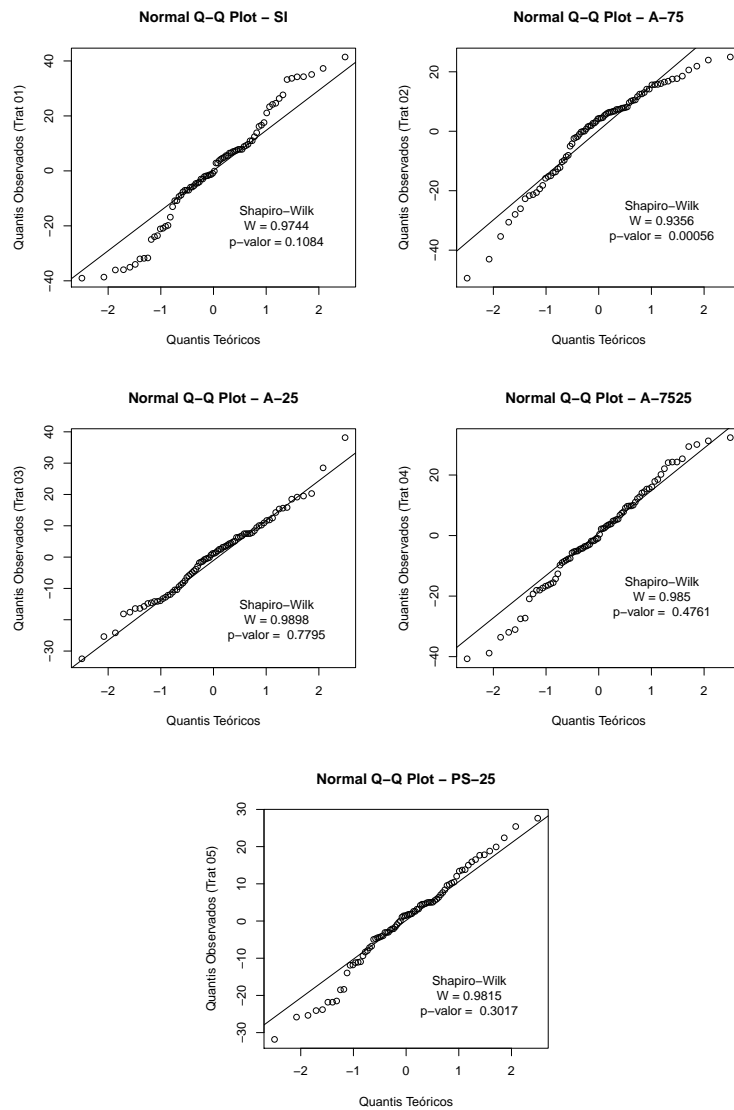


FIGURA 13 Representação gráfica da relação entre os quantis observados e os quantis teóricos esperados relativos aos resíduos do ajuste do modelo não linear de Gompertz e estatística do teste de normalidade de Shapiro-Wilk.

4.4 O ajuste dos modelos não linear com efeitos mistos

4.4.1 O modelo Logístico

O modelo logístico ajustado não apresentou resultados satisfatórios com relação aos pressupostos da análise da variância, principalmente ao que diz respeito a homogeneidade de variância. Foi então feita a modelagem da mesma e os modelos comparados por meio do teste da razão de verossimilhança.

Pelo teste da razão de verossimilhança (Tabela 4), percebe-se que há uma diferença significativa entre os dois modelos, sendo que, o modelo com modelagem da variância (CM) foi o que apresentou menor quadrado médio do resíduo ($\hat{\sigma}$), sugerindo ser este modelo mais adequado.

TABELA 4 Comparação entre o modelo com modelagem da variância (CM) e sem modelagem de variância (SM).

Modelo	GL	$\hat{\delta}$	$\hat{\sigma}$	LV	RV	Pr($\delta = 0$)
SM	31	-	16,016	-8377,177	-	-
CM	32	0,1308	6,959	-8353,953	46,446	<,0001

GL = graus de liberdade; $\hat{\delta}$ = parâmetro da função de variância; $\hat{\sigma}$ = quadrado médio do resíduo; LV = log-verossimilhança; RV = razão de verossimilhança entre os modelos.

Observou-se efeito de tratamento sobre as estimativas dos parâmetros de assíntota (A) e escala (S). Com relação ao parâmetro que representa o ponto de inflexão (I) não houve diferença significativa (Tabela 5).

TABELA 5 Resumo da análise da variância do modelo CM para o crescimento do diâmetro de copa do cafeeiro ao longo do tempo para os tratamentos (Logístico).

Parâmetro	GLnum	GLden	F	Valor-p
Assíntota	4	1846	6,99	<,0001
Inflexão	4	1846	1,22	0,2982
Escala	4	1846	2,55	0,0373

Observa-se que com relação à assíntota (A), conforme Figura 14, todos os

tratamentos com irrigação diferem estatisticamente da testemunha (sem irrigação), indicando que o uso da irrigação é eficiente no desenvolvimento do diâmetro de copa do cafeeiro.

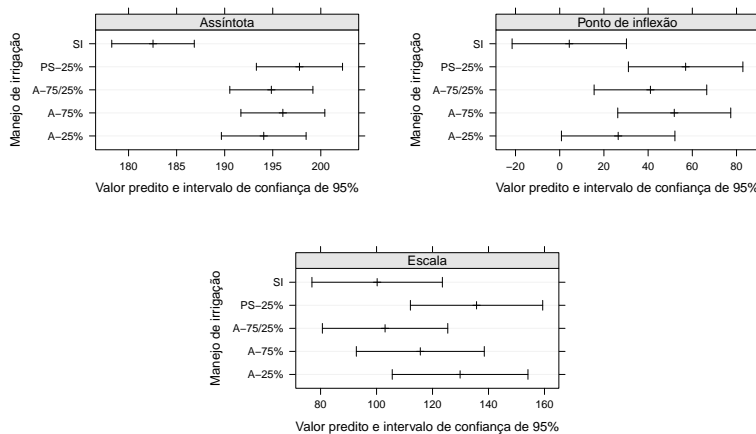


FIGURA 14 Representação gráfica do intervalo de confiança relativo aos parâmetros assíntota (A), Inflexão (I) e Escala (S) do modelo logístico, para os cinco tratamentos.

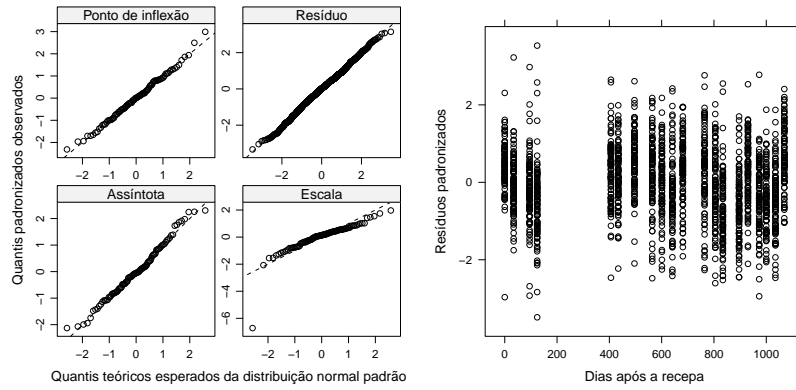
A Tabela 6 mostra os contrastes significativos entre os tratamentos reafirmando o resultado apresentado na Figura 14. Percebe-se que todos os tratamentos diferem do tratamento sem irrigação, com relação à assíntota, sugerindo ao pesquisador que com a irrigação o diâmetro de copa apresenta uma tendência de ser maior do que sem irrigação, o que pode resultar em uma maior produção. Coletando dados de produção destes mesmos indivíduos, Custódio (2009) confirma essa afirmativa, verificando uma produção significativamente maior para os tratamentos com irrigação, os quais, não diferiram entre si com relação a produção.

TABELA 6 Valores dos contrastes que foram significativos entre tratamentos ao nível de 5% para o modelo logístico.

Parâmetro	Contraste	Estimado	Erro padrão	GL	Lim inf.	Lim sup.
Assíntota	SI vs A-75	13,5074	2,4854	1846	8,6329	18,3818
	SI vs A-25	11,5195	2,5048	1846	6,6070	16,4320
	SI vs A-7525	12,3107	2,4607	1846	7,4847	17,1367
	SI vs PS-25	15,2345	2,5372	1846	10,2584	20,2105
Escala	SI vs PS-25	35,5147	13,7329	1846	8,5811	62,4482

A Figura 15(a) representa a verificação dos pressupostos de homogeneidade de variância e normalidade da distribuição dos efeitos aleatórios e resíduos (esquerda). Observa-se que as pressuposições do modelo foram satisfeitas. Na Figura 15(b), observa-se a distribuição dos resíduos padronizados com relação ao tempo; percebe-se que não há indícios de heteroscedasticidade da variância. O teste de Kolmogorov-Smirnov não rejeitou a hipótese nula de normalidade tanto para a distribuição dos parâmetros estimados (assíntota, $p = 0,786$); inflexão, $p = 0,995$) e escala, $p = 0,189$); quanto para o resíduo ($p = 0,398$) para o modelo com modelagem da variância.

A Figura 16 mostra claramente a diferença entre o tratamento SI e os demais, o que já era esperado, dado que, a região de origem dos dados passa por períodos secos no inverno fazendo com que o crescimento da copa das plantas seja prejudicado caso não haja irrigação.



(a) (b)

FIGURA 15 Representação gráfica dos quantis observados padronizados em função dos quantis esperados para a distribuição normal padrão para cada termo aleatório do modelo considerado (a) e diagrama de dispersão dos resíduos em função do tempo (b).

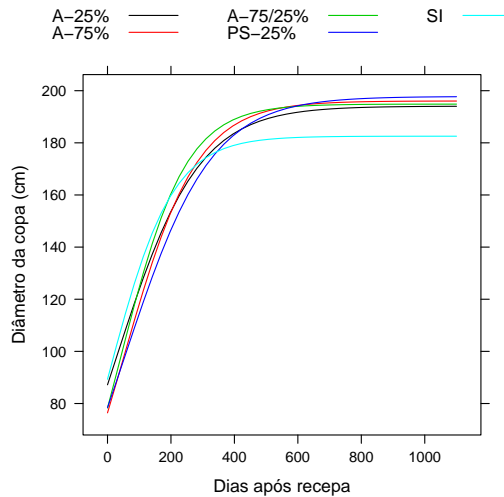


FIGURA 16 Representação gráfica dos valores preditos para o crescimento do diâmetro da copa do cafeeiro em função do tempo com os efeitos fixos do modelo, para cada manejo de irrigação.

De acordo com a Figura 17 observa-se que os valores preditos para cada indi-

víduo estão de acordo com o esperado, indicando que o modelo é adequado para explicar o crescimento do diâmetro da copa do cafeeiro.

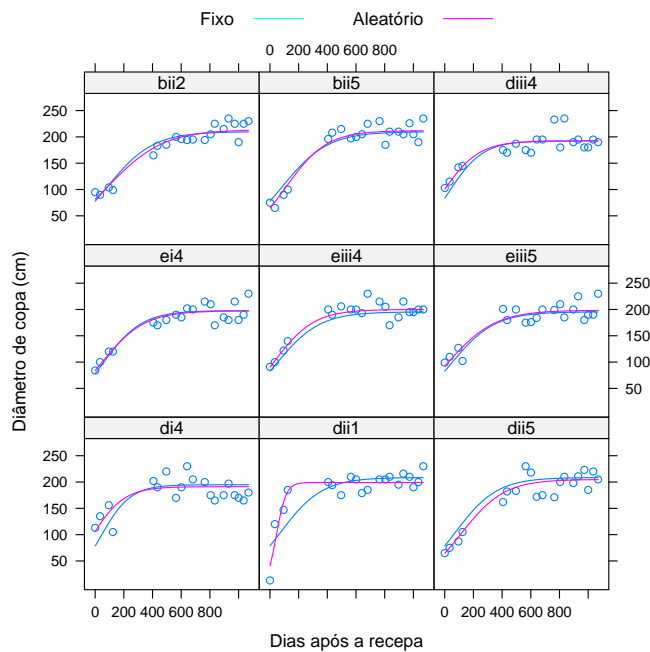


FIGURA 17 Representação gráfica dos valores preditos para o crescimento do diâmetro da copa de nove plantas de cafeeiro em função do tempo, utilizando o modelo logístico com efeitos mistos.

A Tabela 7 mostra as estimativas para os parâmetros do modelo logístico, para os cinco tratamentos; verificou-se que, exceto para o ponto de inflexão do tratamento sem irrigação, todas as demais estimativas foram significativas, pois não incluíram o valor zero (0) no seu intervalo de confiança.

TABELA 7 Estimativas dos parâmetros assíntota (A), ponto de inflexão (I), escala (S) e seus respectivos erros padrão e intervalos de confiança (95%), referentes ao ajuste do modelo logístico com efeitos mistos a dados do diâmetro de copa do cafeeiro.

Parâmetro	Tratamento	Estimativa	Erro padrão	Intervalo de confiança	
				Inferior	Superior
A	SI	182,5447	2,1887	178,2521	186,8374
	A-75	196,0521	2,2258	191,6867	200,4175
	A-25	194,0642	2,2470	189,6573	198,4712
	A-7525	194,8554	2,2040	190,5328	199,1781
	PS-25	197,7792	2,2853	193,2971	202,2613
I	SI	4,3274	13,1891	-21,5398	30,1946
	A-75	51,8294	13,0326	26,2691	77,3897
	A-25	26,4530	13,0896	0,7811	52,1249
	A-7525	41,0448	12,9917	15,5648	66,5247
	PS-25	56,9760	13,2055	31,0768	82,8752
S	SI	100,2133	11,8823	76,9091	123,5175
	A-75	115,6321	11,6538	92,7761	138,4882
	A-25	129,8500	12,3627	105,6036	154,0963
	A-7525	103,0441	11,4219	80,6428	125,4454
	PS-25	135,7279	12,0517	112,0916	159,3643

4.4.2 O modelo de Gompertz

A Tabela 8 mostra os resultados da análise da variância para o modelo Gompertz ajustado com relação a cada parâmetro. Observou-se efeito de tratamento sobre as estimativas de A (assíntota de crescimento do diâmetro da copa quando o tempo tende para o infinito), B (parâmetro numérico relacionado ao valor da função quando Tempo = 0) e C (parâmetro numérico para a escala no eixo do Tempo).

Na Figura 18 observa-se que com relação a assíntota (A), todos os tratamentos diferem estatisticamente da testemunha (sem irrigação). Observa-se também que com relação ao parâmetro C (parâmetro numérico para a escala no eixo do tempo t) a testemunha também difere dos demais tratamentos que não diferem entre si.

TABELA 8 Resumo da análise da variância para o crescimento do diâmetro de copa do cafeeiro ao longo do tempo para os tratamentos (Gompertz).

	GLnum	GLden	F	Valor-p
A	4	1846	683271	<,0001
B	4	1846	1562	<,0001
C	4	1846	5	<,0001

Os valores das estimativas dos contrastes podem ser vistos na Tabela 9.

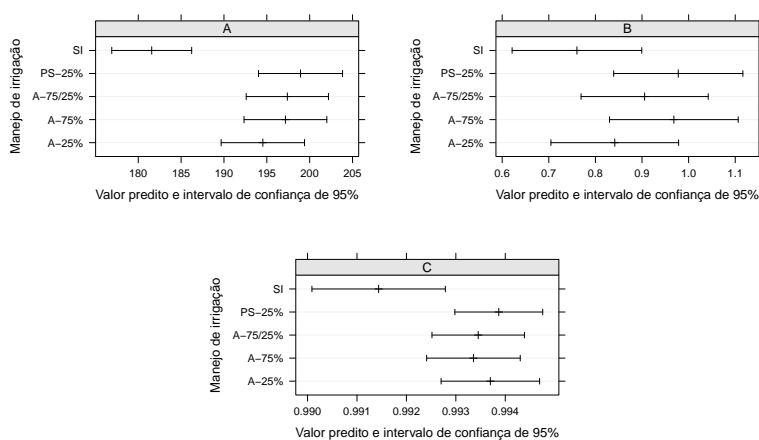
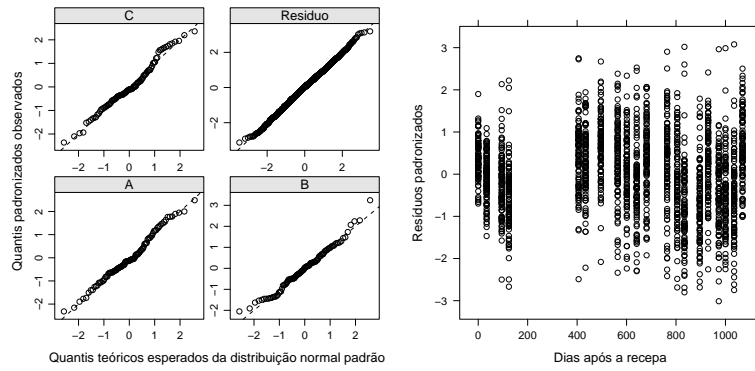


FIGURA 18 Representação gráfica dos intervalos de confiança relativos ao parâmetro assíntota (A), parâmetro numérico relacionado ao valor da função quando o tempo tende a zero $t = 0$ (B) e, parâmetro numérico para a escala no eixo do tempo t (C).

TABELA 9 Valores dos contrastes que foram significativos entre tratamentos em nível de 1% para o modelo Gompertz.

Parâmetro	Contraste	Estimado	Erro Padrão	GL	Lim. inf.	Lim. sup.
A	SI vs A-75	15,601	2,720	1846	10,267	20,934
	SI vs A-25	12,966	2,739	1846	7,594	18,337
	SI vs A-7525	15,823	2,713	1846	10,502	21,145
	SI vs PS-25	17,362	2,763	1846	11,944	22,781
C	SI vs A-75	0,002	0,001	1846	6,01E-04	3,24E-03
	SI vs A-25	0,002	0,001	1846	8,97E-04	3,62E-03
	SI vs A-7525	0,002	0,001	1846	7,27E-04	3,30E-03
	SI vs PS-25	0,002	0,001	1846	1,15E-03	3,71E-03

A Figura 19 representa a verificação dos pressupostos de homogeneidade de variância e normalidade da distribuição dos efeitos aleatórios e resíduos (a). Observa-se que as pressuposições do modelo foram satisfeitas. Em (b), observa-se a distribuição dos resíduos padronizados com relação ao tempo. Percebe-se que não ha indícios de heteroscedasticidade da variância. O teste de Kolmogorov-Smirnov não rejeitou a hipótese nula de normalidade tanto para a distribuição dos parâmetros estimados (A, $p = 0,3187$; B, $p = 0,9856$ e C, $p = 0,3608$); quanto para o resíduo ($p = 0,5078$).



(a)

(b)

FIGURA 19 Representação gráfica dos quantis observados padronizados em função dos quantis esperados para a distribuição normal padrão para cada termo aleatório do modelo considerado (a) e diagrama de dispersão dos resíduos em função dos valores ajustados (b).

A Figura 20 mostra claramente, tanto no ajuste do modelo Logístico quanto no de Gompertz, a diferença entre o tratamento SI e os demais, o que já era esperado, dado que, a região de origem dos dados passa por períodos secos no inverno fazendo com que o crescimento da copa das plantas seja prejudicado caso não haja irrigação.

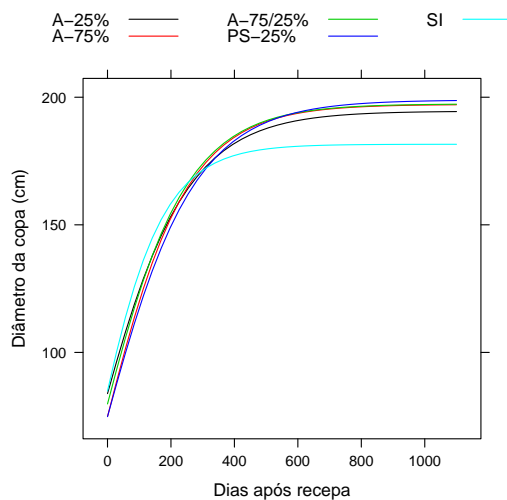


FIGURA 20 Representação gráfica dos valores preditos para o crescimento do diâmetro da copa do cafeeiro em função do tempo com os efeitos fixos do modelo, para cada manejo de irrigação.

De acordo com a Figura 21 observa-se que os valores preditos para cada indivíduo estão de acordo com o esperado, indicando que o modelo é adequado para explicar o crescimento do diâmetro da copa do cafeeiro. Não foram encontrados estudos utilizando-se esta metodologia de análise envolvendo o crescimento do cafeeiro.

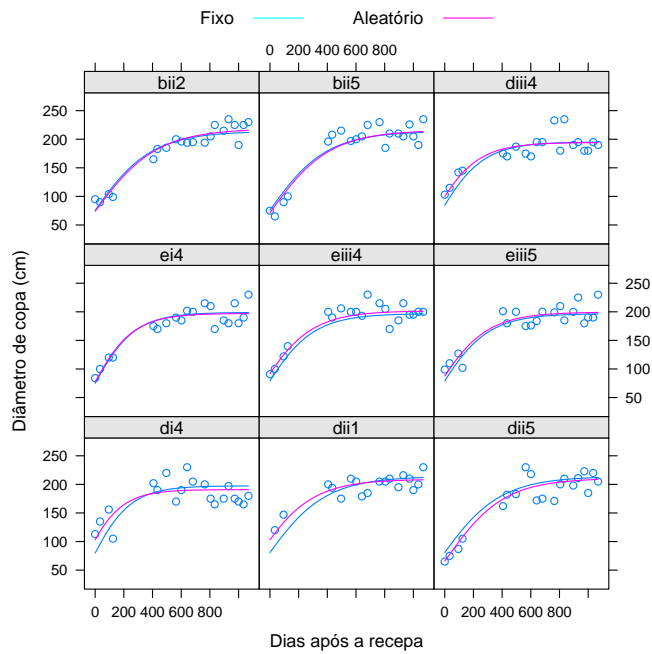


FIGURA 21 Representação gráfica dos valores preditos para o crescimento do diâmetro da copa de nove plantas de cafeeiro em função do tempo, utilizando o modelo de Gompertz com efeitos mistos.

A Tabela 10 mostra que as estimativas para os parâmetros do modelo de Gompertz, para os cinco tratamentos foram significativas a 1%, pelo teste t, pois não incluíram o zero (0) no intervalo de confiança.

TABELA 10 Estimativas dos parâmetros (A), (I), (S) e seus respectivos erros padrão e intervalos de confiança (95%), referentes ao ajuste do modelo gompertz a dados do diâmetro de copa do cafeeiro significativas a 1%.

Parâmetro	Tratamento	Estimativa	Erro padrão	Intervalo de confiança	
				Inferior	Superior
A	SI	181,5707	2,3771	176,9086	186,2329
	A-75	197,1714	2,4622	192,3425	202,0004
	A-25	194,5365	2,4803	189,6721	199,4010
	A-7525	197,3941	2,4468	192,5954	202,1928
	PS-25	198,9331	2,4993	194,0314	203,8348
B	SI	0,7604	0,0709	0,6213	0,8995
	A-75	0,9683	0,0704	0,8303	1,1064
	A-25	0,8416	0,0698	0,7046	0,9785
	A-7525	0,9056	0,0695	0,7692	1,0420
	PS-25	0,9778	0,0707	0,8392	1,1164
C	SI	0,9914	0,0007	0,9901	0,9928
	A-75	0,9934	0,0005	0,9924	0,9943
	A-25	0,9937	0,0005	0,9927	0,9947
	A-7525	0,9935	0,0005	0,9925	0,9944
	PS-25	0,9939	0,0005	0,9930	0,9948

4.5 Considerações finais

Ainda, é interessante aqui ressaltar que, em estudos relacionados a genética, em que existe a variabilidade inerente a cada indivíduo em particular, os modelos mistos podem ser uma boa opção, dado que, apresentam como resultado também estimativas individuais. Sugere-se para trabalhos futuros a proposição de outros modelos e/ou a criação de um modelo direcionado para o cafeeiro. Sugere-se também um estudo da qualidade de ajuste das curvas.

Com relação ao ajuste de modelos de regressão a dados de diâmetro de copa de cafeeiro o modelo linear polinomial sem ou com efeitos mistos ajusta de forma satisfatória o crescimento do diâmetro de copa do cafeeiro. Entretanto, devido à não aceitação do pressuposto da homogeneidade de variância, observou-se a necessidade de transformação logaritmo do tempo para o modelo com efeitos mistos.

Cabe aqui lembrar, como desvantagem desse método, a dificuldade que se tem em fazer-se previsão mais precisa para os tratamentos no tempo quando se utilizam esses modelos. Ademais, os parâmetros estimados não têm nenhuma interpretação direta.

O modelos sigmoidais logístico e o modelo Gompertz considerando-se apenas efeitos fixos apresentaram resultados satisfatórios, com a vantagem de que seus parâmetros possuem uma interpretação biológica.

A modelagem do crescimento do diâmetro de copa do cafeeiro utilizando a tanto a função Logística quanto a Gompertz, incluindo-se efeitos fixos e aleatórios e modelando-se a heterogeneidade da variância, caso seja necessário, apresentou resultados satisfatórios. Assim, sugere-se o uso destes modelos para o estudo do desenvolvimento do diâmetro de copa do cafeeiro. Esta metodologia é recomendada, em detrimento do simples ajuste polinomial ou um modelo misto linear, em estudos de crescimento resultando em boas estimativas de parâmetros que tem a vantagem de serem biologicamente interpretáveis.

5 CONCLUSÃO

Os modelos não lineares logístico e de Gompertz são adequados para a descrição do desenvolvimento do diâmetro de copa do cafeeiro.

A utilização de modelos de efeito misto é viável, tanto para o ajuste linear quanto para o não linear.

A irrigação contribui para um aumento significativo no desenvolvimento vegetativo do cafeeiro.

REFERÊNCIAS BIBLIOGRÁFICAS

ALVES, M. E. B.; FARIA, M. A.; GUIMARÃES, R. J.; MUNIZ, J. A.; SILVA E. L. Crescimento do cafeeiro sob diferentes lâminas de irrigação e fertirrigação. **Revista Brasileira de Engenharia Agrícola e Ambiental**, Campina Grande, v.4, n.2, p.219-225, maio/ago. 2000.

ARAÚJO, J.A.C. de. **Análise do comportamento de uma população de café Icatu (H-4782-7) sob condições de irrigação por gotejamento e quebra-vento artificial**. 1982. 87 p. Dissertação (Mestrado em Irrigação e Drenagem) - Escola Superior de Agricultura "Luiz de Queiroz", Piracicaba.

BATES, D.M.; WATTS, D.G. **Nonlinear Regression Analysis and its Applications**. New York: J. Wiley, 1988. 365p.

CALEGARIO, N.; DANIELS, R. F.; MAESTRI, R.; NEIVA, R. Desenvolvimento de um modelo linear de efeito misto na estimativa do crescimento e produção de povoamentos clonais de Eucalyptus. **Revista Cerne**, Lavras, v.10, n. 1, p. 67-86, jan./jun. 2004.

CALEGARIO, N.; DANIELS, R. F.; MAESTRI, R.; NEIVA, R. Modeling dominant height growth based on nonlinear mixed-effects model: a clonal Eucalyptus plantation case study. **Forest Ecology and Management**, Amsterdam, v.204, n. 1, p. 11-20, Jan. 2005.

CAMARGO, A. P. Necessidades hídricas do cafeeiro. In: CURSO PRÁTICO INTERNACIONAL DE AGROMETEOROLOGIA, 3., 1989, Campinas. Anais... Campinas: IAC, 1989. p. 22.

CARVALHO, C. H. M.; COLOMBO, A.; SCALCO, M. S.; MORAIS, A. R. de. Evolução do cafeeiro (*Coffea arabica* L.) irrigado e não irrigado em duas densidades de plantio. **Ciência e Agrotecnologia**. Lavras, v.30, n.2, p. 243-250, mar/abr, 2006.

CUSTÓDIO, A. A. de P. **Manejo da irrigação na lavoura cafeeira (Coffea arabica L.) recepada**. 2009. 85 p. Dissertação (Mestrado em Engenharia Agrícola) - Universidade Federal de Lavras, Lavras.

DIGGLE, P. J.; LIANG K. Y.; ZEGER S. L. **Analysis of longitudinal data**. Oxford: Clarendon, 1994. 379 p.

- DODGE, Y. **The Concise Encyclopedia of Statistics**. New York: Springer, 2008. 616p.
- DRAPER, N.R.; SMITH, H. **Applied regression analysis**. 3ªed. New York: J. Wiley, 1998. 706p.
- FARAWAY, J. J. **Linear Models with R**. New York: Chapman/CRC, 2004. 229p.
- FAUSTO, M. A.; CARNEIRO, M.; ANTUNES, C. M. F.; PINTO, J.A.; COLOSIMO, E. A. O modelo de regressão linear misto para dados longitudinais: uma aplicação na análise de dados antropométricos desbalanceados. **Cadernos de saúde pública**. Rio de Janeiro, v. 24, n.3, p.513-524, mar. 2008.
- FIGUEIREDO, V. B.; FARIA, M. A.; SILVA, E. L. Crescimento inicial do cafeeiro irrigado com água salina e salinização do solo. **Revista Brasileira de Engenharia Agrícola Ambiental**. Campina Grande, v.10, n.1, p. 50-57, mar. 2006.
- FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS. Disponível em: <<http://faostat.fao.org/>>. Acesso em: 20 ago. 2009.
- GALLANT, A.R. **Nonlinear Statistical Models**. New York: J. Willey, 1987. 624p.
- GRAYBILL, F. A.; IYER, H. K. **Regression Analysis: Concepts and Applications**. Belmont: Duxbury, 2004. 699p.
- GUJARATI, D. **Econometria Básica**. 4ª ed. São Paulo: Campus, 2006. 812p.
- INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Disponível em: <<http://www.sidra.ibge.gov.br/bda/tabela/protabl.asp?c=1613&z=t&o=11&i=P>>. Acesso em: 20 ago. 2009.
- JORGENSEN, B. **The Theory of Linear Models**. New York: Chapman/CRC, 1993. 226p.
- LINDSTRON, M. J.; BATES, D. M. Non linear mixed-effects models for repeated measures data. **Biometrics**, Washington, v. 46, n. 3, p. 673-687, Sept. 1990.
- MARTINS, C. C.; REIS, E. F.; BUSATO, C.; PEZZOPANE, J. E. M. Crescimento inicial do café Conilon (*Coffea canephora* Pierre ex Froehner) sob diferentes lâminas de irrigação. *Revista Engenharia na Agricultura*, Viçosa, v. 14, n. 3, p. 193-201, jul./set. 2006.

MATIELLO, J. B.; DANTAS, S. F. de A. de. Desenvolvimento do cafeeiro e do sistema radicular com e sem irrigação em Brejão, PE. In: CONGRESSO BRASILEIRO DE PESQUISAS CAFEIRAS, 14., 1987, Campinas. **Resumos...** Rio de Janeiro: IBC-Cerca, 1987. p. 165.

MATIELLO, J. B.; SANTINATO, R.; GARCIA, A. W. R.; ALMEIDA, S. R.; FERNANDES, D. R. **Cultura de café no Brasil: novo manual de recomendações.** Rio de Janeiro: Fundação Procafé, 2002. 387 p.

MENDES, A. N. G.; GUIMARÃES, R. J. **Cafeicultura empresarial: produtividade e qualidade (genética e melhoramento de cafeeiro).** Lavras: UFLA/FAEPE, 1996. 99 p.

OLIVEIRA, L. A. M. **Desenvolvimento vegetativo e reprodutivo do cafeeiro (Coffea arabica L.) irrigado em diferentes épocas do ano.** 2003. 54 p. Dissertação (Mestrado em Engenharia Agrícola) - Universidade Federal de Lavras, Lavras.

PATTERSON, H. D.; THOMPSON, R. Recovery of inter-block information when block sizes are unequal. **Biometrika**, Cambridge, v. 58, n. 3, p. 545-54, Dec. 1971.

PEEK, M. S.; RUSSEK-COHEN, D. A. W.; FORSETH, I. N. Physiological response curve analysis using nonlinear mixed models. **Oecologia**, Berlin, v. 132, n. 2, p. 175-180, June 2002.

PINHEIRO, C. J.; BATES, D. M. **Mixed-effects models in S and S-PLUS.** New York: Springer Verlag, 2000. 528p.

RATKOWSKY, D. A. **Nonlinear regression modeling.** New York: M. Bekker, 1983. 276p.

RAVISHANKER, N.; DEY, D. K. **First course in linear model theory.** Boca Raton: CRC, 2001. 473 p.

R DEVELOPMENT CORE TEAM. **R: a language and environment for statistical computing.** Viena: R Foundation for Statistical Computing. 2004. Disponível em: <<http://www.R-project.org>>. Acesso em: 11 out. 2009

RENCHER, A. C.; SCHAALJE G. B. **Linear models in statistics.** 2. ed. New Jersey: J. Wiley, 2008. 672p.

SANTANA, M. S.; OLIVEIRA, C. A. S.; QUADROS, M. Crescimento inicial de duas cultivares de cafeeiro adensado influenciado por níveis de irrigação localizada. **Revista de Engenharia Agrícola**, Jaboticabal, v. 24, n. 3, p. 644-653, set./dez. 2004.

SCHABENBERGER, O.; PIERCE, F. J. **Contemporary statistical models for the plant and soil sciences**. Boca Raton: CRC, 2002. 753p.

SEARLE, S. R. **Linear models for unbalanced data**. New York: J. Wiley, 1987. 536p.

SEBER, G. A. F.; WILD, C. J. **Nonlinear Regression**. New York: J. Wiley, 1989. 752p.

SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality. **Biometrika**, Cambridge, v. 52, n. 3/4, p.591-611, Dec. 1965.

SILVA, W. J.; VOLPE C. A. Crescimento do fruto do cafeeiro (*Coffea arabica* L.) cv. Acaíá, CP 474/19 e suas relações com variáveis metereológicas em dois sistemas de plantio, no cerrado de Uberaba - MG. **Revista Brasileira de Agrometereologia**, Sete Lagoas, v. 13, n. 2, p. 292-302, 2005.

STEPHENS, M. A. EDF Statistics for Goodness of Fit and Some Comparisons. **Journal of the American Statistical Association**, New York, v. 69, n. 347, p.730-737, Sept. 1974.

VENABLES, W. N.; RIPLEY, B. D. **Modern applied statistics with S**. 4. ed. New York: Springer Verlag, 2002. 495p.

WEISBERG, S. **Applied linear regression**. 3. ed. New Jersey: J. Wiley, 2005. 310p.

ANEXOS

ANEXO A Rotina no R utilizado para a obtenção das estimativas dos parâmetros do modelo polinomial.

```
# Plot do crescimento de COPA considerando o TEMPO EM DIAS

### pdf("copa-poli.pdf")
xyplot(COPA~TEMPODIA|factor(TRAT), data=datadia,
type=c("p","smooth"), strip=strip.custom(bg="Grey95",
factor.levels=
c("SI","A-75","A-25","A-7525","PS-25")),col="1",
xlab="Tempo em Dias", ylab="Diâmetro do Copa (cm)",
auto.key=list(lines=TRUE, columns=3))
### dev.off()

# Ajuste polinomial para COPA - Tratamento 01

reg1 <- lm(COPA ~ TEMPODIA + I(TEMPODIA^2)
+ I(TEMPODIA^3) + factor(BLOCO),
data=datadia[datadia$TRAT==1,])

summary(reg1)
anova(reg1)

# Análise dos resíduos

shapiro.test(residuals(reg1, type="pearson"))

ks.test(residuals(reg1), "pnorm",
mean(residuals(reg1)),
sd(residuals(reg1)))

### pdf("copa-t1-pol.pdf")
qqnorm(reg1$residuals,
main="Normal Q-Q Plot - SI",
xlab="Quantis Teóricos",
ylab="Quantis Observados (Trat 01)")
qqline(reg1$residuals, col="1")
text(1.5, -18,
"Shapiro-Wilk
W = 0.9729
p-valor = 0.0875")
op <- par(fig=c(.1,.5,.4,0.9), new=TRUE)
hist(reg1$residuals, probability=T,
col="white", xlab="", ylab="", main="", axes=F)
lines(density(reg1$residuals), col="1", lwd=2)
### dev.off()
```

```

# Os demais tratamentos seguem o mesmo esquema de estimação e
  construção dos gráficos.

###   Cálculo para os tratamentos usando o lmList

require(nlme)
reg.copa <- lmList(COPA ~ (TEMPODIA + I(TEMPODIA^2)+ I(TEMPODIA^3)
  + factor(BLOCO))|factor(TRAT), data=datadia)

summary(reg.copa)
str(intervals(reg.copa))
intervals(reg.copa)[,1:4]

```

ANEXO B Rotina no R utilizado para a obtenção das estimativas dos parâmetros do modelo linear com efeitos mistos.

```

# Lendo os dados

jair <- read.table("mixed-jair.txt", header=TRUE, sep="\t")
attach(jair)
head(jair)

# Estrutura dos dados:

# 'data.frame':  1991 obs. of  6 variables:
# $ trat      : Factor w/ 5 levels "A-25","A-75", ...
# $ bloc      : Factor w/ 4 levels "i","ii","iii", ...
# $ rep       : int  1 2 3 4 5 1 2 3 4 5 ...
# $ individuo: Factor w/ 100 levels "A-25i1","A-25i2", ...
# $ copa      : num  112 77 115 99 87 121 87 90 108 105 ...
# $ tempo2    : num  1 1 1 1 1 1 1 1 1 1 ...

jair$tempo2 <- jair$tempo+1 # cria uma variável chamada
                          # "tempo2" em "jair" para contornar
                          # o valor "zero do primeiro tempo"

head(jair) # verificação da criação da variável

#Retirando indivíduos com observações faltando e as colunas
  #caule e altura
# eliminar as linhas em que as copas são "NA" e as colunas
# 3 (tempo),6 (caule) e 8 (altura) de jair

jair.copa <- subset(jair, copa!="NA")[,-c(3,6,8)]
names(jair.copa); head(jair.copa)

library(lme4)

# Rodando o modelo

modelo1<-lmer(copa ~ factor(bloc) + trat +
  ((tempo2)+I(tempo2^2)+I(tempo2^3))|trat/individuo),

```

```

data=jair.copa3)

summary(modelo1)
anova(modelo1)

pdf("lme2Fitted.pdf", w=5, h=5)
xyplot(fitted(modelo1)~tempo, groups=trat, data=jair.copa2,
  col=c("black", "red", "green4", "blue", "grey"), type="a",
  key=list(space= "top",
  text=list(c("SI", "A-75", "A-25", "A-7525", "PS-25"), col=1),
  lines=list(col= c("black", "red", "green4", "blue", "grey" )),
  columns=3),
  ,xlab="Tempo em dias", ylab="Valores ajustados", lty=1)
dev.off()

pdf("lme2Residuals.pdf", w=5, h=5)
xyplot(residuals(modelo1, level=1, type="pearson")~
  jair.copa3$tempo2, col=1, #type=c("p", "smooth"),
  xlab="Dias após a receita",
  ylab="Resíduos padronizados")
dev.off()

shapiro.test(residuals(modelo1, type="pearson"))

pdf("lme2Qqnorm.pdf", w=5, h=5)
qqnorm(residuals(modelo1),
  main="",
  xlab="Quantis Teóricos", ylab="Quantis Observados")
qqline(residuals(modelo1))
text(1.5, -32,
  "Shapiro-Wilk
W = 0.9991
p-valor = 0.4332")
op <- par(fig=c(.1,.5,.4,0.9), new=TRUE)
hist(residuals(modelo1), probability=T,
col="white", xlab="", ylab="", main="", axes=F)
lines(density(residuals(modelo1)), col="1", lwd=1)
dev.off()

ks.test(residuals(modelo1), "pnorm",
  mean(residuals(modelo1)), sd(residuals(modelo1)))

pdf("lme2BoxplotRes.pdf", w="5", h="5")
boxplot(residuals(modelo1, type="pearson")~
  jair.copa3$trat:jair.copa3$tempo2, xlab="Individuos")
dev.off()

pdf("lme2BplResFixxInd.pdf", w="5", h="5")
boxplot(residuals(modelo1), horizontal=TRUE, ylim=c(-100,100))
dev.off()

pdf("lme2FittedInd.pdf", w="5", h="5")
xyplot(augPred(modelo1, level=0:1), layout=c(10,10),
  xlab="Dias após a receita",
  ylab="Diâmetro de copa (cm)",

```

```

        key=list(text=list(c("Fixo","Aleatório")),
                lines=list(col=c(4,6), lty=1),
                columns=2),
strip=strip.custom(bg="grey95"
)
dev.off()

pdf("lme2ResBrut.pdf",h=5,w=5)
plot(modelo1, resid(.)~tempo2, abline=0, col="1",
      xlab="Tempo em dias", ylab="Residuos brutos")
dev.off()

```

ANEXO C Rotina no R utilizado para a obtenção das estimativas dos parâmetros do modelo não linear logístico e Gompertz.

```

#          MODELO DE GOMPERTZ

require(nlme)
require(lattice)

getInitial(COPA ~ SSgompertz(TEMPODIA, A~trat+bloc,
                             B~trat+bloc, C~trat+bloc),
           data=datadia[datadia$TRAT==1,])

reg1<- nls(COPA ~ A * exp(-B*C^TEMPODIA),
           start = list(A=180, B = 0.75, C = 0.99),
           data=datadia[datadia$TRAT==1,])
resul=summary(reg1)

### pdf("gompertz_t1.pdf")
plot(COPA~TEMPODIA, xlab="Tempo (dias)",
      ylab="Diâmetro da copa (cm)",
      data=datadia[datadia$TRAT==1,],main="Tratamento SI")
x<-seq(1,1100)
curve(resul$coef[1] * exp(-resul$coef[2]*resul$coef[3]^x),add=T)
### dev.off()

# Análise dos resíduos

shapiro.test(residuals(reg1, type="pearson"))

ks.test(residuals(reg1), "pnorm",
        mean(residuals(reg1)), sd(residuals(reg1)))

### pdf("goqqn_T1.pdf")
qqnorm(residuals(reg1),
       main="Normal Q-Q Plot - SI",
       xlab="Quantis Teóricos", ylab="Quantis Observados (Trat 01)")
qqline(residuals(reg1))
text(1.2, -20,
     "Shapiro-Wilk

```

```

        W = 0.9744
        p-valor = 0.1084")
### dev.off()

# Para os demais tratamentos procede-se de maneira semelhante.

#      GOMPERTZ PARA TODOS TRATAMENTOS

data <- read.table("data.txt", header=TRUE)
attach(data)

getInitial(COPA ~ SSgompertz(tempo, A~trat+bloc, B~trat+bloc,
                             C~trat+bloc), data=datadia)

reggomp<- nlsList(copa ~ A * exp(-B*C^tempo)|trat,
                 start = list(A=191, B = 0.89, C = 0.99),
                 data=data)
summary(reggomp)

# nlsList

### pdf("gomp_interv.pdf")
plot(intervals(reggomp))
### dev.off()

# Modelo logístico

getInitial(COPA ~ SSlogis(TEMPODIA, Assíntota~trat+bloc,
                          Inflexão~trat+bloc, Escala~trat+bloc),
           data=datadia[datadia$TRAT==1,])

rel<- nls(COPA ~ Assíntota/(1+exp((Inflexão-TEMPODIA)/Escala)),
         start = list(Assíntota=179, Inflexão = 9,
                     Escala = 109), data=datadia[datadia$TRAT==1,])

rel1=summary(rel)
names(rel)

### pdf("log_t1.pdf")
plot(COPA~TEMPODIA, xlab="Tempo (dias)",
     ylab="Diâmetro da copa (cm)",
     data=datadia[datadia$TRAT==1,],main="Tratamento SI")
x<-seq(1,1100)
curve(rel1$coef[1]/(1+exp((rel1$coef[2]-x)/rel1$coef[3])),add=T)
### dev.off()

# Análise dos resíduos

shapiro.test(residuals(rel, type="pearson"))

ks.test(residuals(rel), "pnorm",
        mean(residuals(rel)),
        sd(residuals(rel)))

```

```

### pdf("loqqn_T1.pdf")
qqnorm(residuals(re1),
      main="Normal Q-Q Plot - SI",
      xlab="Quantis Teóricos",
      ylab="Quantis Observados (Trat 01)")
qqline(residuals(re1))
text(1.2, -20,
     "Shapiro-Wilk
W = 0.973
p-valor = 0.0887")
### dev.off()

# Para os demais tratamentos procede-se de maneira semelhante.

# Modelo logístico para todos os tratamentos

getInitial(copa ~ SSlogis(tempo, Assíntota~trat+bloc, Inflexão~
      trat+bloc, Escala~trat+bloc), data=data)

reglog<- nlsList(copa ~ Assíntota/(1+exp((Inflexão-tempo)/Escala))
  |trat, start = list(Assíntota=190, Inflexão = 39, Escala = 123),
  data=data)
summary(reglog)

### pdf("log_interv.pdf")
plot(intervals(reglog))
### dev.off()

```

ANEXO D Rotina no R utilizado para a obtenção das estimativas dos parâmetros do modelo não linear com efeitos mistos.

```

# Bibliotecas utilizadas

require(lattice)
require(nlme)
require(gregmisc)
require(stats)

#plotando os dados

xyplot(copa-tempo|trat, groups=bloc, data=jair,
      type=c("p","smooth"))

# Rodando o modelo logístico com o conjunto de dados jair.copa3

logis12 <- nlsList(copa~SSlogis(tempo, Asym, xmid, scal),
  data=jair.copa3)
logis12

logis13 <- nlme(logis12)
logis13

```



```

getInitial(copa ~ SSlogis(tempo, Asym, xmid, scal),
           data=jair.copa3)

logis14 <- update(logis13, fixed=list(Asym~trat+bloc,
                                     xmid~trat+bloc, scal~trat+bloc),
                 start=c(190,0,0,0,0,0,0, 52,0,0,0,0,0,0,0,
                        118,0,0,0,0,0,0,0))

plot((logis14),col="1", xlab="Valores ajustados",
     ylab="Resíduos padronizados")

summary(logis14)

qqnorm(ranef(logis14)[,1])
qqline(ranef(logis14)[,1])

qqnorm(ranef(logis14)[,2])
qqline(ranef(logis14)[,2])

qqnorm(ranef(logis14)[,3])
qqline(ranef(logis14)[,3])

qqnorm(residuals(logis14))
qqline(residuals(logis14))

shapiro.test(ranef(logis14)[,1])
shapiro.test(ranef(logis14)[,2])
shapiro.test(ranef(logis14)[,3])
shapiro.test(residuals(logis14))

ks.test(residuals(logis11), "pnorm",
        mean(residuals(logis11)),
        sd(residuals(logis11)))

# Rodando o modelo com modelagem da variância

logis15 <- update(logis14, weights=varPower(0.18, form=~tempo2))

logis15
anova(logis15)
intervals(logis15)

anova(logis15, logis14)

plot(logis15, col="1")

qqnorm(ranef(logis15)[,1])
qqline(ranef(logis15)[,1])

qqnorm(ranef(logis15)[,2])
qqline(ranef(logis15)[,2])

qqnorm(ranef(logis15)[,3])
qqline(ranef(logis15)[,3])

```

```
qqnorm(residuals(logis15))
qqline(residuals(logis15))

shapiro.test(ranef(logis15)[,1])
shapiro.test(ranef(logis15)[,2])
shapiro.test(ranef(logis15)[,3])

ks.test(ranef(logis15)[,1], "pnorm",
        mean(ranef(logis15)[,1]),
        sd(ranef(logis15)[,1]))

ks.test(ranef(logis15)[,2], "pnorm",
        mean(ranef(logis15)[,2]),
        sd(ranef(logis15)[,2]))

ks.test(ranef(logis15)[,3], "pnorm",
        mean(ranef(logis15)[,3]),
        sd(ranef(logis15)[,3]))

shapiro.test(residuals(logis15))

ks.test(residuals(logis15), "pnorm",
        mean(residuals(logis15)),
        sd(residuals(logis15)))

# Para o modelo de gompertz segue-se o mesmo esquema de análise,
# utilizando-se o SSgompertz para encontrar os valores iniciais
# dos parâmetros.
```