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

**Análise de resíduos projetados em modelos não-lineares de
regressão: uma aplicação em crescimento de frutos**

Maria Cristina Martins

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

**Piracicaba
2013**

Maria Cristina Martins
Licenciada em Matemática

**Análise de resíduos projetados em modelos não-lineares de
regressão: uma aplicação em crescimento de frutos**

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Orientadora:
Prof^a Dr^a **TACIANA VILLELA SAVIAN**

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

**Piracicaba
2013**

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

Martins, Maria Cristina

Análise de resíduos projetados em modelos não-lineares de regressão: uma aplicação em crescimento de frutos / Maria Cristina Martins. - - versão revisada de acordo com a resolução CoPGr 6018 de 2011. -- Piracicaba, 2013.
90 p: il.

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

1. Curvas de crescimento 2. Logística - Modelos 3. Mínimos quadrados 4. Modelos matemáticos 5. Modelos não lineares 6. Pera I. Título

CDD 581.10184
M386a

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

DEDICATÓRIA

Aos meus pais,
Francisco Martins Bruno e
Maria Aparecida Liberali Martins

que sempre acreditaram na minha
capacidade e me apoiaram em todos os
momentos.

Com amor, DEDICO.

AGRADECIMENTOS

À Deus por sempre estar presente em cada momento de minha vida, mesmo quando eu não o via. Ele estava lá, me dando forças e me livrando das tribulações.

Aos meus pais, pelo amor, ensinamentos, apoio e por acreditarem na minha capacidade, deixando muitas vezes de realizar suas vontades em detrimento das minhas.

Aos meus irmãos que, sempre respeitam minhas escolhas e mais do que ninguém, entendem minhas dificuldades e esforços.

À minha orientadora de mestrado, que além de grande dedicação no direcionamento dos trabalhos, demonstra grande amizade. Também a cada docente e funcionário do LCE/ESALQ/USP que, com empenho, ajudam e instruem na construção dos meus conhecimentos.

Ao meu namorado que em todas as ocasiões, com amor, me ampara e auxilia.

À todos os familiares e amigos que mesmo de longe, de forma especial, torcem por mim. Ainda, aos colegas de turma com quem tive momentos sérios e de intenso estudo, mas também, momentos de descontração e lazer. Também aos amigos do LCE, que, em muitos momentos, se tornaram minha família, dando apoio e compartilhando momentos de alegria e tristeza.

Aos meus professores de graduação, com quem pude aprender muito. E a cada pessoa, que ao longo de minha jornada, contribuiu direta ou indiretamente à minha formação.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela bolsa de estudo.

O meu muito obrigada!

“Que os vossos esforços desafiem as impossibilidades, lembrai-vos de que as grandes coisas do homem foram conquistadas do que parecia impossível.”

Charles Chaplin

“Se você tem um sonho, tem que correr atrás dele, as pessoas não conseguem vencer e dizem que você também não vai vencer. Se você quer uma coisa, corra atrás, ponto.”

A Procura da Felicidade

SUMÁRIO

RESUMO	11
ABSTRACT	13
LISTA DE FIGURAS	15
LISTA DE TABELAS	17
1 INTRODUÇÃO	19
2 REVISÃO BIBLIOGRÁFICA	23
2.1 Crescimento e desenvolvimento vegetal	23
2.2 Modelos não lineares	24
2.2.1 Métodos de estimação dos parâmetros	27
2.2.1.1 Método dos mínimos quadrados ponderados	28
2.2.1.2 Método dos mínimos quadrados generalizados	29
2.2.2 Diagnóstico	30
2.3 Resíduo projetado	31
2.4 Estudo de simulação	33
3 MATERIAL E MÉTODOS	35
3.1 Banco de dados	35
3.2 Ajustes e pressuposições dos modelos	35
3.3 Método iterativo	37
3.4 Avaliadores da qualidade de ajuste	39
3.4.1 Diagnóstico do ajuste	39
3.4.2 Avaliadores no estudo de simulação	39
3.4.3 Seleção de modelos	40
3.5 Resíduo Projetado	41
3.6 Estudo de simulação	42
4 RESULTADOS E DISCUSSÃO	43
4.1 Ajuste do Modelo Logístico para a média dos dados	46
4.2 Ajuste do Modelo de Gompertz para a média dos dados	51
4.3 Algumas considerações a respeito dos ajustes dos modelos	55
4.4 Estudo de simulação	57
5 CONCLUSÃO	67
REFERÊNCIAS	69

APÊNDICES	75
ANEXOS	79

RESUMO

Análise de resíduos projetados em modelos não-lineares de regressão: uma aplicação em crescimento de frutos

Alguns aspectos devem ser considerados nos trabalhos com frutíferas como o ponto adequado de colheita e estratégias de conservação, sendo fundamental a análise do crescimento e desenvolvimento dos frutos. Vários tipos de modelos estatísticos podem descrever os processos de crescimento envolvidos no sistema de produção vegetal, sendo os modelos não lineares considerados mais adequados para esse tipo de estudo. Para que os resultados obtidos no ajuste do modelo sejam válidos é necessário verificar a qualidade de ajuste por meio de uma análise dos resíduos. A distribuição dos resíduos ordinários para os modelos de regressão não linear, é matematicamente trabalhosa e os critérios de diagnóstico são falhos, principalmente em pequenas amostras. Diante de tal fato, Cook e Tsai (1985) definiram o resíduo projetado. O comportamento dos resíduos projetados é melhor do que o dos resíduos ordinários, pois suas propriedades são mais próximas das correspondentes ao resíduo ordinário da regressão normal linear. O ganho é substancial se o teste para a medida de não linearidade for significativo. Por meio dos ajustes dos modelos logístico e de Gompertz aos dados de comprimento de peras foi realizada a análise dos resíduos ordinários e dos resíduos projetados, a fim de evidenciar suas vantagens. Os resíduos projetados se mostraram viáveis para diagnóstico dos modelos e descartaram possíveis tendências apresentadas pelos resíduos normalizados.

Palavras-chave: Critérios de diagnóstico; Modelo Logístico; Modelo de Gompertz

ABSTRACT

Residual Analysis designed in non-linear regression: an application to fruit growth

Some aspects must be considered in the works with fruit tree such as the appropriate point of harvest and conservation strategies, being fundamental the analysis of growth and fruit developments. Several types of statistical models can describe the growth processes involved in plant production system and the non-linear models are considered more suitable for this type of study. For the results obtained from the model fit to be valid is necessary to check the quality of fit through a residual analysis. The distribution of the ordinary residuals in the non-linear regression models is mathematically complicated and the diagnostic criteria are not precise, mainly in small samples. To overcome this fact, Cook and Tsai (1985) defined the projected residual. The behaviour of the projected residuals is better than the ordinary residuals since their properties are closer to those corresponding to the ordinary residuals for the normal linear regression. The gain is substantial if the test for non-linearity is significant. Through the fit of the logistic and Gompertz models to the pears length data, it was performed an analysis of the ordinary and projected residuals in order to demonstrate their advantages. The projected residuals showed more feasible for the diagnostic of the models and ruled out possible trends presented by normalized residuals.

Palavras-chave: Diagnostic criteria; Logistic model; Gompertz model

LISTA DE FIGURAS

Figura 1 - Fruto da pereira asiática (<i>Pyrus pyrifolia</i>) da cultivar <i>Shinsseiki</i> e a respectiva medida de seu comprimento (y)	35
Figura 2 - Esquema do modelo logístico com $\alpha = \phi_1$ assíntota do modelo, $\beta = \frac{\phi_2}{\phi_3}$ relacionado com o ponto de inflexão da curva (sem interpretação biológica), e $\gamma = \frac{1}{\phi_3}$ representa a taxa de crescimento	36
Figura 3 - Gráfico de dispersão dos dados de comprimento de peras ao longo do tempo	43
Figura 4 - Gráfico boxplot do comprimento de peras (mm) em função do tempo (dias)	45
Figura 5 - Ajuste do Modelo Logístico aos dados de comprimento de frutos de pereira e seu intervalo com 95% de confiança	47
Figura 6 - (a) Histogramas dos resíduos normalizados e gráfico de densidade de uma distribuição normal (b) Envelope simulado com 95% de confiança	48
Figura 7 - Resíduos do ajuste do modelo logístico aos dados de comprimento de peras, (a) Normalizado (b) Projetado	50
Figura 8 - Gráfico de índices das observações <i>versus</i> , (a) Distância de Cook e (b) elementos da diagonal da matriz de projeção	51
Figura 9 - Ajuste do Modelo de Gompertz aos dados de comprimento de frutos de pereira e seu intervalo com 95% de confiança	52
Figura 10 - (a) Histograma dos resíduos normalizados e gráfico de densidade e (b) Envelope simulado com 0,95 de probabilidade	53
Figura 11 - Resíduos do ajuste do modelo de Gompertz aos dados de comprimento de peras: (a) Normalizado (b) Projetado	55
Figura 12 - Ajuste dos modelos logístico e de Gompertz à média dos dados de comprimento de peras	56
Figura 13 - Histogramas das estimativas dos parâmetros para as 1000 amostras simuladas a partir do modelo logístico, (a) Alfa (b) Beta (c) Gama	60
Figura 14 - Histogramas das estimativas dos parâmetros para as 1000 amostras simuladas a partir do modelo de Gompertz, (a) Alfa (b) Beta (c) Gama	61
Figura 15 - Gráfico normal de probabilidade dos resíduos projetados calculados para o ajuste do modelo logístico aos dados reais, e envelope simulado	64
Figura 16 - Gráfico normal de probabilidade dos resíduos projetados calculados para o ajuste do modelo de Gompertz aos dados reais, e envelope simulado	65

LISTA DE TABELAS

Tabela 1 - Média dos dados de comprimento de peras ao longo dos dias de observação	46
Tabela 2 - Estimativas dos parâmetros do modelo logístico com os respectivos erros padrões . .	46
Tabela 3 - Estimativas dos parâmetros do modelo de Gompertz com os respectivos erros padrões	52
Tabela 4 - Critério de informação de Akaike (AIC), critério de informação bayesiano (BIC), medida de curvatura de efeito do parâmetro, medida de curvatura intrínseca do modelo e quadrado médio do resíduo	56
Tabela 5 - Amostras dos comprimentos dos frutos de pereira, simulados para o modelo logístico nas 11 idades	58
Tabela 6 - Amostras dos comprimentos dos frutos de pereira, simulados para o modelo de Gom- pertz nas 11 idades	58
Tabela 7 - Estimativas dos parâmetros dos modelos logístico e de Gompertz para as seis primei- ras amostras simuladas	59
Tabela 8 - Média, variância e erro quadrático médio das estimativas dos parâmetros do ajuste dos modelos para as 1000 amostras simuladas	59
Tabela 9 - Excesso de variância dos parâmetros obtidos dos ajustes do modelo para cada simulação	60
Tabela 10 - Porcentagem de viés dos parâmetros obtidos dos ajustes dos modelos	61
Tabela 11 - Resíduos projetados, dos modelos logístico e de Gompertz, para as primeiras seis amostras simuladas	62
Tabela 12 - Estatísticas para os resíduos projetados dos modelos logístico e de Gompertz, consi- derando as 1000 amostras simuladas	63
Tabela 13 - Estatísticas para as curvaturas dos modelos logístico e de Gompertz, considerando as 1000 amostras simuladas	64
Tabela 14 - Valores dos comprimentos de frutos de peras (mm) ao longo do tempo (dias)	81

1 INTRODUÇÃO

Nos trabalhos com frutíferas, um aspecto importante a ser considerado é o momento adequado de colheita, que, na maioria das vezes, é determinado de maneira subjetiva, considerando-se a coloração externa e o tamanho do fruto.

A análise do crescimento e desenvolvimento dos frutos é fundamental para descrever as mudanças ocorridas nas características físicas e nos componentes químicos, possibilitando estabelecer as bases para definir o ponto mais adequado de colheita e estratégias para sua conservação, visando ao aumento da vida útil pós-colheita.

É de grande importância o conhecimento do crescimento de plantas e frutos, medido por meio da produção de biomassa, que começa com o processo fotossintético (PEREIRA, 2007), e se acumula ao longo do tempo. Também é fundamental, para o manejo adequado, a detecção de fatores que influenciam o desenvolvimento das plantas, tais como: ocorrências de deficiências hídricas e nutricionais, competição de plantas daninhas, solos compactados ou mal drenados, entre outros (TERRA; MUNIZ; SAVIAN, 2010).

O estudo de curvas de crescimento é importante uma vez que pode ser realizado em situações em que a informação contida numa seqüência de pontos “tamanho-tempo”, do crescimento de uma espécie seja sintetizada num conjunto de parâmetros que tenham interpretação prática e preditiva.

Vários tipos de modelos estatísticos podem ser usados para estudar os processos de crescimento envolvidos no sistema de produção vegetal. Em diversas situações, as pesquisas que avaliam crescimento e desenvolvimento de frutos, utilizam modelos de regressão polinomial, e, em alguns casos não têm sido adequados para descrever os fenômenos. Estudos têm revelado que o crescimento e desenvolvimento de frutos possuem comportamento sigmoidal e podem, satisfatoriamente, ser descritos por modelos não lineares.

Em estatística, utiliza-se a regressão para estudar a relação existente entre duas variáveis X e Y por meio de uma função que explique essa relação. Os objetivos principais da regressão são: prever a variável resposta, selecionar variáveis, estimar parâmetros e inferir com relação aos mesmos.

Segundo Maroco (2007), o modelo clássico de regressão teve origem nos trabalhos de astronomia, elaborado por Gauss no período de 1809 a 1821. É a técnica mais adequada para modelar relações entre variáveis e prever o valor de uma variável

dependente (resposta) a partir de um conjunto de variáveis, conhecidas como preditoras ou regressoras. Como, existem situações em que não é desejável ou mesmo possível descrever o fenômeno por meio de um modelo de regressão linear, já que em muitas situações práticas, algumas de suas suposições, como a normalidade e a linearidade nos parâmetros, não são satisfeitas, alavancou-se o desenvolvimento de novas técnicas estatísticas para os modelos de regressão, dando início aos modelos de regressão não linear que se tornaram, em muitos casos, uma alternativa mais vantajosa.

Dentre as vantagens da utilização dessa classe de modelos destaca-se a obtenção de parâmetros que são biologicamente interpretáveis. Além disso, em diversas situações, necessita-se de menos parâmetros nos modelos não lineares do que nos lineares, o que simplifica e facilita a interpretação. Em alguns casos, a interpretação dos parâmetros é direta, mas, em outros, é necessário recorrer à dedução da equação para uma compreensão exata de sua relação com fenômenos fisiológicos subjacentes (FORNI, 2007).

Nos modelos classificados como lineares em relação aos parâmetros, o problema de estimação consiste em minimizar a soma de quadrados dos resíduos, e resolver um sistema de equações lineares com relação aos coeficientes de regressão desconhecidos e, por existir solução única para esse sistema de equações, é possível obter uma forma analítica para os estimadores dos parâmetros. Para os modelos não lineares, a estimação dos parâmetros também consiste em minimizar a soma de quadrados dos resíduos. Entretanto para esses modelos, o sistema de equações normais, não possui solução analítica e as estimativas dos parâmetros devem ser obtidas por meio de processos iterativos. Na literatura existem vários métodos numéricos iterativos, por exemplo, método de Newton, método Newton-Raphson, método Gauss-Newton (HARTLEY, 1961), algoritmo EM.

Mesmo após a escolha cuidadosa de um modelo e subsequente ajuste a um conjunto de observações é possível que se obtenha um resultado insatisfatório. Isso pode ocorrer em função de algum desvio sistemático entre as observações e os valores ajustados pela função escolhida ou, então, porque uma ou mais observações são discrepantes em relação às demais. Na prática, em geral, há uma combinação de diferentes tipos de falhas e a verificação da adequação de um modelo para um determinado conjunto de observações torna-se um processo difícil, porém, imprescindível para que os resultados obtidos sejam válidos.

No modelo clássico de regressão, os erros aleatórios são as diferenças entre os valores observados e aqueles esperados pelo modelo escolhido. Esses termos representam a variação natural dos dados, mas, também, podem ser interpretados como o efeito cumulativo de fatores que não foram considerados no modelo. Se as pressuposições do modelo de que os erros são normais, independentes e homogêneos são violadas, a análise resultante pode conduzir a resultados duvidosos.

Para os modelos de regressão não linear a distribuição dos resíduos é matematicamente trabalhosa, principalmente em pequenas amostras. Além disso, os resíduos, em geral, têm esperança diferente de zero e distribuição dependendo fortemente dos valores ajustados que pode conduzi-los a não refletirem exatamente a distribuição dos erros (CORDEIRO; PRUDENTE; DEMÉTRIO, 2009). Logo, nesse caso, os critérios de diagnóstico da regressão normal linear podem falhar.

Cook e Tsai (1985) definiram um novo resíduo, denominado resíduo projetado, que apesar de algebricamente mais complexo, tem propriedades mais próximas das propriedades correspondentes do resíduo ordinário da regressão linear.

Diante do exposto, os objetivos do estudo foram investigar o crescimento de frutos de pereira descrito por modelos de regressão não linear, e avaliar a qualidade do ajuste por meio de resíduos projetados.

O estudo permitiu concluir que a aproximação linear para os resíduos é aceitável uma vez que as medidas de curvatura sejam pequenas. Além disso, o modelo logístico foi o mais adequado para modelar a curva de crescimento de peras da cultivar Shinsseiki (*Pyrus pyrifolia*).

2 REVISÃO BIBLIOGRÁFICA

Para desenvolvimento deste trabalho foi realizada uma revisão sobre modelos de regressão não linear, avaliadores da qualidade do ajuste, estudo de simulação, e também, foi feito um estudo da utilização desses modelos no crescimento e desenvolvimento vegetal.

2.1 Crescimento e desenvolvimento vegetal

Faoro e Orth (2010) citam o ponto ideal de colheita como alternativa para garantir as exigências do mercado e aumentar o valor comercial dos frutos. Sua importância também se dá para definir estratégias de conservação do produto (TRENTO; SEPULCRI; MORIMOTO, 2011).

A presença de cultivares de frutíferas com características que atendam às expectativas do consumidor, como tamanho, forma, cor e sabor adequados, pequeno número de sementes, resistência a doenças e oferta de frutos nas épocas mais quentes do ano é fundamental para o crescimento do setor produtivo (SILVA et al., 2009).

Avila et al. (2009) estudaram a maturação fisiológica e a qualidade sanitária de frutos e sementes na determinação da época adequada de colheita de frutos de pitanga, e concluíram que o momento adequado da coleta das sementes ocorreu entre 56 e 63 dias após a antese.

Lombardi, Moraes e Camelatto (2000) citam que, no Brasil, entre as espécies de clima temperado, a pera é a terceira mais consumida, perdendo apenas para a maçã e o pêssego. Como alternativa para a diversificação na fruticultura da região nordeste do RS e no planalto catarinense, surge o cultivo da pereira asiática (*Pyrus pyrifolia*) que, possui sua exploração econômica reduzida sendo as regiões Sul e Sudeste onde ocorrem sua maior produção.

Segundo Argenta et al. (1995) as modificações mais notórias que ocorrem durante a maturação são a intensificação da cor da epiderme, a redução na firmeza da polpa, a liberação de substâncias voláteis, entre outros. Todas essas medidas são feitas de maneira subjetiva, sendo assim, há a necessidade de um estudo com relação ao crescimento e desenvolvimento de frutos.

A utilização de modelos não lineares em estudos envolvendo o desenvolvimento de frutos, possibilita ao pesquisador comparar o comportamento do crescimento

em diversas situações, visto que possuem parâmetros biologicamente interpretáveis.

Diversos autores utilizaram os modelos não lineares na descrição do desenvolvimento vegetal. Silveira et al. (2001) descreveram a influência da temperatura de molhamento e concentração do inóculo de fungos na incidência de podridões pós-colheita em frutos de tomateiro, em que são utilizados dois modelos não lineares a fim de descrever esse processo. Souza (2007) utilizaram o modelo não linear (sigmoidal) para descrever o crescimento de frutos de mangueira. Cunha e Volpe (2011) ajustaram os modelos de crescimento exponencial e o sigmoidal, para a descrição da massa seca em função da massa fresca do fruto do cafeeiro.

Iossi, Saber e Moro. (2007) utilizaram os modelos logístico e exponencial decrescente, em função das épocas de colheitas dos frutos, na obtenção da maturidade fisiológica das sementes de tamareira de jardim, e concluíram que o ponto de maturidade ocorre aos 138 dias após antese, mas o período de colheita dos frutos pode ser estendido até os 194 dias após a antese, sem perda da qualidade fisiológica das sementes.

Souza (2007) estudaram a extração de micro e macro nutrientes e o crescimento de frutos de mangueira, e utilizaram o modelo logístico para descrever seu crescimento ao longo do tempo. Em um estudo com tamareira-anã, Terra, Muniz e Savian (2010) utilizaram os modelos, logístico e de Gompertz.

2.2 Modelos não lineares

Uma importante tarefa em estatística é encontrar relações, se for o caso, que existem em um conjunto de variáveis quando pelo menos uma é aleatória, estando sujeito a flutuações aleatórias e, possivelmente, aos erros de medição (SEBER; WILD, 1988).

Em regressão, tipicamente, uma das variáveis é denominada variável resposta ou variável dependente, e, geralmente, denotada por Y . A outra variável, X , denominada variável regressora, ou variável independente, é usada, principalmente, para prever ou explicar o comportamento de Y (MONTGOMERY; PECK; VINING, 2006). Busca-se relacionar as variáveis Y e X e assim expressar essa relação por meio de uma função $f(\cdot)$, ou seja,

$$Y \simeq f(X).$$

O modelo a ser adotado depende de vários fatores, por exemplo, natureza das variáveis, relação linear entre os parâmetros ou não, homogeneidade de variâncias ou não, tipos de erros, independência dos erros entre outros. O que se busca com a utilização das técnicas de regressão, de acordo com Vieira (2006), é encontrar a expressão que melhor descreve o fenômeno em estudo.

Segundo Draper e Smith (1998), os modelos de regressão são classificados basicamente como: modelos lineares; modelos linearizáveis e modelos não lineares. O termo modelo de regressão linear pode ter dois significados. O primeiro uso refere-se a uma relação linear entre as variáveis, dependente e regressora, e o segundo refere-se à um modelo em que os parâmetros, quantidades a serem estimadas, aparecem de forma linear no modelo (RATKOWSKY, 1983). Os modelos de regressão linearizáveis são aqueles que podem ser convertidos em lineares por meio de uma transformação. Segundo Mazucheli e Achcar (2002), os modelos de regressão não linear são aqueles em que pelo menos uma das derivadas parciais da função em relação aos parâmetros depende de pelo menos um dos parâmetros do modelo, ou seja, pode-se dizer que um modelo é não linear quando ele é não linear nos parâmetros.

A não linearidade do modelo pode decorrer da forma como as perturbações aleatórias são nele integradas, por um lado, e, por outro, não são necessariamente idênticos o número de coeficientes de regressão e o número de regressores, ao contrário do que acontece no modelo linear de regressão (OLIVEIRA, 2004).

Na literatura, vários modelos não lineares são utilizados, e os mais comuns na descrição de curvas de crescimento são: o logístico (NELDER, 1961), o de Gompertz (LAIRD, 1965), o de Richards (RICHARDS, 1959), o de Von Bertalanffy (VON BERTALANFFY, 1957) e o de Brody (Brody, 1945), sendo os três últimos mais utilizados na descrição de curvas de crescimento animal.

Silva et al. (2011) utilizaram esses cinco modelos no estudo de curvas de crescimento de vacas de corte em diferentes tipos biológicos. Diante da variedade de modelos, alguns autores tentam encontrar o modelo mais indicado. McManus et al. (2003) recomendaram o modelo logístico em favor do de Richards e Brody. Sarmento et al. (2006) e Braccini Neto et al. (1996) mencionaram que o modelo Richards, apesar de apresentar bom ajuste, durante o processo iterativo, apresenta dificuldade quanto a convergência. Mazzini et al. (2005) comentaram a respeito da importância do estudo da heterogeneidade

de variâncias (heterocedasticidade) no processo do ajuste do modelo, podendo ocorrer a obtenção de estimativas viesadas caso ela seja ignorada.

Na área vegetal alguns autores têm trabalhado com modelos não lineares em detrimento dos lineares. Vieira Junior et al. (1999) estudaram o potencial total de água em sementes; Correa, Araújo e Afonso Júnior (2003), na determinação dos parâmetros de secagem de sementes de milho doce (*Zea mays* L.); Paixão et al. (2004) estimaram a infiltração de água no solo por meio de ajustes de funções não lineares e de modelos empíricos; Fideles Filho et al. (2005) avaliaram o índice de massa foliar e biomassa aérea da cultura do feijoeiro irrigado; Carlesso et al. (2007) compararam cinco modelos não lineares com o objetivo de determinar o modelo que melhor descreve a secagem por convecção em camada fina de sementes de maracujá-amarelo. Em um estudo de crescimento em vitro aplicado a diferentes extratos vegetais, Souza et al. (2011) utilizaram o modelo não linear Chapman-Richards.

De maneira geral, os modelos não lineares podem ser escritos da seguinte forma:

$$y_i = f(x_i, \boldsymbol{\theta}) + \varepsilon_i, i = 1, \dots, n, \quad (1)$$

em que y_i é a observação da variável dependente; $f(x_i, \boldsymbol{\theta})$ é a função resposta não linear em $\boldsymbol{\theta}$; x_i representa a variável independente; $\boldsymbol{\theta}$ é um vetor de parâmetros e ε_i é o erro experimental não observável diretamente, suposto independente e identicamente distribuído de uma distribuição normal com média 0 e variância σ^2 desconhecida.

A obtenção das estimativas dos parâmetros pode ser realizada de várias maneiras, dentre elas, por meio do método da máxima verossimilhança ou pelo método dos mínimos quadrados, objetiva minimizar a soma de quadrados residual na resolução do sistema de equações normais (GALLANT, 1987).

Para os modelos lineares, a estimação dos parâmetros é simples, obtida pela resolução de um sistema linear. Rawlings, Pantula e Dickey (1998) relataram que para os modelos de regressão não lineares, o sistema de equações normais não é resolvido facilmente, sendo necessários métodos iterativos na obtenção dessas estimativas.

2.2.1 Métodos de estimação dos parâmetros

A maioria dos algoritmos computacionais para obtenção das estimativas de mínimos quadrados $\hat{\theta}$ e os métodos inferenciais para modelos não lineares são baseados em métodos iterativos que consideram uma aproximação linear local para o modelo (BATES; WATTS, 1988).

Greene (2003) cita que o método de Gauss-Newton, também conhecido como método da linearização, usa uma expansão em série de Taylor para aproximar o modelo de regressão não linear em termos lineares, facilitando a aplicação dos mínimos quadrados ordinários para a estimação dos parâmetros.

Os modelos utilizados na descrição do crescimento vegetal são classificados como modelos de regressão não linear e apresentam algumas peculiaridades quanto ao procedimento de ajuste. Sob o ponto de vista da inferência estatística clássica, Santos e Ribeiro (2005) dizem que, para os modelos de regressão linear, os estimadores de mínimos quadrados apresentam propriedades ótimas entre qualquer outra classe de estimadores, isto é, são não viesados, normalmente distribuídos, e ainda possuem variância mínima. Essas propriedades são consideradas como as melhores que uma classe de estimadores podem apresentar (SEARLE, 1971). Já no caso de modelos não lineares, as propriedades somente são válidas assintoticamente, isto é, quando o tamanho da amostra é suficientemente grande. Em geral, em pequenas amostras essas propriedades não são satisfeitas.

Souza (1998) considera o modelo de regressão não linear em sua forma matricial da seguinte maneira:

$$\mathbf{Y} = f(\mathbf{X}, \boldsymbol{\theta}) + \boldsymbol{\varepsilon}, \quad (2)$$

em que:

$\mathbf{Y}$ é um vetor $(n \times 1)$ de observações da uma variável aleatória Y ;

$\mathbf{X}$ é uma matriz $(n \times p)$ de observações de p variáveis não aleatórias (natureza fixa);

$\boldsymbol{\theta}$ é um vetor $(p \times 1)$ de parâmetros desconhecidos, $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)^T$;

$f(\mathbf{X}, \boldsymbol{\theta})$ é um vetor $(n \times 1)$ de funções não lineares em $\boldsymbol{\theta}$, contínuas e diferenciáveis no espaço paramétrico $\boldsymbol{\Theta}$;

$\boldsymbol{\varepsilon}$ é um vetor $(n \times 1)$ de resíduos.

O método dos mínimos quadrados ordinários consiste da estimação dos parâmetros, por meio, da minimização da soma de quadrados dos resíduos (SQRes). Essa soma, associada ao modelo, é dada pela seguinte expressão:

$$\text{SQRes}(\boldsymbol{\theta}) = \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} = [\mathbf{Y} - f(\mathbf{X}, \boldsymbol{\theta})]^T [\mathbf{Y} - f(\mathbf{X}, \boldsymbol{\theta})].$$

Logo,

$$\begin{aligned} \text{SQRes}(\boldsymbol{\theta}) &= \{\mathbf{Y}^T - [f(\mathbf{X}, \boldsymbol{\theta})]^T\} [\mathbf{Y} - f(\mathbf{X}, \boldsymbol{\theta})] \\ &= \mathbf{Y}^T \mathbf{Y} - \mathbf{Y}^T f(\mathbf{X}, \boldsymbol{\theta}) - [f(\mathbf{X}, \boldsymbol{\theta})]^T \mathbf{Y} + [f(\mathbf{X}, \boldsymbol{\theta})]^T f(\mathbf{X}, \boldsymbol{\theta}) \\ &= \mathbf{Y}^T \mathbf{Y} - [f(\mathbf{X}, \boldsymbol{\theta})]^T \mathbf{Y} - [f(\mathbf{X}, \boldsymbol{\theta})]^T \mathbf{Y} + [f(\mathbf{X}, \boldsymbol{\theta})]^T f(\mathbf{X}, \boldsymbol{\theta}) \\ &= \mathbf{Y}^T \mathbf{Y} - 2[f(\mathbf{X}, \boldsymbol{\theta})]^T \mathbf{Y} + [f(\mathbf{X}, \boldsymbol{\theta})]^T f(\mathbf{X}, \boldsymbol{\theta}). \end{aligned}$$

Para encontrar o estimador de mínimos quadrados, é necessário fazer a diferenciação de $\text{SQRes}(\boldsymbol{\theta})$ com relação a cada um dos parâmetros, isto é,

$$\frac{\partial \text{SQRes}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = -2 \left[\frac{\partial f(\mathbf{X}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right]^T [\mathbf{Y} - f(\mathbf{X}, \boldsymbol{\theta})]$$

tal que,

$\frac{\partial f(\mathbf{X}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ é uma matriz $(n \times k)$ de elementos genéricos $\frac{\partial f(X_i, \boldsymbol{\theta})}{\partial \theta_j}$;

$f(X_i, \boldsymbol{\theta})$ é a i -ésima linha do vetor $f(\mathbf{X}, \boldsymbol{\theta})$, com $i = 1, \dots, n$ e

θ_j é o j -ésimo elemento do vetor de parâmetros $\boldsymbol{\theta}$, com $j = 1, \dots, k$.

A solução do sistema de equações normais dado por $\left[\frac{\partial f(\mathbf{X}, \hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}} \right] [\mathbf{Y} - f(\mathbf{X}, \hat{\boldsymbol{\theta}})] = 0$ pode ser realizada por meio do algoritmo de Gauss-Newton, como método iterativo, e seu sucesso na estimação dos parâmetros depende da escolha apropriada da função resposta e de bons valores iniciais. O processo de escolha é um procedimento decidido pelo pesquisador.

Ao trabalhar com modelos de regressão, assume-se que os erros sejam normais, independentes e homogêneos. Entretanto, quando pelo menos um desses itens é violado são necessárias algumas considerações.

2.2.1.1 Método dos mínimos quadrados ponderados

O estimador de mínimos quadrados penaliza, igualmente, os erros para cada observação. Em algumas situações práticas, poderá ser desejado dar pesos diferentes para

alguns valores de erros. Segundo Hoffmann e Vieira (1998), em presença de heterogeneidade de variâncias, o método dos quadrados mínimos ponderados é mais adequado em relação ao método dos mínimos quadrados por fornecer estimadores não tendenciosos e de variância mínima. É assumido que $\varepsilon \sim N(0, \mathbf{V}\sigma^2)$, em que $\mathbf{V}$ é uma matriz diagonal, positiva definida, que representa as variâncias associadas a cada ε_i , em que, seus elementos são os pesos V_j , com $E(\varepsilon) = 0$ e

$$(\varepsilon^T \varepsilon) = \mathbf{V}\sigma^2 = \begin{bmatrix} V_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & V_n \end{bmatrix} \sigma^2.$$

O fato de serem nulos os elementos fora da diagonal principal da matriz $\mathbf{V}$ significa que é assumida a pressuposição de independência das várias observações, isto é, $E(\varepsilon_j, \varepsilon_h) = 0$ para $j \neq h$. Definindo-se uma matriz diagonal $\mathbf{\Lambda}$, cujos elementos são dados por $\lambda_i = \frac{1}{\sqrt{V_j}}$, tem-se que: $\mathbf{\Lambda}\mathbf{\Lambda} = \mathbf{V}^{-1}$ e $\mathbf{V} = \mathbf{\Lambda}^{-1}\mathbf{\Lambda}^{-1}$. Pré multiplicando cada um dos termos de um modelo de regressão não linear, $\mathbf{Y} = f(\mathbf{X}, \boldsymbol{\theta}) + \varepsilon$, por $\mathbf{\Lambda}$, obtém-se o modelo: $\mathbf{\Lambda Y} = \mathbf{\Lambda}f(\mathbf{X}, \boldsymbol{\theta}) + \mathbf{\Lambda}\varepsilon$ com resíduos homocedásticos. O método dos quadrados mínimos ponderados fornece o sistema de equações normais (SEN) dado por: $\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X} \hat{\boldsymbol{\theta}} = \mathbf{X}^T \mathbf{V}^{-1} \mathbf{Y}$.

2.2.1.2 Método dos mínimos quadrados generalizados

Para os modelos de regressão mais simples, assume-se que os erros sejam independentes, o que não é apropriado quando se trabalha com séries cronológicas de dados, em que o erro da observação relativa a um período está correlacionado com o erro da observação anterior (HOFFMANN; VIEIRA, 1998). Crocci (1984) expõe que o método dos mínimos quadrados generalizados é mais eficiente do que o método dos mínimos quadrados ponderados, ou seja, o estimador da variância do primeiro está mais próximo de um valor ótimo, não sendo necessário procurar um estimador de menor variância.

A característica geral da dependência dos resíduos é a de existir uma variação sistemática dos valores em observações sucessivas. Quando isso ocorre, diz-se que os resíduos são autocorrelacionados. Além da autocorrelação residual pode ocorrer heterogeneidade de variâncias. Neste caso, é mais indicada a utilização do método dos quadrados mínimos generalizados em detrimento do método dos quadrados mínimos ponderados e

ordinários.

Seja $\boldsymbol{\varepsilon} \sim N(0, \mathbf{W}\sigma^2)$, em que $\mathbf{W}$ é uma matriz simétrica, positiva definida, que representa as variâncias e covariâncias dos erros e, admite-se que os erros sejam autocorrelacionados na forma de um processo autoregressivo estacionário de ordem p AR(p), em que, $\varepsilon_i = \phi_1\varepsilon_{i-1} + \phi_2\varepsilon_{i-2} + \dots + \phi_p\varepsilon_{i-p} + u_i$. Em particular, para o caso de primeira ordem AR(1), tem-se que $\varepsilon_i = \phi_1\varepsilon_{i-1} + u_i$ e portanto, $E(u_i) = 0$, $E(u_i^2) = \sigma_u^2$ e $E(u_i, u_{i-h}) = 0$ para $h \neq 0$. Nessas condições, $Var(\varepsilon) = \frac{\sigma_u^2}{1-\phi_1^2}$ e $Cov(\varepsilon_i, \varepsilon_h) = \frac{\sigma_u^2}{1-\phi_1^2}\phi_1^h$.

Segundo e Draper e Smith (1998) de maneira análoga ao método dos quadrados mínimos ponderados, o vetor de estimativas dos parâmetros é dado por: $\hat{\boldsymbol{\theta}} = (\mathbf{X}^T\mathbf{W}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}^{-1}\mathbf{Y}$. Outros métodos possíveis para incorporar correlações, segundo Morettin e Toloi (2004), são o de médias móveis (MA), autorregressivo e médias móveis (ARMA) e modelos mistos.

Nos estudos de curvas de crescimento que envolvem medidas repetidas no tempo, tomadas num mesmo indivíduo, é comum ocorrer heterogeneidade de variâncias o que pode exigir algum tipo de ponderação. Mazzini et al. (2003, 2005) e Silva, Aquino e Oliveira (2002) utilizaram a ponderação pelo inverso da variância dos pesos, juntamente com a estrutura de erros autorregressivos, verificando redução na estimativa da variância das estimativas dos parâmetros para as funções de crescimento animal.

Em estudos de modelos não lineares que descrevem o crescimento vegetal é bastante razoável incorporar a autocorrelação, tendo em vista que as medidas de crescimento são tomadas em uma mesma unidade experimental, planta ou fruto, ao longo do tempo, estando, portanto, provavelmente correlacionadas.

2.2.2 Diagnóstico

Métodos de diagnósticos são utilizados para acessar adequadamente pressupostos subjacentes do modelo e para identificar características inesperadas dos dados. Uma diversidade de métodos de diagnósticos está disponível para auxiliar nas análises e, em geral, são formulados para modelos de regressão linear.

Os resíduos usuais (ordinários, normalizados, estudentizados, entre outros) formam a base mais frequentemente utilizada para a construção de métodos de diagnósticos de regressão não linear, e sua definição se enquadra no âmbito geral fornecido por Cox e Snell (1968). Vários outros tipos de resíduos têm sido sugeridos para utilização

em regressão linear, e alguns deles podem ser facilmente estendidos para a regressão não linear. Entretanto, o desenvolvimento desses métodos é dependente de um estudo intensivo do comportamento exato dos resíduos, principalmente quando se trata de pequenas amostras.

Algumas investigações são necessárias para determinar a melhor abordagem em regressão não linear, já que, tipos alternativos de resíduos podem apresentar deficiências assim como os resíduos usuais. Diante disso, Bates e Watts (1980) associaram duas medidas de curvaturas que pudessem descrever o quão distante está um modelo não linear das propriedades de um modelo linear na vizinhança de $\hat{\theta}$.

As estatísticas que indicam a adequação de uma aproximação linear e seus efeitos nas inferências são chamadas medidas de não linearidade, e o que Bates e Watts (1980) mostraram foi que a não linearidade de um modelo pode ser separada em dois componentes: (i) uma não linearidade intrínseca associada com a curvatura no local da solução e (ii) uma de efeito do parâmetro, associada com o fato de que as projeções não são, em geral, nem retas, nem paralelas, nem equidistantes (SOUZA et al., 2010).

Bates e Watts (1980) sugeriram a utilização das curvaturas máximas no local da solução esperada, $\eta(\hat{\theta})$. Se a maior das curvaturas for considerada pequena, por transitividade, todas as outras também o serão. Os autores consideraram todas as curvaturas em todas as direções possíveis no local da solução $\eta(\hat{\theta})$ (SOUZA et al., 2010). Um problema levantado pelos autores foi a não existência de fórmulas analíticas para computar tais curvaturas, sendo sugeridos algoritmos alternativos que viabilizem esses resultados.

Foi sugerida por Bates e Watts (1980), a curvatura para o estudo da não linearidade, que apresentaram as medidas RMS (*Root Mean Square*) para a realização dos cálculos e elaboraram o programa computacional para a resolução dessas operações. Essas medidas, além de avaliadoras da qualidade do ajuste, serviram como critério para a utilização dos resíduos projetados descritos adiante.

2.3 Resíduo projetado

De acordo com Cook e Tsai (1985) pode-se empregar uma expansão em série de Taylor para aproximar e investigar o comportamento dos resíduos ordinários na regressão não linear. Entretanto, em alguns casos, esses resíduos podem produzir

resultados enganosos quando usados em métodos de diagnósticos análogos aos da regressão linear, não refletindo de forma correta a distribuição dos erros. Sugere-se, então um novo tipo de resíduo que supera a potencial deficiência dos resíduos ordinários.

A obtenção de derivadas de expressões para a esperança e a variância do vetor de resíduos usuais, e discussões do porquê desses resíduos não poderem ser uma base adequada para métodos de diagnósticos são de grande importância. Cook e Tsai (1985) apresentaram um novo tipo de resíduo que pode superar muitas das deficiências dos resíduos usuais.

A seguir está apresentado parte do procedimento que foi realizado por Cook e Tsai (1985) na definição desses resíduos. Para denotar o subespaço gerado pelas colunas de $\tilde{\mathbf{X}}$ é utilizado $C(\tilde{\mathbf{X}})$ e o complemento ortogonal é dado por $\overline{C}(\tilde{\mathbf{X}})$.

Seja $\hat{\boldsymbol{\theta}}$ o estimador de máxima verossimilhança de $\boldsymbol{\theta}$, $f_i = f(x_i, \boldsymbol{\theta})$ e $\tilde{\mathbf{X}}$ denotam uma matriz $n \times p$ com elementos $f_i^r = \frac{\partial f_i}{\partial \boldsymbol{\theta}_r}$ ($i = 1, \dots, n; r = 1, \dots, p$), em que n é o número de observações e p a quantidade de parâmetros do modelo. A menos que indicado, todas as derivadas são avaliadas no valor do parâmetro $\boldsymbol{\theta}^*$. Várias expansões quadráticas envolvem matrizes $\mathbf{W}_i$ $p \times p$ ($i = 1, \dots, n$); os elementos de $\mathbf{W}_i$ são $f_i^{rs} = \frac{\partial^2 f_i}{\partial \boldsymbol{\theta}_r \partial \boldsymbol{\theta}_s}$ ($r, s = 1, \dots, p$). Essas matrizes são utilizadas para formar faces de uma matriz $\mathbf{W}$ ($n \times p$) $\times p$ descrita por Bates e Watts (1980). A (k, j) -ésima coluna de $\mathbf{W}$ é a segunda derivada para o (k, j) -ésimo termo do vetor $\mathbf{w}_{kj}$ com elementos f_i^{kj} ($i = 1, \dots, n$).

É apresentada por Cook e Tsai (1985) uma aproximação para o resíduo ordinário $\mathbf{r}$, dado da seguinte maneira:

$$\mathbf{r} \cong (\mathbf{I} - \mathbf{H})\boldsymbol{\varepsilon} - \tilde{\mathbf{X}} \sum_{i=1}^n \mathbf{r}_i \mathbf{W}_i \boldsymbol{\delta} - \frac{1}{2}(\mathbf{I} - \mathbf{H})\boldsymbol{\delta}^T \mathbf{W} \boldsymbol{\delta}.$$

tal que, $\boldsymbol{\delta} = \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}$, $\boldsymbol{\delta}^T \mathbf{W} \boldsymbol{\delta}$ representa $\boldsymbol{\delta}^T \mathbf{W}_i \boldsymbol{\delta}$ para produzir a i -ésima componente de $\mathbf{r}$.

A matriz $\mathbf{W}$ apresenta p faces, e cada uma de suas faces é uma matriz $n \times p$, em que, n é o número de observações e p o de parâmetros. Projeta-se cada coluna de derivadas de segunda ordem de $\mathbf{W}$ sobre $\overline{C}(\tilde{\mathbf{X}})$, tem-se $\mathbf{H}_1 = \mathbf{S}(\mathbf{S}^T \mathbf{S})^{-1} \mathbf{S}$ que é o operador de projeção ortogonal do espaço gerado por sobre essa projeção. $\mathbf{S} = (\mathbf{I} - \mathbf{H})\mathbf{T}$, em que, $\mathbf{T}$ é uma matrix $n \times q$ obtida pelos vetores não nulos, de derivadas de segunda ordem e $\mathbf{H} = \tilde{\mathbf{X}}(\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^T$.

Apresenta-se $\mathbf{V}$ uma matriz $n \times (p + q)$ definida como $\mathbf{V} = (\tilde{\mathbf{X}}, \mathbf{S})$ e

$\mathbf{H}_2 = \mathbf{V}(\mathbf{V}^T \mathbf{V})^{-1} \mathbf{V}^T$ o operador de projeção ortogonal em $C(\mathbf{V})$.

Utilizando as matrizes descritas anteriormente, Cook e Tsai (1985) definem o resíduo projetado dado por

$$(\mathbf{I} - \mathbf{H}_2)\mathbf{r} = (\mathbf{I} - \mathbf{H})\boldsymbol{\varepsilon} - (\mathbf{I} - \mathbf{H}_1)\boldsymbol{\varepsilon}. \quad (3)$$

O primeiro termo de (3) é a aproximação linear para o resíduo ordinário $\mathbf{r}$, enquanto o segundo termo retrata a perda de informação necessária para remover os componentes não lineares. Se $q = \text{posto}(\mathbf{H}_1)$ for pequeno comparado a $n - p$, essa perda será pequena. Livre disso, se as medidas de não linearidade intrínseca e de efeito do parâmetro forem significativas, ou seja, maiores do que $2F^{-\frac{1}{2}}$, o ganho será considerável. Além disso, mostra-se que

$$E\{(\mathbf{I} - \mathbf{H}_2)\mathbf{r}\} = \mathbf{0}; \text{Var}\{(\mathbf{I} - \mathbf{H}_2)\mathbf{r}\} = \sigma^2 2(\mathbf{I} - \mathbf{H}_2)$$

e

$$E\{\mathbf{r}^T(\mathbf{I} - \mathbf{H}_2)\mathbf{r}\} = \sigma^2 \text{tr}(\mathbf{I} - \mathbf{H}_2).$$

Em geral, os resíduos projetados apresentam maior qualidade de diagnóstico do que os ordinários, e, algumas das técnicas de diagnóstico utilizadas na regressão linear, são válidas para esses resíduos.

Cordeiro, Prudente e Demétrio (2009), empregaram esses resíduos a fim de diagnosticar o ajuste do modelo exponencial utilizado para explicar a fração de cloro disponível em um produto manufaturado.

2.4 Estudo de simulação

Outra forma de avaliar o grau de não linearidade do modelo é por meio do estudo de simulação. Esse procedimento poderá indicar se o grau de não linearidade é pequeno o suficiente para justificar a utilização de resultados usuais da teoria dos modelos lineares como aproximação dos não lineares (PERRE et al., 1989).

Se o modelo é linear nos parâmetros, não há viés dos estimadores de mínimos quadrados, e, eles possuem variância mínima. Entretanto, quando o modelo é não linear, essas propriedades são válidas apenas assintoticamente, sendo necessárias amostras muito

grandes. Com o intuito de avaliar esse problema, utiliza-se a técnica da simulação, que, de acordo com Silva et al. (2009), Silva, Bergamasco e Vendite (2002) é uma técnica matemática destinada a resolver problemas cuja solução analítica é difícil.

Quando o estimador de mínimos quadrados para modelos não lineares possuir viés pequeno, variância mínima e possuir uma distribuição próxima da normal, diz-se que apresenta um comportamento próximo do linear.

Para um estudo de simulação, a distribuição dos resíduos e_i é especificada e $e_1, \dots, e_n$ são gerados por uma rotina. Um valor para o estimador é fixado produzindo-se um conjunto de n observações Y_i , sendo gerados N conjuntos. Se ocorre convergência do algoritmo de estimação, haverá N valores de estimativas e a partir dessas estimativas são analisadas as distribuições marginais de cada um dos estimadores de $\hat{\theta}$.

Ao serem efetuados os cálculos do excesso de variância e da porcentagem de viés, se os resultados não forem significativos e a distribuição dos estimadores próxima da normal, existem evidências de que os estimadores de mínimos quadrados terão um comportamento aproximadamente linear.

3 MATERIAL E MÉTODOS

3.1 Banco de dados

No presente estudo, foram adaptados os dados de Lombardi, Moraes e Camelatto (2000) referentes ao comprimento de frutos de pereira asiática, da cultivar *Shinsseiki*. O experimento foi conduzido no município de Canguçu (RS) da seguinte forma: foram selecionadas 10 plantas ao acaso, na floração plena. De cada planta foram colhidos 2 frutos dos ramos da região externa e mediana, totalizando 20 frutos a cada época de avaliação. A primeira medição do comprimento foi realizada aos 14 dias e estendeu-se até 154 dias após a floração, a cada duas semanas, até que o fruto estivesse totalmente desenvolvido e maduro, totalizando 11 observações.

Figura 1 - Fruto da pereira asiática (*Pyrus pyrifolia*) da cultivar *Shinsseiki* e a respectiva medida de seu comprimento (y)

3.2 Ajustes e pressuposições dos modelos

Para os dados de comprimento de frutos de pereira, foram ajustados os modelos não lineares logístico e de Gompertz. Esses modelos são descritos pelas seguintes expressões:

Modelo logístico:

$$y_i = \frac{\alpha}{[1 + \exp(\beta - \gamma x_i)]} + \varepsilon_i \quad (4)$$

Modelo de Gompertz:

$$y_i = \alpha \exp(-\exp(\beta - \gamma x_i)) + \varepsilon_i \quad (5)$$

em que y_i é a medida do comprimento do fruto (mm); x_i uma medida de tempo (dias); α representa, em ambos modelos, uma assíntota horizontal superior, ou seja, o valor de estabilização da variável dependente em relação ao tempo; o parâmetro γ a taxa de expansão ou velocidade de crescimento e ε_i é o erro experimental com média zero e variância σ^2 . De forma geral, não há uma interpretação biológica direta para o parâmetro β , sendo este, o parâmetro de locação relacionado ao ponto de inflexão. Essas medidas foram descritas por Pinheiro e Bates (2000) que apresentaram o modelo logístico e a interpretação de seus parâmetros. O modelo de Gompertz, diferentemente do logístico, não apresenta simetria em relação ao ponto de inflexão, entretanto, a interpretação de seus parâmetros é análoga.

Figura 2 - Esquema do modelo logístico com $\alpha = \phi_1$ assíntota do modelo, $\beta = \frac{\phi_2}{\phi_3}$ relacionado com o ponto de inflexão da curva (sem interpretação biológica), e $\gamma = \frac{1}{\phi_3}$ representa a taxa de crescimento

Para verificar a existência de heterogeneidade de variâncias, foram calculadas, a cada tempo, a variância da variável resposta, em cada uma das idades dos frutos estudados, possibilitando a obtenção do $F_{máximo}$ de Hartley (HARTLEY, 1950) dado pelo

quociente entre a maior e menor variância encontrada, em que a não significância mostra que existem evidências de homogeneidade de variâncias. Além disso, foi realizado o teste de Bartlett, em que a hipótese nula é a igualdade de variâncias ao longo do tempo.

Segundo Morettin e Toloi (2004), para a constatação da presença de autocorrelação residual os modelos autorregressivos e de médias móveis podem ser ajustados. Diante disso, foram verificados os modelos ARMA(p, q) de ordem 1 e 2 para os parâmetros p e q , em que, AR(p) representa um processo autorregressivo e MA(q) um processo de médias móveis. Para análise da normalidade dos resíduos foi utilizado o teste de Shapiro-Wilk (SHAPIRO;WILK, 1965).

3.3 Método iterativo

Como padrão, foi utilizado o método de Gauss-Newton, na forma original, que consiste no desenvolvimento em série de Taylor até o termo de primeira ordem da função, $f(\mathbf{X}_i, \boldsymbol{\theta})$ em torno do ponto $\boldsymbol{\theta}^0$.

O primeiro passo consiste em atribuir valores iniciais para o vetor de parâmetros $\boldsymbol{\theta}^0 = (\alpha^0, \beta^0, \gamma^0)$. Esses valores podem ser obtidos de estudos anteriores, conhecimentos teóricos, por inspeção de uma grade de valores que minimizem a soma de quadrados residuais ou, algumas vezes, graficamente.

Os valores iniciais foram obtidos por meio da linearização dos modelos logístico e de Gompertz. Esses modelos são intrinsecamente não lineares, ou seja, não são linearizáveis. Entretanto, como o parâmetro α representa uma assíntota horizontal em ambos modelos, é atribuído a ele o valor máximo dos dados, e assim, é feita a linearização desconsiderando a estrutura aditiva dos erros.

Para o modelo logístico tem-se

$$\begin{aligned} y_i &= \frac{\alpha}{1 + \exp(\beta - \gamma x_i)} \\ \frac{\alpha}{y_i} &= 1 + \exp(\beta - \gamma x_i) \\ \frac{\alpha}{y_i} - 1 &= \exp(\beta - \gamma x_i) \\ \ln\left(\frac{\alpha}{y_i} - 1\right) &= \beta - \gamma x_i \\ Y^* &= \beta - \gamma x_i \end{aligned}$$

$$\text{com, } Y^* = \ln\left(\frac{\alpha}{y_i} - 1\right).$$

Para o modelo de Gompertz tem-se

$$\begin{aligned} y_i &= \alpha \exp[-\exp(\beta - \gamma x_i)] \\ \frac{y_i}{\alpha} &= \exp[-\exp(\beta - \gamma x_i)] \\ -\ln\left(\frac{y_i}{\alpha}\right) &= \exp(\beta - \gamma x_i) \\ \ln\left[\ln\left(\frac{\alpha}{y_i}\right)\right] &= \beta - \gamma x_i \\ Y^{**} &= \beta - \gamma x_i \end{aligned}$$

com, $Y^{**} = \ln\left[\ln\left(\frac{\alpha}{y_i}\right)\right]$.

As estimativas dos parâmetros, obtidas por meio de uma regressão linear, foram atribuídas como valores iniciais dos parâmetros dos modelos não lineares, necessário ao processo iterativo. Foi feita uma aproximação da função esperada, $f(X_i, \theta)$, por termos lineares da expansão em série de Taylor, em uma vizinhança de $\theta^0 \in \Theta$, da seguinte forma:

$$f(\mathbf{X}, \theta) \approx f(\mathbf{x}, \theta) + \mathbf{Z}^0(\theta - \theta^0) \quad (6)$$

$$f(\mathbf{X}, \theta) \approx f(\mathbf{x}, \theta) + \mathbf{Z}^0\theta - \mathbf{Z}^0\theta^0$$

em que, $\mathbf{Z}^0$ é a matriz de derivadas parciais, $\frac{\partial f(\mathbf{X}, \theta^0)}{\partial \theta^0}$, do modelo avaliada em $\theta = \theta^0$.

Dessa maneira, o modelo não linear $\mathbf{Y} = f(\mathbf{X}, \theta) + \varepsilon$ pode ser reescrito como:

$$\begin{aligned} \mathbf{Y} &= f(\mathbf{X}, \theta) + \varepsilon \\ \mathbf{Y} &= f(\mathbf{X}, \theta) + \mathbf{Z}^0(\theta - \theta^0) \\ Y - f(\mathbf{X}, \theta) &= \mathbf{Z}^0(\varphi^0) \end{aligned}$$

e assim,

$$\mathbf{Y}^0 = \mathbf{Z}^0\varphi^0 + \varepsilon. \quad (7)$$

O propósito de ajustar o modelo de regressão linear é estimar os coeficientes de regressão φ^0 pelo método dos mínimos quadrados ordinários e usar essas estimativas para corrigir as estimativas iniciais dos parâmetros de regressão como indicado a seguir

$$\hat{\varphi}^0 = (\mathbf{Z}^{0'}\mathbf{Z}^0)^{-1}\mathbf{Z}^{0'}\mathbf{Y}^0. \quad (8)$$

Ao término da primeira iteração, a atualização do vetor de estimativas dos parâmetros é dada por: $\boldsymbol{\theta}^1 = \boldsymbol{\theta}^0 + \hat{\boldsymbol{\varphi}}^0$. Se o algoritmo estiver na direção correta, a soma quadrática residual de $\boldsymbol{\theta}^1$ será menor que a de $\boldsymbol{\theta}^0$. As iterações ocorrem sucessivamente até que a soma de quadrados de resíduos seja mínima.

3.4 Avaliadores da qualidade de ajuste

3.4.1 Diagnóstico do ajuste

Para verificar a adequação do ajuste dos modelos aos dados de comprimento dos frutos foram utilizadas as técnicas de diagnóstico utilizando o resíduo projetado proposto por Cook e Tsai (1985) e avaliadores da qualidade do ajuste dados a seguir.

Distância de Cook

Trata-se do afastamento do vetor de estimativas provocado pela retirada da i -ésima observação. Seu cálculo pode ser realizado da seguinte forma:

$$D_{(i)} = \frac{h_i}{1 - h_i} \frac{1}{p} rsi_i^2 \quad (9)$$

em que: h_i é o i -ésimo elemento da diagonal da matriz de projeção $\mathbf{H}$, rsi_i são os resíduos estudantizados internamente e p o número de parâmetros.

Leverage

É a distância de uma observação em relação às demais, representada por

$$h_i = \frac{1}{n} + \frac{x_i^2}{\sum_{i=1}^n x_i^2} \quad (10)$$

em que h_i são os elementos da diagonal de $\mathbf{H}$. O critério para análise desses valores é $\frac{2p}{n}$.

3.4.2 Avaliadores no estudo de simulação

Para o estudo de simulação de dados, Perre et al. (1989) citam duas medidas na análise da distribuição marginal dos estimadores dos parâmetros: o excesso de variância (Ev) e porcentagem de viés (%v)

Excesso de variância (%)

É a porcentagem na qual a variância amostral excedeu a variância assintótica do parâmetro, dada por:

$$Ev = \frac{\tilde{s}_i^2 - \hat{s}_i^2}{\hat{s}_i^2} 100, \quad (11)$$

em que $\hat{s}_i^2$ é o i -ésimo elemento da matriz de covariâncias, $\tilde{s}_i^2 = \frac{1}{N-1} \sum_{j=1}^N (\tilde{\theta}_{ij} - \hat{\theta}_i)^2$ é a variância amostral e N o número de simulações. A estatística $Z_i = \frac{(n_{sim}-1)\hat{s}_i^2}{cov(s_i, s_{i'})}$ é utilizada na verificação da significância do excesso de variância.

Porcentagem de viés

A porcentagem de viés de θ_i é dada pela diferença entre a média amostral, $\tilde{\theta}_i$, e o verdadeiro valor do parâmetro, θ_i , ou seja,

$$(\%)viés = \frac{\tilde{\theta}_i - \theta_i}{\theta_i} 100. \quad (12)$$

Foi utilizada a estatística do teste $Z_i = \frac{\tilde{\theta}_i - \theta_i}{\sqrt{\frac{\hat{s}_i^2}{N}}}$ de uma distribuição normal, com média 0 e variância 1, a fim de verificar se o teste para o viés foi ou não significativo.

3.4.3 Seleção de modelos

Detectado o bom ajuste do modelo aos dados, a seleção do modelo mais adequado para explicar o crescimento dos frutos teve como base a interpretação biológica dos parâmetros e precisão dos ajustes, e obedeceu aos seguintes critérios:

Critério de informação de Akaike (AIC)

Permite utilizar o princípio da parcimônia na escolha do melhor modelo, ou seja, de acordo com esse critério nem sempre o modelo mais parametrizado é melhor (Burnham e Anderson, 2004). Menores valores de AIC refletem um melhor ajuste (Akaike, 1974).

$$AIC = -2\log L + 2p$$

Critério de informação bayesiano (BIC)

Assim como o AIC, também leva em conta o grau de parametrização do modelo, e da mesma forma, quanto menor for o valor de BIC (SCHWARZ, 1978), melhor será o ajuste do modelo. Este último é mais rigoroso, já que penaliza mais o modelo em estudo.

$$BIC = -2\log L + p \log n$$

Percentual de Convergência (C%)

É usado quando se têm ajustes individuais, ou seja, o interesse reside em ajustar uma curva para cada indivíduo. Sendo assim, é possível observar para os modelos estudados, a quantidade de ajustes que convergiram.

Medidas de curvatura

Bates e Watts (1980,1988) sugeriram uma medida simplificadora da não linearidade por meio da curvatura. Os mesmos autores apresentam a medida RMS (*Root Mean Square*), dada pela relação:

$$C^2 = \frac{1}{A} \int_{||\mathbf{u}||=1} \sum (\mathbf{u}^T \mathbf{C}_t \mathbf{u})^2, \quad (13)$$

em que A é a área de uma esfera unitária p -dimensional e $\mathbf{C}_t$ é a t -ésima face de um *array* de curvatura, e $\mathbf{u}$ é um vetor unitário.

A notação utilizada pelos autores para a curvatura RMS de efeitos paramétricos é c^θ e a curvatura intrínseca RMS por c^l . O cálculo dessas medidas não é simples. Entretanto, os autores elaboraram um programa computacional para a obtenção dessas medidas.

Essas medidas servem para diagnóstico e comparação de modelos, indicando que quanto menor a curvatura melhor o ajuste. Além disso, são utilizadas como critério para avaliar se uma aproximação linear para os resíduos é aceitável, ou se seria mais aconselhável a utilização dos resíduos projetados. A fim de saber se a curvatura encontrada é grande, Bates e Watts (1980) recomendaram o critério $C > 2F^{-\frac{1}{2}}$, em que, F é o quantil da distribuição F com p e $n - p$ graus de liberdade.

3.5 Resíduo Projetado

Foram aplicadas as técnicas de diagnóstico, utilizando o resíduo projetado proposto por Cook e Tsai (1985), para os ajustes dos modelos aos dados de comprimento de frutos de pereira.

Seja $\tilde{\mathbf{X}} = \{\frac{\partial \mu(\boldsymbol{\beta})}{\partial \beta_r}\}$, $i = 1, \dots, n$ e $r = 1, \dots, p$ a matriz de derivadas de primeira ordem avaliada em θ (Jacobiana) também denominada por matriz modelo local. A matriz $\mathbf{W}$ é uma matriz de dimensão $(n \times p) \times p$ de derivadas de segunda ordem avaliada em $\boldsymbol{\theta}$ (Hessiana).

Para o caso em estudo $\mathbf{W}$ apresenta 3 faces, e cada uma de suas faces é

uma matriz 11×3 , ou seja, $n = 11$ observações e $p = 3$ parâmetros. As matrizes $\mathbf{H}$, $\mathbf{H}_1$ e $\mathbf{H}_2$ são as matrizes que foram definidas na seção 2.3, $\mathbf{r}$ é o resíduo estimado e $\boldsymbol{\varepsilon}$ o erro experimental.

Foram calculadas as estimativas dos resíduos projetados

$$(\mathbf{I} - \mathbf{H}_2)\mathbf{r} = (\mathbf{I} - \mathbf{H})\boldsymbol{\varepsilon} - (\mathbf{I} - \mathbf{H}_1)\boldsymbol{\varepsilon}$$

a fim de avaliar a qualidade de ajuste.

3.6 Estudo de simulação

No presente estudo, para os dois modelos empregados, foram simulados 1000 conjuntos de dados com 11 observações cada. Os valores paramétricos utilizados na simulação, foram gerados a partir das estimativas dos ajustes dos modelos, e $\boldsymbol{\varepsilon}_i \sim N(0, QMRes)$. Para cada amostra simulada, foram efetuados os ajustes dos modelos aos conjuntos de dados gerados e assim, obtidas estimativas dos parâmetros para cada uma das amostras.

Todas as análises apresentadas neste trabalho, foram realizadas por meio do *software* R, e os testes, obtidos com nível de 5% de significância.

4 RESULTADOS E DISCUSSÃO

Inicialmente, é mostrado, na Figura 3, o diagrama de dispersão de dados de crescimento de 20 frutos pereira extraídos de 10 árvores (2 frutos por planta), a cada tempo no decorrer de 154 dias.

Figura 3 - Gráfico de dispersão dos dados de comprimento de peras ao longo do tempo

No caso em estudo, é possível observar por meio da Figura 3 que ocorre uma estabilização do comprimento do fruto, valor de aproximadamente 70 mm. Também pode ser notado que as peras apresentam tendência de crescimento até aproximadamente 100 dias, a partir do qual decorre um período de menores incrementos em seu comprimento, o que mostra uma disposição sigmoidal do comportamento, concordando com os resultados obtidos por Lallu (1990) e Box (1992).

A utilização dos modelos não lineares possibilita ao pesquisador analisar tais comportamentos em diversas situações e estabelecer um padrão de resposta, crescimento-tempo, ao longo do ciclo de cada fruto. No caso de modelos sigmoidais é possível descrever o ponto de estabilização por meio de suas assíntotas superiores.

A seguir é mostrada a matriz de correlação (Cor), que apresenta valores baixos fora de sua diagonal, indicando fraca autocorrelação no tempo.

$$\text{Cor} = \begin{bmatrix} 1,000 & -0,408 & 0,098 & 0,172 & 0,232 & 0,155 & -0,118 & 0,180 & 0,004 & 0,149 & -0,185 \\ -0,408 & 1,000 & -0,068 & 0,257 & -0,166 & -0,269 & -0,267 & -0,030 & 0,246 & -0,090 & -0,335 \\ 0,098 & -0,068 & 1,000 & 0,324 & 0,261 & -0,501 & -0,373 & 0,085 & -0,390 & -0,135 & 0,292 \\ 0,172 & 0,257 & 0,324 & 1,000 & 0,453 & -0,391 & -0,513 & -0,060 & -0,301 & -0,012 & -0,239 \\ 0,232 & -0,166 & 0,261 & 0,453 & 1,000 & -0,312 & -0,456 & -0,128 & -0,222 & 0,079 & -0,034 \\ 0,155 & -0,269 & -0,501 & -0,391 & -0,312 & 1,000 & 0,418 & -0,139 & -0,063 & -0,165 & 0,210 \\ -0,118 & -0,267 & -0,373 & -0,513 & -0,456 & 0,418 & 1,000 & 0,004 & 0,115 & -0,088 & 0,317 \\ 0,180 & -0,030 & 0,085 & -0,060 & -0,128 & -0,139 & 0,004 & 1,000 & -0,066 & 0,120 & 0,047 \\ 0,004 & 0,246 & -0,390 & -0,301 & -0,222 & -0,063 & 0,115 & -0,066 & 1,000 & 0,252 & -0,230 \\ 0,149 & -0,090 & -0,135 & -0,012 & 0,079 & -0,165 & -0,088 & 0,120 & 0,252 & 1,000 & -0,423 \\ -0,185 & -0,335 & 0,292 & -0,239 & -0,034 & 0,210 & 0,317 & 0,047 & -0,230 & -0,423 & 1,000 \end{bmatrix}$$

Foi verificada a necessidade de inclusão de parâmetros autorregressivos e de médias móveis de ordem 1 e 2, constatando-se a ausência de autocorrelação residual. Uma possível explicação para esse cenário seria o fato de que foram medidos diferentes frutos, de diferentes árvores, a cada época de observação.

Ainda por meio do diagrama de dispersão, Figura 3, não são encontradas evidências de heterogeneidade de variâncias, ou seja, a cada tempo de observação a variabilidade dos dados de comprimento de peras se mantém aproximadamente constante. A seguir é apresentada a matriz de covariância (Cov) dos dados, que pode reforçar esses indícios, uma vez que os valores de sua diagonal não diferem muito.

No gráfico boxplot, apresentado na Figura 4, observa-se que a variabilidade dos dados pouco se altera ao longo do tempo. Tal constatação pode ser comprovada pelo teste Bartlett para homogeneidade de variâncias que apresentou nível descritivo igual a 0,59. O valor F calculado para o teste de Hartley foi de 2,46 enquanto o valor estabelecido para o $F_{máximo}$ de Hartley correspondente foi igual a 4,69 o que leva a não rejeição da hipótese de nulidade em que as variâncias são iguais.

$$\mathbf{Cov} = \begin{bmatrix} 12,319 & -4,367 & 1,351 & 1,576 & 2,573 & 2,137 & -1,042 & 2,095 & 0,039 & 1,637 & -1,817 \\ -4,367 & 9,311 & -0,811 & 2,043 & -1,597 & -3,228 & -2,044 & -0,307 & 2,177 & -0,862 & -2,866 \\ 1,351 & -0,811 & 15,485 & 3,323 & 3,243 & -7,749 & -3,676 & 1,110 & -4,443 & -1,664 & 3,217 \\ 1,576 & 2,043 & 3,323 & 6,803 & 3,734 & -4,014 & -3,356 & -0,520 & -2,275 & -0,102 & -1,752 \\ 2,573 & -1,597 & 3,243 & 3,734 & 9,997 & -3,876 & -3,610 & -1,344 & -2,030 & 0,783 & -0,300 \\ 2,137 & -3,228 & -7,749 & -4,014 & -3,876 & 15,470 & 4,119 & -1,804 & -0,716 & -2,025 & 2,313 \\ -1,042 & -2,044 & -3,676 & -3,356 & -3,610 & 4,119 & 6,283 & 0,032 & 0,838 & -0,686 & 2,231 \\ 2,095 & -0,307 & 1,110 & -0,520 & -1,344 & -1,804 & 0,032 & 10,954 & -0,635 & 1,242 & 0,432 \\ 0,039 & 2,177 & -4,443 & -2,275 & -2,030 & -0,716 & 0,838 & -0,635 & 8,387 & 2,280 & -1,868 \\ 1,637 & -0,862 & -1,664 & -0,102 & 0,783 & -2,025 & -0,686 & 1,242 & 2,280 & 9,760 & -3,704 \\ -1,817 & -2,866 & 3,217 & -1,752 & -0,300 & 2,313 & 2,231 & 0,432 & -1,868 & -3,704 & 7,863 \end{bmatrix}$$

Figura 4 - Gráfico boxplot do comprimento de peras (mm) em função do tempo (dias)

Posteriormente, ajustou-se o modelos logístico e de Gompertz para a média dos dados, apresentados na Tabela 1.

Tabela 1 - Média dos dados de comprimento de peras ao longo dos dias de observação

Dias (x)	Comprimento (y)	Dias (x)	Comprimento (y)
14	13,06	98	61,47
28	19,86	112	63,59
42	27,25	126	66,17
56	37,96	140	68,29
70	47,05	154	68,91
84	54,76		

4.1 Ajuste do Modelo Logístico para a média dos dados

Inicialmente, para realização do método iterativo de Gauss-Newton, obtiveram-se os valores iniciais dos parâmetros por meio da linearização do modelo como descrito em 3.3. Em seguida, ajustou-se o modelo logístico aos dados. Pode-se comprovar por meio dos resultados da Tabela 2 que o teste t para as estimativas de todos os parâmetros foi significativo considerando um nível nominal de significância de 5%.

Sabe-se também que $\alpha = 70,110$ é o máximo do comprimento do fruto e $\gamma = 0,040$ é o quanto o fruto cresce (mm) por unidade de tempo (dia). É possível calcular o ponto de inflexão, que é quando o fruto começa a desacelerar seu crescimento, para o caso do modelo logístico, ocorre quando o fruto atinge a metade do seu comprimento ($\frac{\alpha}{2} \approx 35,055$ mm) no tempo $\frac{\beta}{\gamma} \approx 51,725$ dias.

Tabela 2 - Estimativas dos parâmetros do modelo logístico com os respectivos erros padrões

Parâmetros	Estimativa	Erro Padrão	Valor t
α	70,110 *	0,467	150,100
β	2,069 *	0,043	48,600
γ	0,040 *	0,001	42,950

* significativo a 5% pelo teste t

Na Figura 5, pode-se observar a curva ajustada do modelo e seu respectivo intervalo com 95% de confiança, obtido por meio dos valores preditos pelo modelo. Nota-se que todos os pontos estão dentro do intervalo, indicando um bom ajuste.

Entretanto, o fato de os parâmetros serem todos significativos não é suficiente na validação do ajuste do modelo, existe a necessidade de verificação da qualidade do ajuste por meio de uma análise de diagnóstico.

Figura 5 - Ajuste do Modelo Logístico aos dados de comprimento de frutos de pereira e seu intervalo com 95% de confiança

Na Figura 6 está representado o histograma dos resíduos normalizados e a densidade de uma distribuição normal com mesma média e variância dos dados (Figura 6a), e o envelope simulado com 95% de confiança dos resíduos normalizados (Figura 6b) proveniente do ajuste do modelo logístico. Por meio da inspeção dos gráficos da Figura 6, verifica-se que, aparentemente, os resíduos seguem a distribuição normal. Este resultado é

confirmado pelo teste de Shapiro-Wilk para normalidade cujo nível descritivo foi de 0,87, ou seja, não existem evidências para rejeitar a hipótese nula de que os resíduos seguem distribuição normal.

Figura 6 - (a) Histogramas dos resíduos normalizados e gráfico de densidade de uma distribuição normal
(b) Envelope simulado com 95% de confiança

Além disso, foram obtidos os resíduos projetados do ajuste do modelo com a finalidade de compará-lo com o resíduo normalizado. Para obtenção desses resíduos foi necessária a estruturação das matrizes $(\tilde{\mathbf{X}})$ e $(\mathbf{W}_p)$ descritas na seção 2.3.

A seguir estão apresentadas as matrizes $\tilde{\mathbf{X}}$, com dimensões 11×3 , e $\mathbf{W}$ que apresenta 3 faces, uma para cada parâmetro, e cada uma de suas faces é uma matriz 11×3 ($n = 11$ observações e $p = 3$ parâmetros).

$$\tilde{\mathbf{X}} = \begin{bmatrix} 0,1806 & -10,3743 & 145,2405 \\ 0,2778 & -14,0662 & 393,8537 \\ 0,4018 & -16,8498 & 707,6926 \\ 0,5397 & -17,4162 & 975,3051 \\ 0,6717 & -15,4588 & 1082,1140 \\ 0,7818 & -11,9798 & 1006,3030 \\ 0,8619 & -8,3502 & 818,3210 \\ 0,9159 & -5,4027 & 605,0991 \\ 0,9500 & -3,3301 & 419,5951 \\ 0,9707 & -1,9919 & 278,8678 \\ 0,9830 & -1,1702 & 180,2094 \end{bmatrix} \quad \mathbf{W}_\alpha = \begin{bmatrix} 0 & -0,1480 & 2,0717 \\ 0 & -0,2006 & 5,6180 \\ 0 & -0,2403 & 10,0947 \\ 0 & -0,2484 & 13,9119 \\ 0 & -0,2205 & 15,4355 \\ 0 & -0,1709 & 14,3541 \\ 0 & -0,1191 & 11,6727 \\ 0 & -0,0771 & 8,6312 \\ 0 & -0,0475 & 5,9852 \\ 0 & -0,0284 & 3,9778 \\ 0 & -0,0167 & 2,5705 \end{bmatrix}$$

$$\mathbf{W}_\beta = \begin{bmatrix} -0,1480 & 6,6272 & -92,7807 \\ -0,2006 & 6,2500 & -175,0005 \\ -0,2403 & 3,3106 & -139,0462 \\ -0,2484 & -1,3813 & 77,3526 \\ -0,2205 & -5,3096 & 371,6753 \\ -0,1709 & -6,7393 & 566,1039 \\ -0,1191 & -6,0420 & 592,1183 \\ -0,0771 & -4,4935 & 503,2668 \\ -0,0475 & -2,9971 & 377,6343 \\ -0,0284 & -1,8753 & 262,5430 \\ -0,0167 & -1,1305 & 174,0895 \end{bmatrix} \quad \mathbf{W}_\gamma = \begin{bmatrix} 2,0717 & -92,7807 & 1298,9300 \\ 5,6180 & -175,0005 & 4900,0150 \\ 10,0947 & -139,0462 & 5839,9400 \\ 13,9119 & 77,3526 & -4331,7470 \\ 15,4355 & 371,6753 & -26017,2680 \\ 14,3541 & 566,1039 & -47552,7280 \\ 11,6727 & 592,1183 & -58027,5970 \\ 8,6312 & 503,2668 & -56365,8850 \\ 5,9852 & 377,6343 & -47581,9160 \\ 3,9778 & 262,5430 & -36756,0220 \\ 2,5705 & 174,0895 & -26809,7760 \end{bmatrix}$$

Na Figura 7, são apresentados os resíduos normalizados (Figura 7a) e os resíduos projetados (Figura 7b) do ajuste do modelo.

Pode-se notar que, assim como os resíduos normalizados, os resíduos projetados estão dispostos de maneira aleatória em torno de zero. Além disso, em ambos os casos, não foram verificados resíduos fora do intervalo $(-2, 2)$ indicando um bom ajuste do modelo.

Figura 7 - Resíduos do ajuste do modelo logístico aos dados de comprimento de peras, (a) Normalizado (b) Projetado

Assim como para os resíduos normalizados, o teste de normalidade de Shapiro-Wilk foi aplicado para os resíduos projetados, resultou em um nível descritivo de 0,2449, permitindo concluir que, também para os resíduos projetados, a pressuposição de normalidade foi verificada.

Ao se calcularem as medidas de curvatura de efeito do parâmetro e intrínseca do modelo, obtiveram-se, respectivamente $c^\theta = 0,0716$ e $c^l = 0,0145$. O valor $2F^{-\frac{1}{2}}$ calculado com 3 e 8 graus de liberdade, foi de 0,9914 conforme sugerido por Bates e Watts (1980). Como ambos valores, c^θ e c^l , foram inferiores ao recomendado pelos autores, pode ser concluído que uma aproximação linear dos resíduos é aceitável.

Ainda como diagnóstico de ajuste do modelo logístico, são apresentados, na Figura 8, os gráficos da distância de Cook (Figura 8a) e dos elementos da diagonal da matriz de projeção (Figura 8b) do modelo. Por meio do gráfico da distância de Cook, verifica-se a presença de possíveis pontos influentes (3ª e 7ª observações). Entretanto, a retirada desses pontos apresentou uma variação muito pequena nas estimativas dos parâmetros, sendo esses pontos não considerados influentes, mantidos. O gráfico dos elementos da diagonal da matriz de projeção do modelo, confirmou esse resultado, uma vez que nenhum ponto foi superior ao valor crítico de $\frac{2p}{n} \approx 0,55$.

Figura 8 - Gráfico de índices das observações *versus*, (a) Distância de Cook e (b) elementos da diagonal da matriz de projeção

4.2 Ajuste do Modelo de Gompertz para a média dos dados

Para verificação do ajuste do modelo de Gompertz à média dos dados de comprimento de peras, procedeu-se de forma semelhante ao logístico. A seguir, é possível observar os resultados para este segundo modelo. Também, foi utilizado o método iterativo de Gauss-Newton, com a obtenção dos valores iniciais por meio da linearização do modelo conforme item 3.3. Pode-se comprovar por meio dos resultados da Tabela 3 que o teste t para as estimativas de todos os parâmetros foi significativo considerando um nível nominal de significância de 5%.

Da mesma forma que para o modelo logístico, o modelo de Gompertz apresenta o máximo do comprimento do fruto em $\alpha \approx 73,739$, sua taxa de crescimento $\gamma \approx 0,026$ (mm) por unidade de tempo (dia). E o ponto de inflexão para esse modelo ocorre quando o fruto atinge o comprimento $\frac{\alpha}{e} \approx 27,31$ (mm) no tempo $\frac{\beta}{\gamma} \approx 38,423$ dias.

Tabela 3 - Estimativas dos parâmetros do modelo de Gompertz com os respectivos erros padrões

Parâmetros	Estimativa	Erro Padrão	Valor t
α	73,739 *	1,583	46,570
β	0,999 *	0,058	17,120
γ	0,026 *	0,002	15,800

* significativo a 5% pelo teste t

É possível observar, na Figura 9, a curva ajustada do modelo de Gompertz aos dados de crescimento e seu respectivo intervalo com 95% de confiança. Para esse modelo, não há pontos fora do intervalo, sendo assim, há evidências de um bom ajuste.

Figura 9 - Ajuste do Modelo de Gompertz aos dados de comprimento de frutos de pereira e seu intervalo com 95% de confiança

Mesmo com a figura indicando um bom ajuste, existe a necessidade de verificação e diagnóstico do comportamento dos resíduos. Com base no ajuste do modelo, foram encontrados os resíduos normalizados.

Na Figura 10 estão representados os gráficos de densidade (Figura 10a), e o envelope simulado com 95% de confiança dos resíduos normalizados (Figura 10b), provenientes do ajuste do modelo de Gompertz.

Por meio da inspeção dos gráficos da Figura 10, verifica-se que, aparentemente os resíduos seguem distribuição normal. Esse resultado é confirmado por meio do teste de Shapiro-Wilk para normalidade cujo nível descritivo foi de 0,7053, ou seja, não existem evidências para rejeitar a hipótese nula de que os resíduos seguem distribuição normal.

Figura 10 - (a) Histograma dos resíduos normalizados e gráfico de densidade e (b) Envelope simulado com 0,95 de probabilidade

Foram calculados os resíduos projetados do modelo a fim de compará-los com os resíduos normalizados. Para obtenção desses resíduos, assim como para o modelo logístico, é necessária a estruturação das matrizes $(\tilde{\mathbf{X}})$ e $(\mathbf{W}_p)$. Essas matrizes que se seguem são obtidas da mesma forma e possuem as mesmas dimensões das calculadas para o modelo logístico, já que os modelos apresentam o mesmo número de parâmetros.

$$\tilde{\mathbf{X}} = \begin{bmatrix} 0,1506 & -21,0246 & 294,3446 \\ 0,2672 & -26,0034 & 728,0950 \\ 0,3985 & -27,0356 & 1135,4958 \\ 0,5265 & -24,9046 & 1394,6564 \\ 0,6394 & -21,0852 & 1475,9611 \\ 0,7322 & -16,8316 & 1413,8588 \\ 0,8046 & -12,8963 & 1263,8388 \\ 0,8594 & -9,6026 & 1075,4872 \\ 0,8997 & -7,0089 & 883,1257 \\ 0,9290 & -5,0452 & 706,3330 \\ 0,9499 & -3,5967 & 553,8908 \end{bmatrix} \quad \mathbf{W}_\alpha = \begin{bmatrix} 0 & -0,2851 & 3,9917 \\ 0 & -0,3526 & 9,8730 \\ 0 & -0,3666 & 15,3989 \\ 0 & -0,3377 & 18,9134 \\ 0 & -0,2859 & 20,0160 \\ 0 & -0,2283 & 19,1739 \\ 0 & -0,1749 & 17,1394 \\ 0 & -0,1302 & 14,5851 \\ 0 & -0,0950 & 11,9764 \\ 0 & -0,0684 & 9,5788 \\ 0 & -0,0488 & 7,5115 \end{bmatrix}$$

$$\mathbf{W}_\beta = \begin{bmatrix} -0,2851 & 18,7749 & -92,7807 \\ -0,3526 & 8,3141 & -175,0005 \\ -0,3666 & -2,1609 & -139,0462 \\ -0,3377 & -8,9297 & 77,3526 \\ -0,2859 & -11,6561 & 371,6753 \\ -0,2283 & -11,5841 & 566,1039 \\ -0,1749 & -10,0933 & 592,1183 \\ -0,1302 & -8,1475 & 503,2668 \\ -0,0951 & -6,2685 & 377,6343 \\ -0,0684 & -4,6737 & 262,5430 \\ -0,0488 & -3,4120 & 174,0895 \end{bmatrix} \quad \mathbf{W}_\gamma = \begin{bmatrix} 3,9917 & -262,8494 & 3679,8920 \\ 9,8740 & -232,7957 & 6518,2790 \\ 15,3989 & 90,7586 & -3811,8630 \\ 18,9134 & 500,0659 & -28003,6890 \\ 20,0160 & 815,9252 & -57114,7660 \\ 19,1738 & 973,0663 & -81737,5660 \\ 17,1394 & 989,1408 & -96935,7990 \\ 14,5851 & 912,5183 & -102202,0540 \\ 11,9764 & 789,8309 & -99518,6930 \\ 9,5788 & 654,3119 & -91603,6620 \\ 7,5115 & 525,4507 & -80919,4120 \end{bmatrix}$$

Na Figura 11, são apresentados os resíduos normalizados (Figura 11a) e os resíduos projetados (Figura 11b) do ajuste do modelo. Pode ser observado que os resíduos projetados se apresentaram de forma um pouco mais aleatória, descartando uma possível autocorrelação indicada pela análise gráfica dos resíduos normalizados.

Figura 11 - Resíduos do ajuste do modelo de Gompertz aos dados de comprimento de peras: **(a)** Normalizado **(b)** Projetado

O teste de normalidade de Shapiro-Wilk foi aplicado para os resíduos projetados, resultou em um nível descritivo de 0,2148, permitindo concluir que a pressuposição de normalidade foi verificada.

No caso do modelo de Gompertz, as medidas de curvatura de efeito do parâmetro e intrínseca foram respectivamente $c^\theta = 0,2689$ e $c^l = 0,0273$. Pela avaliação sugerida por Bates e Watts (1980), o valor $2F^{-\frac{1}{2}}$, com 3 e 8 graus de liberdade, também foi 0,9914. Como ambos os valores, c^θ e c^l foram inferiores ao sugerido pelos autores, pode ser concluído que uma aproximação linear dos resíduos é aceitável.

4.3 Algumas considerações a respeito dos ajustes dos modelos

Na Figura 12, estão representados os valores médios de comprimento das peras, bem como o ajuste dos modelos logístico e de Gompertz. Nota-se uma pequena diferença entre os ajustes dos dois modelos, sendo o modelo logístico levemente superior em qualidade de ajuste no início e final do banco de dados.

Figura 12 - Ajuste dos modelos logístico e de Gompertz à média dos dados de comprimento de peras

A fim de comparar os ajustes dos dois modelos e diagnosticar qual é o modelo mais adequado para a descrição do fenômeno, já que ambos se ajustam bem aos dados, foram obtidos os valores pelos critérios que estão apresentados na Tabela 4.

Tabela 4 - Critério de informação de Akaike (AIC), critério de informação bayesiano (BIC), medida de curvatura de efeito do parâmetro, medida de curvatura intrínseca do modelo e quadrado médio do resíduo

Modelo	AIC	BIC	c^θ	c^l	QMRes
logístico	1138,661	1152,236	0,072	0,015	0,362
Gompertz	1162,916	1176,491	0,269	0,027	1,967

Segundo Souza et al. (2010), as medidas de curvatura, além de estabelecerem um critério de utilização dos resíduos projetados, também auxiliam na comparação de

modelos. Cordeiro, Prudente e Demétrio (2009) afirmaram que quanto menor a medida de curvatura, mais indicado é o ajuste do modelo e quanto maiores as medidas de curvatura maior é a indicação da utilização dos resíduos projetados nas análises de diagnóstico. Os modelos também podem ser comparados por meio do critério de informação de Akaike (AIC), e pelo critério bayesiano (BIC).

Um fato a ser observado é o quadrado médio dos resíduos (QMRes) de ambos os modelos, que é menor para o ajuste do primeiro do que do segundo. É notório, pelas Figuras 5 e 9, que o intervalo de confiança do modelo logístico é menor que o do modelo de Gompertz, indicando que o primeiro é mais criterioso que o segundo.

Com base nos critérios apresentados na Tabela 4, é possível indicar que o modelo mais adequado para ajuste aos dados é o modelo logístico, pois os valores de todos os critérios avaliados para esse modelo foram inferiores quando comparados ao modelo de Gompertz.

4.4 Estudo de simulação

Os estimadores dos parâmetros, em modelos de regressão não linear, não possuem algumas das propriedades apresentadas pelos estimadores dos parâmetros em modelos de regressão linear. Assim sendo, existe a necessidade de estudar o comportamento do modelo e, segundo Ratkowsky (1983), isso pode ser feito por meio de estudos de simulação.

Para a obtenção das amostras simuladas foram consideradas, como valores paramétricos, as estimativas dos parâmetros dos modelos logístico e de Gompertz ajustados aos dados experimentais de comprimento de pera. Foram gerados 1000 amostras a partir de cada um dos modelos, cada uma com 11 observações, e considerando $\varepsilon_i \sim N(0, \hat{\sigma}^2)$ em que $\hat{\sigma}^2$ é a estimativa da variância residual obtida, também, dos ajustes dos modelos aos dados experimentais ($\sigma_L^2 = 0,6014$ e $\sigma_G^2 = 1.4042$). Para cada uma das 1000 amostras, foram novamente ajustados os modelos logístico e de Gompertz, obtendo-se 1000 estimativas para cada parâmetro, uma vez que o percentual de convergência para os ajustes dos modelos para as amostras foi de 100%.

Nas Tabelas 5 e 6, estão apresentadas as seis primeiras amostras simuladas por meio dos modelos logístico e de Gompertz, respectivamente. Na Tabela 7, encontram-se as estimativas dos parâmetros para cada uma das amostras bem como o

valor paramétrico utilizado no estudo de simulação. São apresentados, média, variância e erro quadrático médio (EQM) dos parâmetros, para cada um dos modelos ajustados (Tabela 8).

Tabela 5 - Amostras dos comprimentos dos frutos de pereira, simulados para o modelo logístico nas 11 idades

Idades (dia)	Amostras					
	1	2	3	4	5	6
14	12,6111	12,2729	11,8469	12,5603	13,0855	11,9906
28	19,7746	19,0688	18,4115	19,5987	19,0015	20,7633
42	28,6374	27,8201	29,0713	26,4822	29,1023	28,2741
56	37,9151	36,2642	37,6073	38,1507	37,7681	37,6218
70	47,8728	47,6233	48,3031	47,4685	47,5997	46,6816
84	54,1159	54,2698	54,3956	54,9370	55,9697	54,7216
98	60,3012	60,5296	60,4699	59,6258	60,0938	59,8763
112	64,6511	65,0352	64,2666	64,9203	63,6105	63,9654
126	66,2815	66,5907	66,3184	67,5659	65,4306	66,4373
140	68,4210	67,8778	68,8071	68,4474	67,3712	67,9241
154	68,5543	68,9753	69,4726	68,6374	68,8248	69,1739

Tabela 6 - Amostras dos comprimentos dos frutos de pereira, simulados para o modelo de Gompertz nas 11 idades

Idades(dia)	Amostras					
	1	2	3	4	5	6
14	10,9906	10,2019	9,20850	10,8721	12,0968	9,5435
28	20,3955	18,7496	17,2170	19,9853	18,5927	22,7012
42	30,4842	28,5783	31,4960	25,4583	31,5683	29,6371
56	39,0177	35,1680	38,2999	39,5671	38,6747	38,3336
70	48,9700	48,3883	49,9735	48,0272	48,3332	46,1923
84	52,4580	52,8168	53,1102	54,3728	56,7809	53,8706
98	59,0654	59,5982	59,4589	57,4906	58,5819	58,0747
112	64,4071	65,3029	63,5106	65,0350	61,9806	62,8082
126	65,6031	66,3241	65,6890	68,5982	63,6188	65,9665
140	69,3598	68,0931	70,2603	69,4213	66,9117	68,2011
154	69,2066	70,1884	71,3480	69,4003	69,8374	70,6515

Tabela 7 - Estimativas dos parâmetros dos modelos logístico e de Gompertz para as seis primeiras amostras simuladas

Parâmetros	Modelo logístico			Parâmetros	Modelo de Gompertz		
	α	β	γ		α	β	γ
θ_1	70,0103	2,0470	0,0397	θ_1	73,2083	0,9801	0,0260
θ_2	70,2292	2,1299	0,0403	θ_2	74,2418	1,0662	0,0261
θ_3	70,1204	2,1211	0,0408	θ_3	73,3999	1,0846	0,0276
θ_4	70,3723	2,1158	0,0402	θ_4	74,5638	1,0506	0,0261
θ_5	69,1602	2,0645	0,0408	θ_5	70,8745	1,0042	0,0278
θ_6	70,2470	2,0275	0,0389	θ_6	74,3044	0,9496	0,0245
θ_t	70,1100	2,0700	0,0400	θ_G	73,7390	0,9990	0,0260

Tabela 8 - Média, variância e erro quadrático médio das estimativas dos parâmetros do ajuste dos modelos para as 1000 amostras simuladas

Parâmetros	Modelo logístico			Parâmetros	Modelo de Gompertz		
	Média	Variância	EQM		Média	Variância	EQM
α	70,1218	0,2320	0,2320	α	75,0553	3,1438	3,1569
β	2,0724	0,0019	0,0019	β	0,9817	0,0032	0,0033
γ	0,0398	0,0000	0,0000	γ	0,0248	0,0000	0,0000

Como os testes para porcentagem de viés (Tabela 9) e para o excesso de variância (Tabela 10) não foram significativos (n.s.), considerando um nível de significância de 5%, podemos supor que para esse conjunto de dados os estimadores dos parâmetros apresentam comportamento próximo do linear.

Pode-se dizer que a distribuição dos estimadores dos parâmetros convergem para uma distribuição normal. Este fato pode ser confirmado pelos histogramas apresentados na Figuras 13 e 14 e pelo teste de normalidade de Shapiro-Wilk, em que, para todos os parâmetros o nível descritivo foi maior que 0,05, não rejeitando a hipótese nula de que os estimadores seguem distribuição normal.

Tabela 9 - Excesso de variância dos parâmetros obtidos dos ajustes do modelo para cada simulação

Parâmetros	Modelo logístico		Parâmetros	Modelo de Gompertz	
	Ev (%)	Z		Ev (%)	Z
α	6,358	1,410 (n.s.)	α	8,556	1,884 (n.s.)
β	2,852	0,644 (n.s.)	β	2,164	0,492 (n.s.)
γ	0,756	0,180 (n.s.)	γ	0,721	0,172 (n.s.)

Figura 13 - Histogramas das estimativas dos parâmetros para as 1000 amostras simuladas a partir do modelo logístico, (a) Alfa (b) Beta (c) Gama

Tabela 10 - Porcentagem de viés dos parâmetros obtidos dos ajustes dos modelos

Parâmetros	Modelo logístico		Parâmetros	Modelo de Gompertz	
	Viés(%)	Z		Viés(%)	Z
α	0,023	1,059 (n.s.)	α	1,785	23,479 (n.s.)
β	0,143	2,173 (n.s.)	β	-1,718	-9,538 (n.s.)
γ	0,101	1,367 (n.s.)	γ	-3,733	-19,042 (n.s.)

Figura 14 - Histogramas das estimativas dos parâmetros para as 1000 amostras simuladas a partir do modelo de Gompertz, (a) Alfa (b) Beta (c) Gama

Para cada uma das 1000 amostras simuladas foram obtidos os resíduos projetados para os modelos logístico e de Gompertz (Tabela 11). Algumas estatísticas descritivas (Tabela 12) para a construção do gráfico normal de probabilidade com envelope simulado.

O gráfico normal de probabilidade para os resíduos projetados provenientes do ajuste dos modelos aos dados reais e envelope de simulação são apresentados nas Figuras 15 e 16 para os modelos logístico e de Gompertz, respectivamente. Pode-se notar que todos os resíduos projetados calculados para os dados reais se encontram dentro do envelope, indicando evidência de um bom ajuste dos modelos.

Tabela 11 - Resíduos projetados, dos modelos logístico e de Gompertz, para as primeiras seis amostras simuladas

	Modelo logístico						Modelo de Gompertz					
	rp ₁	rp ₂	rp ₃	rp ₄	rp ₅	rp ₆	rp ₁	rp ₂	rp ₃	rp ₄	rp ₅	rp ₆
x ₁	-0,77	-0,75	-0,74	-0,76	-0,74	-0,78	-1,10	-1,19	-1,04	-1,18	-0,93	-1,24
x ₂	1,00	1,00	0,99	1,00	0,98	1,01	1,18	1,26	1,12	1,25	1,02	1,30
x ₃	-1,32	-1,30	-1,33	-1,31	-1,35	-1,29	-1,30	-1,35	-1,28	-1,34	-1,20	-1,36
x ₄	0,78	0,77	0,80	0,76	0,83	0,75	0,81	0,63	0,87	0,67	1,01	0,59
x ₅	-0,31	-0,34	-0,31	-0,33	-0,28	-0,34	-0,04	-0,03	-0,08	-0,03	-0,14	0,00
x ₆	-0,66	-0,65	-0,67	-0,65	-0,70	-0,63	-0,49	-0,36	-0,54	-0,38	-0,67	-0,32
x ₇	1,85	1,89	1,86	1,88	1,81	1,88	1,51	1,58	1,54	1,56	1,53	1,53
x ₈	-1,15	-1,15	-1,14	-1,15	-1,14	-1,16	-1,26	-1,27	-1,24	-1,27	-1,22	-1,28
x ₉	-0,46	-0,48	-0,46	-0,48	-0,43	-0,49	-0,43	-0,54	-0,44	-0,52	-0,37	-0,52
x ₁₀	0,86	0,84	0,85	0,85	0,86	0,85	1,05	0,98	1,02	1,00	1,02	1,02
x ₁₁	-0,51	-0,43	-0,49	-0,45	-0,57	-0,46	-0,99	-0,54	-0,83	-0,63	-0,95	-0,69

Tabela 12 - Estatísticas para os resíduos projetados dos modelos logístico e de Gompertz, considerando as 1000 amostras simuladas

	Modelo logístico						Modelo de Gompertz					
	mín	q(25)	média	q(50)	q(975)	máx	mín	q(25)	média	q(50)	q(975)	máx
x ₁	-1,37	-1,36	-1,31	-1,31	-1,24	-1,19	-1,44	-1,40	-1,32	-1,33	-1,20	-1,17
x ₂	-1,16	-1,16	-1,15	-1,15	-1,13	-1,12	-1,38	-1,36	-1,27	-1,27	-1,17	-1,09
x ₃	-0,81	-0,79	-0,76	-0,76	-0,73	-0,72	-1,29	-1,28	-1,13	-1,15	-0,90	-0,81
x ₄	-0,72	-0,71	-0,66	-0,65	-0,62	-0,61	-1,06	-0,96	-0,78	-0,81	-0,56	-0,54
x ₅	-0,59	-0,56	-0,50	-0,50	-0,47	-0,47	-0,77	-0,70	-0,49	-0,50	-0,13	-0,09
x ₆	-0,49	-0,48	-0,45	-0,45	-0,38	-0,38	-0,48	-0,46	-0,35	-0,39	-0,08	-0,04
x ₇	-0,38	-0,37	-0,32	-0,32	-0,28	-0,27	-0,26	-0,17	-0,03	-0,03	0,19	0,29
x ₈	0,72	0,73	0,78	0,77	0,84	0,85	0,17	0,31	0,71	0,72	0,99	1,01
x ₉	0,80	0,83	0,85	0,85	0,87	0,88	0,53	0,73	0,99	1,02	1,07	1,10
x ₁₀	0,96	0,97	1,00	1,00	1,02	1,02	1,01	1,03	1,22	1,23	1,44	1,47
x ₁₁	1,79	1,82	1,87	1,87	1,97	1,93	1,47	1,49	1,56	1,54	1,75	1,98

O valor do quadrado médio residual obtido para os dados reais no ajuste do modelo logístico (QMResl= 0,3617) foi menor que o valor obtido no ajuste do modelo de Gompertz (QMResg= 1,9669). Como a simulação foi realizada considerando-se esses valores encontrados para os dados reais, da mesma forma que para os intervalos de confiança, a amplitude dos resíduos encontrados para todas as curvas simuladas foi menor no logístico que no de Gompertz (Figuras 15 e 16). Sendo assim, a simulação e os resíduos projetados confirmam que para o caso em estudo, o ajuste do modelo logístico foi melhor que o de Gompertz.

Na Tabela 13, são apresentadas as estatísticas (mínimo, máximo, média e

quantis) das medidas de curvatura das 1000 amostras simuladas, podendo ser observado que nenhum dos valores foi maior que o estabelecido pelo critério ($2F^{-\frac{1}{2}} = 0,9914$) de Bates e Watts (1980). Com isso, pode-se concluir que, para os dados simulados, as medidas de curvatura foram pequenas indicando que, pode ser realizada uma aproximação linear para diagnóstico do modelo apesar de os resíduos projetados serem melhores para descrever os resultados obtidos.

Tabela 13 - Estatísticas para as curvaturas dos modelos logístico e de Gompertz, considerando as 1000 amostras simuladas

	Modelo Logístico						Modelo de Gompertz					
	mín	q(25)	média	q(50)	q(975)	máx	mín	q(25)	média	q(50)	q(975)	máx
c^θ	0,0215	0,0371	0,0705	0,0699	0,1058	0,1289	0,0688	0,1325	0,2671	0,2587	0,4191	0,5970
c^l	0,0047	0,0075	0,0142	0,0140	0,0214	0,0261	0,0096	0,0142	0,0269	0,0265	0,0410	0,0518

Figura 15 - Gráfico normal de probabilidade dos resíduos projetados calculados para o ajuste do modelo logístico aos dados reais, e envelope simulado

Figura 16 - Gráfico normal de probabilidade dos resíduos projetados calculados para o ajuste do modelo de Gompertz aos dados reais, e envelope simulado

5 CONCLUSÃO

No presente estudo, tanto o modelo logístico quanto o de Gompertz se ajustaram bem aos dados, entretanto, o primeiro foi melhor. Portanto, para o intervalo de tempo estudado, os modelos se mostraram adequados para descrever o comportamento do comprimento das peras da cultivar *Shinsseiki* (*Pyrus pyrifolia*) ao longo do tempo.

Os resíduos projetados se mostraram viáveis para análise de diagnóstico em modelos não lineares de regressão. Além disso, pôde ser observado com maior clareza no diagnóstico do modelo de Gompertz que os resíduos normalizados apresentaram uma tendência que poderia indicar a presença de autocorrelação. Ao serem calculados os resíduos projetados, esses apresentaram maior aleatoriedade em torno de zero descartando tal suspeita, o que fez com que se mostrassem melhores.

Como as medidas de não linearidade dos modelos foram pequenas, pode-se concluir que tanto o resíduo normalizado quanto o projetado podem ser utilizados. Entretanto, quanto maior a medida de curvatura mais indicada a utilização do resíduo projetado. Isso pôde ser comprovado por meio do estudo de simulação.

REFERÊNCIAS

- AKAIKE, H. A new look at the statistical model identification. **IEEE Transation on Automatic Control**. New York, v. 19, p. 716-723, 1974.
- ARGENTA, L. C.; BENDER, R. J.; KREUS, C. L.; MONDARMO, M. Padrões de maturação e índices de colheita de maçãs cvs. Gala, Golden Delicious e Fuji. **Pesquisa Agropecuária Brasileira**. Brasília, v. 30, n. 10, p. 1259-1266, out. 1995.
- AVILA, A. L., ARGENTA, M. S. da; MUNIZ, M. F. B. de; PELETO, I.; BLUME, E. Maturação fisiológica e colheita de sementes de *Eugenia uniflora* L. (Pitanga), RS. **Ciência Florestal**. Santa Maria, v. 19, n.1, jan./mar. 2009.
- BATES, D. M.; WATTS, D. G.; Relative Curvature Measures of Nonlinearity. **Journal of the Royal Statistical Society**. Serie B. Methodological, London, v. 42, p. 1-25, 1980.
- BATES, D. M.; WATTS, D. G. **Nonlinear regression analysis and its applications**. New York: Jonh Wiley & Sons. 1988. 365p.
- BOX, J. M. M. **El nashi**. Madrid: Mundi, 1992. 131p.
- BRACCINI NETO, J.; DIONELLO, J. L.; JÚNIOR, P. S.; BONGALHARDO, D.; XAVIER, E. Análise de curvas de crescimento de aves de postura. **Revista Brasileira de Zootecnia**. Brasília, v. 25, n.6, p. 1062-1073, 1996.
- BRODY, S. **Bioenergetcs and growth**: with special reference to the efficiency complex of domestic animals. New York: Hafner, 1945. 1023p.
- BURNHAM, K. P.; ANDERSON, D. R. Multimodel Inferennce: Understanding AIC and BIC in Model Selection. **Sociological Methods Research**. Washington, v. 33, n. 2, p. 261-304, 2004.
- CALLIOLI, C. A.; DOMINGUES, H.H.; COSTA, R. C. **Álgebra Linear e Aplicações**, 6. ed. São Paulo: Atual, 1997. 351p.
- CARLESSO, V. O.; BERBERT, P. A.; SILVA, R. S. da; DETMANN, E. Avaliação de modelos de secagem em camada fina de sementes de maracujá amarelo. **Revista Brasileira de Sementes**. Londrina, v. 29, n. 2, p. 28-37, 2007.
- COOK, R. D.; TSAI, C. L. Residuals in Nonlinear Regression. **Biometrika**. London, v. 72, p. 23-29, 1985.
- CORDEIRO, G. M.; PRUDENTE, A. A.; DEMETRIO, C. G. B.; Uma revisão dos modelos normais não lineares. **Revista Brasileira Biometria**. São Paulo, v. 27, n.3, p. 360-393, 2009.
- CORREA, P. C.; ARAÚJO, E. F.; AFONSO JÚNIOR, P. C. Determinação dos parâmetros de secagem em camada delgada de sementes de milho doce (*Zea mays* L.). **Revista Brasileira de Milho e Sorgo**. Sete Lagoas, v. 2, n. 2, p. 110-119, 2003.

COX, D. R.; SNELL, E. J. A General Definition of Residuals **Journal of the Royal Statistical Society. Series B(Methodological)**. London, v. 20, n. 2, p. 248-275, 1968.

CROCCI, A. J. **Modelo autorregressivo, para análise de experimentos com vacas em lactação**. 1984. 86p. Tese (Doutorado em Agronomia) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 1984.

CUNHA, A. R.; VOLPE, C. A. Curvas de crescimento do fruto de cafeeiro cv. Obatã IAC 1669-20 em diferentes alinhamentos de plantio. **Semina: Ciências Agrárias**. Londrina, v. 32, n. 1, p. 49-62, jan./mar., 2011.

DRAPER, N. R.; SMITH, H; **Applied regression analysis**. 3rd ed., New York: Jonh Wiley & Sons, 1998. 706p.

FAORO, I. D.; ORTH, A. I.; Qualidade de frutos da pereita-japonesa colhidos em duas regiões de Santa Catarina, Brasil **Revista Brasileira de Fruticultura**. Jaboticabal - SP, v. 32, n. 1, p. 308-315, 2010.

FIDELES FILHO, J.; NOBREGA, J. Q.; SOUSA, J. T.; DANTAS, J. P. Monitoramento de área foliar e biomassa do feijoeiro usando índice de vegetação por diferença normalizada. **Revista Brasileira de Agrometeorologia**. Piracicaba, v. 13, n. 2, p. 2012-218, 2005.

FORNI, S. **Análise da curva de crescimento de bovinos da raça Nelore, utilizando modelos não-lineares em análises bayesianas**. 2007. 650. Tese (Doutorado em Zootecnia) Jaboticabal, Brasil: Faculdade de Ciências Agrárias e Veterinárias, UNESP, 2007.

GALLANT, A. R. **Nonlinear statistical models**. New York: John Wiley & Sons, 1987. 610p.

HARTLEY, H. O. The maximum F-ratio as a short-cut test for heterogeneity of variance. **Biometrika**. Cambridge, v. 37, p. 308-312, 1950.

HARTLEY, H. O. The modified Gauss-Newton method for the fitting of non-linear regression functions by least squares. **Technometrics**. Alexandria, v. 3, n. 2, p. 269-280, 1961.

HOFFMAN, R.; VIEIRA, S. **Análise de regressão: uma introdução à econometria**. 3. ed. São Paulo: Hucitec, 1998. 372p.

IOSSI, E.; SABER R.; MORO, Maturação fisiológica de sementes *Phoenix roebelenii* O'BRIEN. **Revista Brasileira de Sementes**. Pelotas, vol. 29, n. 1, p. 147-154, 2007.

LAIRD, A. K. Dynames of relative growth, **Growth**. Bar Harbor, v. 29, n. 9, p. 249-263, 1965.

LALLU, N. Fruit growth and development. **Nashi-Asian pear in New Zealand**. Wellington: Division of Horticulture and Processing, 1990. 85p.

LOMBARDI, S. R. B.; MORAES, D. M.; CAMELATTO, D. Avaliação do crescimento e maturação pós colheita de peras da cultivar Shinsseiki. **Pesquisa Agropecuária Brasileira**. Brasília, v. 35, n. 12, p. 2399-2405, dez 2000.

MAROCO, J. **Análise Estatística com a utilização do SPSS**. 3. ed. Silabo. Lisboa, 2007. 822p.

MAZUCHELI, J.; ACHCAR, J. Algumas considerações em regressão não linear. **Acta Scientiarum**, Maringá, v. 24, n. 6, p. 1761-1770, 2002.

MAZZINI, A. R. A.; MUNIZ, J. A.; AQUINO, L. H.; SILVA, F. F. Análise da curva de crescimento de machos Hereford. **Ciência e Agrotecnologia**. Lavras, v. 27, p. 1105-1112, 2003.

MAZZINI, A. R. A.; MUNIZ, J. A.; SILVA, F. F.; AQUINO, L. H. Curva de crescimento de novilhos Hereford: heterocedasticidade e resíduos autorregressivos. **Ciência Rural**. Santa Maria, v. 35, n. 2, p. 422-427, 2005.

McMANUS, C.; EVANGELISTA, L. A. C.; MIRANDA, R. M. de; MORENO BERNAL, F. E.; SANTOS, N. R. dos Curvas de crescimento de Ovinos Bergamácia Criados no Distrito Federal. **Revista Brasileira de Zootecnia**. Viçosa, v. 32, n.5, p. 1207-1212, 2003.

MONTGOMERY, D. C.; PECK, E. A.; VINING, G. G. **Introdution to linear regression analysis**. 4th ed. New York: Wiley, 2006. 621p.

MORETTIN, P. A.; TOLOI, C. M. de D. **Previsão de séries temporais**. 3. ed. São Paulo: Atual, 2004, 436p.

NELDER, J. A. The fitting of a generation of the logistic curve. **Biometrics**. Washington, v. 17, p. 89-110, 1961.

OLIVEIRA, M. M.; **Modelos de Regressão não Lineares**, 2004 Disponível em: <<http://www.fep.up.pt/disciplinas/2E103/nlin.pdf>>. Acesso em: 20 nov. 2012.

PAIXÃO, F. J. R. da; ANDRADE, A. R. S.; AZEVEDO, C. A. V. de; SILVA, J. M. da; COSTA, T. L.; FEITOSA, R. M. Estimativa da infiltração da água no solo através de modelos empíricos e funções não-lineares. **Revista de Biologia e Ciências da Terra**. Campina Grande, v. 5, n. 1, p. 1-12, 2005.

PEREIRA, A. R.; ANGELOCCI, L. R.; SENTELHAS, P. C. Meteorologia Agrícola. Apostila. Escola Superior de Agricultura “Luiz de Queiroz”, Piracicaba - SP, fev. 2007.

PERRE, M. A.; LIMA, E. A.; ASANOME, C. R.; LUDOVICO, J. Um estudo de simulação para avaliação do comportamento do modelo de Mitscherlich. **Semina**. Londrina, v. 10, p. 221-224, 1989.

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

RATKOWSKY, D. A. **Nonlinear regression modeling, a unified practical approach**. New York: Marcel Dekker, 1983. 241p.

RAWLINGS, J. O.; PANTULA, S.G.; DICKEY, D. A. **Applied regression analysis**. 2nd ed. New York: Springer, 1998. 659 p.

RICHARDS, J. F. A flexible growth function for empirical use. **Journal of Experimental Botany**. Oxford, v. 10, p. 290, 1959.

SANTOS, R. V.; RIBEIRO, E. P. Desigualdade de Rendimentos e discriminação por gênero no Brasil em 1999. 2005. 87p. Dissertação (Mestrado em Economia). Porto Alegre, Universidade Federal do Rio Grande do Sul, 2005.

SARMENTO, J. L. R.; REGAZZI A. J.; SOUSA, W. H.; TORRESE, R. de; BREDA, F. C.; MENEZES, G. R. O. de Estudo da curva de crescimento de ovinos Santa Inês. **Revista Brasileira de Zootecnia**. Viçosa, v.35, p. 435-442, 2006.

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

SEARLE, S. R. **Linear Models**. New York: John Wiley & Sons, 1971. 532p.

SEBER, G. A. F.; WILD, C. J. **Nonlinear Regression**. New York: John Wiley & Sons, Auckland, may 1988.

SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete samples). **Bimetrika**. Oxford, v. 52, n. 3/4, p. 591-611, 1965.

SILVA, F. C.; BERGAMASCO, A. F.; VENDITE, L. L. Modelos de Simulação para Análise e Apoio à Decisão em Agrossistemas. **Revista de Biociências**. Taubaté, v. 8, n. 2, p. 7-17, jun./dez. 2002.

SILVA, F. F.; AQUINO, L. H.; OLIVEIRA, A. I. G. Estimativas de parâmetros genéticos de curva de crescimento de gado Nelore (*Bos indicus*). **Ciência e Agrotecnologia**. Lavras, Edição Especial, p. 1562-1567, dez. 2002.

SILVA, F. L.; ALENCAR, M. M. de; FREITAS, A. R. de; PACKER, I. U.; MOURÃO, G. B. Curvas de crescimento em vacas de corte de diferentes tipos biológicos. **Pesquisa Agropecuária Brasileira**. Brasília, v. 46, n. 3, p. 262-271, 2011.

SILVA, S. R.; OLIVEIRA, J. C. de; STUCHI, E. S.; REIFF, E. T. Qualidade e maturação de tangerinas e seus híbridos em São Paulo. **Revista Brasileira de Fruticultura**. Jaboticabal-SP, v. 31, n. 4, p. 977-986, dez. 2009.

SILVEIRA, N. S. S.; MICHEREFF, S. J.; MARIANO, R. L. R.; TAVARES, L. A.; MAIA, L. C. Influência da temperatura, período de molhamento e concentração do inóculo de fungos na incidência de podridões pós-colheita em frutos de tomateiro **Fitopatologia Brasileira**. Brasília, v. 26, mar. 2001.

SOUZA, E. M.; MUNIZ, J. A.; MARCHI, G.; GUILHERME, L. R. G. Modelagem não linear da extração de zinco de um solo tratado com lodo de esgoto **Acta Scientiarum, Technology**. Maringá, v. 32, n. 2, p. 193-199, 2010.

SOUZA, F. V. **Curva de crescimento e exportação de nutrientes e sódio por frutos de mangueira palmer, haden e tommy atkins**. 2007. 54p. Dissertação (Mestrado em Agronomia) - Faculdade de Ciências Agrárias e Veterinária “Júlio de Mesquita Filho”, Universidade Estadual Paulista, Jaboticabal, 2007.

SOUZA, G. S. **Introdução aos modelos de regressão linear e não-linear**. Brasília: Embrapa, 1998. 505p.

SOUZA, M. D.; PERES FILHO, O.; CALDEIRA, S. F.; PELISSARI, A. L.; DORVAL, A. Análise de agrupamento e regressão não-linear aplicados ao crescimento *in vitro* de *Leucoagaricus gongylophorus* (Singer) Möller em meios de cultura acrescido com diferentes extratos vegetais. **Revista Biotermas Brasil**: Cuiabá-MT, v. 22, n. 4, dez. 2011, 85p.

TERRA, M. F.; MUNIZ, J. A.; SAVIAN, T. V. Ajuste de modelos não-lineares aos dados de crescimento de frutos de tamareira-anã *Phoenix roebelenii* O'BRIEN. **Magistra**. Cruz das Almas - BA, v. 22, n.1, p. 01-07, jan./mar., 2010.

TRENTO, E. J.; SEPULCRI, O.; MORIMOTO, F. Comercialização de frutas, legumes e verduras **EMATER**. Curitiba - PR, 2011. 40p.

VIEIRA JUNIOR, P. A.; SILVA, W. R.; VILLA NOVA, N. A.; DIAS, C. T. dos S.; SMIDERLE, O. J.; DOURADO NETO, D.; CICERO, S. M. Estimativas do potencial total de água em sementes usando modelos não-lineares. **Revista Brasileira de Sementes**. Brasília, v. 21, n. 1, p. 87-92, 1999.

VIEIRA, S. **Análise de Variância (ANOVA)**. São Paulo: Atlas, 2006. 216p.

VON BERTELANFY, L. Quantitative laws in metabolism and growth. **The Quartely Review of Biology**. Los Angeles, v. 32, p. 217-230, 1957.

APÊNDICES

APÊNDICE A - Espaço e Subespaço Vetorial

A fim de compreender o cálculo dos resíduos projetados, vejamos as definições de espaço e subespaço vetorial dadas por Callioli, Domingues e Costa (1997).

Definição 1: Um conjunto $V \neq \emptyset$ é um espaço vetorial sobre $\mathbb{R}$ quando, e somente quando

I) Existe uma adição $(u, v) \mapsto u + v$ em V em que

1. $u + v = v + u, \forall u \in V$;
2. $u + (v + w) = (u + v) + w, \forall u, v, w \in V$;
3. $\exists$ em V um elemento neutro: $\exists 0 \in V; u + 0 = u, \forall u \in V$;
4. $\forall u$ de V $\exists$ o oposto indicado por $(-u)$: $\forall u \in V, \exists (-u) \in V; u + (-u) = 0$.

II) Está definida a multiplicação de $\mathbb{R} \times V$ em V , o que implica que (α, μ) de $\mathbb{R} \times V$ está associado a um unico elemento de V , αu .

1. $\varphi(\omega u) = (\varphi\omega)u$;
2. $(\varphi + \omega)u = \varphi u + \omega u$;
3. $\varphi(u + v) = \varphi u + \varphi v$;
4. $1u = u$.

Definição 2: Seja V um espaço vetorial sobre $\mathbb{R}$. Um subespaço vetorial de V é um subconjunto de $W \subset V$ tal que,

1. $0 \in W$;
2. $\forall u, v \in W, (u + v) \in W$;
3. $\forall \alpha \in \mathbb{R}$ e $\forall u \in W, \alpha u \in W$.

Dizemos que dois vetores são ortogonais se, e somente se, seu produto escalar é zero, ou seja, dados os vetores $\mathbf{u}$ e $\mathbf{v}$, eles serão ortogonais quando $\mathbf{u}^T \mathbf{v} = 0$

ANEXOS

ANEXO A - Banco de dados

Tabela 14 - Valores dos comprimentos de frutos de peras (mm) ao longo do tempo (dias)

x	14	28	42	56	70	84	98	112	126	140	154
y ₁	18,06	20,28	31,76	44,44	49,22	50,43	57,59	67,45	66,98	65,11	68,50
y ₂	10,51	25,89	28,11	37,01	43,26	54,17	61,03	64,10	71,49	68,44	68,55
y ₃	11,70	21,57	28,20	39,60	45,72	46,83	58,35	65,09	65,74	69,89	64,68
y ₄	14,88	21,03	28,65	37,35	46,24	58,67	62,83	61,08	63,50	60,43	72,18
y ₅	6,11	21,80	27,83	38,66	49,94	49,22	61,51	62,40	67,74	71,20	68,47
y ₆	15,16	17,45	28,55	35,62	44,74	59,17	62,42	62,94	67,50	68,48	65,75
y ₇	10,86	20,74	20,97	35,45	42,93	58,44	64,79	67,61	67,49	67,60	67,54
y ₈	14,38	12,85	31,08	33,55	48,82	56,05	59,87	61,57	65,99	67,09	73,74
y ₉	12,67	22,59	30,90	42,16	51,69	53,44	55,02	58,73	62,18	70,79	65,49
y ₁₀	15,95	14,76	32,38	39,44	46,15	54,58	64,29	66,11	59,95	67,76	71,06
y ₁₁	17,11	19,63	24,87	38,19	48,14	55,37	60,81	67,74	63,24	72,38	64,79
y ₁₂	11,80	23,90	26,17	36,95	47,09	51,52	59,09	63,94	67,87	64,71	67,33
y ₁₃	5,68	19,20	28,04	37,83	46,94	57,40	61,38	67,29	61,32	64,91	74,71
y ₁₄	9,54	20,72	27,85	35,03	39,30	52,74	64,67	62,41	67,83	69,94	71,02
y ₁₅	16,61	21,48	30,64	37,37	48,58	55,50	61,41	68,71	66,37	71,66	71,37
y ₁₆	13,05	20,21	21,30	40,95	47,51	59,38	62,71	58,19	66,88	67,79	69,76
y ₁₇	13,95	18,07	17,76	34,59	47,79	59,25	63,32	63,63	68,59	68,38	67,51
y ₁₈	17,61	15,82	27,65	38,17	52,09	50,48	62,05	64,74	69,73	72,58	69,36
y ₁₉	15,02	18,32	22,48	37,78	44,10	60,87	62,25	59,94	68,19	71,57	67,39
y ₂₀	10,45	20,83	29,86	39,13	50,77	51,61	63,97	58,12	64,92	65,20	68,94

ANEXO B - Programação no *software* R

A seguir está apresentada a programação para análises envolvendo o modelo logístico. Para o ajuste do modelo Gompertz, o procedimento é similar.

```
rm(list=ls(all=TRUE))

#####
## Análise exploratória dos dados
#####
comprimento<-read.table("dados.txt",h=T); comprimento
a=comprimento$dias;a #dias em que ocorreram as observações
b=comprimento$y;b # variável resposta

## Diagrama de dispersão
plot(a,b,xlab="Tempo",ylab="Comprimento",main='(a)')

## Gráfico boxplot
par(mfrow=c(1,1),pty="s")
boxplot(b~a)

#####
## Ajuste do Modelo Logístico
#####

## Obtenção de valores iniciais meio da linearização do modelo logístico
inlog <- function(x,y){
A0=max(y)+1
yl=log((A0/y)-1)
modl=lm(yl~x)
summary(modl)
B0=as.numeric(coef(modl)[1])
g=as.numeric(coef(modl)[2])
K0=as.numeric(-1*g)
vinlog=c(A0,B0,K0)
return(vinlog)
}

## Obtenção dos valores iniciais
A1=inlog(a,b)[1];A1
B1=inlog(a,b)[2];B1
K1=inlog(a,b)[3];K1

## Verificação da existência de auto-correlação residual: ARMA(0,0), ou seja, não
## possui auto-correlação residual
require(nlme)
regr1l=gnls(b~((A)/(1 + exp(B - K*a))), start=c(A=A1,B=B1,K=K1),
correlation=corARMA(form=~a|rept,p=1,q=0),data=comprimento)# verificação da
# autorrelação de ordem 1
res1l=regr1l$residuals
regr2l=gnls(b~((A)/(1 + exp(B - K*a))), start=c(A=A1,B=B1,K=K1),
```

```

correlation=corARMA(form=~a|rept,p=2,q=0),data=comprimento)# verificação da
# autocorrelação de ordem 2

## Resíduos normalizados do modelo para todos os dados
regrl=gnls(b~((A)/(1 + exp(B - K*a))), start=c(A=A1,B=B1,K=K1))
dl=residuals(regrl,"normalized")

## Teste de homogeneidade de variâncias F máx
with(comprimento, tapply(dl,dias,var))
## Fmax=4,69
1.5290255 /0.6204245 #=2.464483< Fmáx. Portanto os resíduos possuem
# homogeneidade de variâncias

## Teste de normalidade para os resíduos
shapiro.test(dl)
pl=predict(regrl)

#####
## Resíduos Projetados
#####
rproject=function(model,v)
{
V=attr(eval(v),"gradient")
W=attr(eval(v),"hessian")
pm=length(coef(model))
n=length(y)
Wn=matrix(W,n*pm,pm)
H=V%%solve(crossprod(V))%%t(V)
ro=summary(model)$resid
I=diag(n)
xij=matrix(0,n,1)
T=matrix(0,n,pm*pm)
for (j in 1:(pm*pm))
{
l=(1+((j-1)*n)):(n+((j-1)*n))
xij=W[l]
T[,j]= as.vector(xij)
}
(I-H)%%T
Tc=cbind(T[,5],T[,6],T[,9])
S=(I-H)%%Tc
H1=S%%solve(crossprod(S))%%t(S)
P=cbind(V,S)
H2=P%%solve(crossprod(P))%%t(P)
sig_2=(t(ro)%%(I-H2)%%ro)/sum(diag(I-H2))
sig=sqrt(sig_2)
pr=((I-H2)%%ro)/(sig*diag(sqrt(I-H2)))
m0 = lm(y~-1+V)
par(mfrow=c(2,2), mar=c(5.1,4.1,4.1,2.1),pty="s")
plot(m0,pty="s")
mtext("Residual Analysis for Nonlinear Regression Models",

```

```

    outer=TRUE, line=-2, cex=1.4)
return(pr)
}

#####
## Média dos dados
#####
x=c(14, 28, 42, 56, 70, 84, 98, 112, 126, 140, 154)
# x dias em que ocorreram as observações
y=c(13.06, 19.86, 27.25, 37.96, 47.05, 54.76, 61.47, 63.59, 66.17, 68.29, 68.91)
# y média da variável resposta

#####
# Modelo Logístico para a média dos dados
#####

## Valores iniciais para ajuste do modelo
Alm=inlog(x,y)[1];Alm
Blm=inlog(x,y)[2];Blm
Klm=inlog(x,y)[3];Klm

## Diagrama de dispersão e ajuste do modelo logístico
regrlm=nls(y~(A/(1+exp(B-K*x))),start=c(A=Alm,B=Blm,K=Klm),trace=T)
summary(regrlm)
QME1=summary(regrlm)$sigma^2;QME1
plot(x, y, xlab="Tempo", ylab="Comprimento",pch=19)
curve((coef(regrlm)[1])/(1+exp((coef(regrlm)[2])-(coef(regrlm)[3])*x)),0,160,
add=TRUE,col="red")
## Intervalo de confiança
yesplm=predict(regrlm,newdata=list(x=seq(0,160,length=1000)))
upiclm=yesplm+qt(0.975,summary(regrlm)$df[2])*summary(regrlm)$sigma
loiclm=yesplm-qt(0.975,summary(regrlm)$df[2])*summary(regrlm)$sigma
lines((upiclm~seq(0,160,length=1000)),col="blue",lty=2)
lines((loiclm~seq(0,160,length=1000)),col="blue",lty=2)
legend("bottomright",legend=c('modelo logístico','IC modelo logístico'),
col=c('red','blue'),lty=c(1,2))

# Histogramas para os resíduos da média dos dados
reglm=gnls(y~((A)/(1 + exp(B - K*x))), start=c(A=Alm,B=Blm,K=Klm))
d1m=residuals(reglm,"normalized")
plm=predict(reglm)
xx=seq(-4,4,0.01)
hist(d1m,prob=T,main='(a)',xlab='',ylab="densidade")
lines(xx,dnorm(xx,mean(d1m),sqrt(var(d1m))),col="blue")
points(d1m,rep(0,11),col="red",pch=19)

## Gráfico de normalidade dos resíduos para a média dos dados
require(car)
par(mfrow=c(1,1),pty='s')
qqPlot(d1m,xlab='quantis normais',ylab='resíduos',main='(b)')

```

```

# Teste de normalidade para os resíduos
shapiro.test(dlm)

# Utilização das estimativas como valores paramétricos
Apl=as.numeric(coef(regrlm)[1]);Apl
Bpl=as.numeric(coef(regrlm)[2]);Bpl
Kpl=as.numeric(coef(regrlm)[3]);Kpl
p<-length(coef(regrlm))
n<-length(y)

# Primeira e segunda derivada em relação ao vetor de parâmetros
ul<- expression(Apl/(1+exp(Bpl-(Kpl*x))))
ul<-deriv(ul, c("Apl","Bpl","Kpl"), hessian = TRUE)
Xl<-attr(eval(ul),"gradient")
Xlt=t(Xl)
XltXl<- crossprod(Xl)
Wl<-attr(eval(ul),"hessian")
Wnl<-matrix(Wl,n*p,p)
Hl<-Xl%*%solve(XltXl)%*%Xlt
hl<-diag(Hl)
rl<-summary(regrlm)$resid
sl<-summary(regrlm)$sigma
tl<-rl/(sl*sqrt(1-hl))

# p matrizes n x p
# matriz W com dimensão (n*p) x p
# Matriz de projeção
# Resíduo ordinário
# Desvio padrão
# Resíduo estudentizado

# Análise de diagnóstico
k=data.frame(x=c(14, 28, 42, 56, 70, 84, 98, 112, 126, 140, 154),
y=c(13.06, 19.86, 27.25, 37.96, 47.05, 54.76, 61.47, 63.59, 66.17, 68.29, 68.91) )
n0=nls(y~(A/(1+exp(B-K*x))),start=c(A=Alm,B=Blm,K=Klm),trace=T,data=k)
summary(regrlm)
m0 = lm(y~1+Xl,data=k)
cbind(fitted(m0), fitted(n0))      # valores ajustados
cbind(residuals(m0), residuals(n0)) # valores preditos
c(summary(m0)$sig, summary(n0)$sig) # estimativa do desvio padrão residual
vcov(m0); vcov(n0)                 # matriz de covariância das estimativas
anova(m0)
anova(n0, lm(y~1, k))

# Gráfico dos resíduos normalizados
plot(x,dlm,pch=19,xlab="Tempo",ylab="Resíduo normalizado",ylim=c(-3,3),main='(a)')
abline(-2,0,lty=2, col="red", lwd=2)
abline(2,0,lty=2, col="red", lwd=2)

# Resíduos projetados para o modelo logístico
rprl=rproject(regrlm,ul)

# Gráfico dos resíduos projetados
plot(x,rprl, pch=19,xlab="Tempo", ylab="Resíduo projetado",ylim=c(-3,3),main='(b)')
abline(-2,0,lty=2, col="red", lwd=2)
abline(2,0,lty=2, col="red", lwd=2)

# Teste de normalidade para os resíduos projetados

```

```

shapiro.test(rpr1)

# Envelope simulado
qqPlot(rpr1, ylab='Tempo',xlab='Quantis Normais')

# F calculado para medida de curvatura
2/sqrt(4.07)

## Distância de Cook
D<-(t1^2/p)*(h1/(1-h1))

## Gráfico da distância de Cook
plot(D, pch=19, xlab="Índice", ylab="Distância de Cook",main='(a)')
identify(D, cex=1, col="darkblue")

## Gráfico dos elementos da diagonal da matriz de projeção H
plot(h1, xlab="Índice", ylab="Diagonal de H", pch=19, main="(b)", ylim=c(0,0.6))
abline((2*p)/n,0,lty = 2, lwd = 2, col="red")

#####
## Simulação para os resíduos projetados
#####

nexp=1000 ## Número de simulações
m=length(seq(14,154,14)) # Número de observações
y1=matrix(0,nexp,m)
x1=matrix(sort((rep((seq(14,154,14)),nexp))),nexp,m)
e=matrix(0,nexp,m)
set.seed(23456) #fixando semente

## Simulação dos dados
for(j in 1:nexp)
{
  for(i in 1:m)
  {
    e[j,i]= rnorm(1,0,sqrt(QME1))
    y1[j,i]=(Ap1/(1+exp(Bp1-(Kp1*x1[j,i]))))+e[j,i]
  }
}

# Medidas de curvatura para os modelos ajustados de cada amostra
library(MASS)
y3 <- NULL
for(i in 1:nrow(y1)){
  y3[[i]] <- y1[i,]
}
curv1=deriv3(~A/(1+exp(B-(K*x))), c("A","B","K"), function(A,B,K,x)NULL)
reg1<-NULL
curv<-NULL
for(i in 1:nrow(y1))
{

```

```

regl[[i]]=nls(y3[[i]]~curv1(A,B,K,x),start=list(A=Alm,B=Blm,K=Klm))
curv[[i]]=rms.curv(regl[[i]])
}
curvmat <- matrix(0,1000,2)
for(i in 1:length(curv)){
  curvmat[i,1] <- as.numeric(curv[[i]][3])
  curvmat[i,2] <- as.numeric(curv[[i]][4])
}
ordcurv=apply(curvmat,2,sort)
mincurv=apply(ordcurv,2,min)
meancurv=apply(ordcurv,2,mean)
maxcurv=apply(ordcurv,2,max)
q025=apply(ordcurv,2,quantile,prob=0.025,na.rm=TRUE)
q50=apply(ordcurv,2,quantile,prob=0.5,na.rm=TRUE)
q975=apply(ordcurv,2,quantile,prob=0.975,na.rm=TRUE)

## Matriz contendo os valores de X, Y e identificação da amostra simulada
aux=data.frame(cbind(rep(seq(14,154,14),nexp),matrix(t(y1),m*nexp,1),
sort(rep(seq(,nexp),m))))

## Função para verificar o percentual de convergência
nconver0 =
sapply(1:nexp,
function(qual, silent=FALSE){
da = subset(aux, X3==qual)
n0 = try(nls(X2~A/(1+exp(B-(K*X1))), data=da,
start=list(A=Alm, B=Blm, K=Klm), trace=FALSE),silent=silent) # se n0 não nulo,
# houve convergência
if(class(n0)=="nls") c(1, coef(n0)) else rep(NA, 4)
}
)

## Porcentual de convergência
pconv_y1=100*sum(!is.na(nconver0[1,]))/nexp;pconv_y1
eq_y1 =apply(nconver0[2:4,!is.na(nconver0[1,])], 2,
function(x1){(c(x1)-c(Ap1, Bp1, Kp1))^2})

## Erro quadrático médio
eqm_y1 = apply(eq_y1, 1, mean);eqm_y1

## Matriz com as estimativas dos parâmetros em cada amostra simulada
vfl=nconver0[2,]
kfl=nconver0[3,]
lfl=nconver0[4,]
dadosl=nconver0[2:4,]

## Excesso de variância
varvf<-vcov(regrlm)[1,1];varvf
varkf<-vcov(regrlm)[2,2];varkf
varlf<-vcov(regrlm)[3,3];varlf
dados<-data.frame(vfl,kfl,lfl)

```



```

varvs<-var(dados)[1,1];varvs
varks<-var(dados)[2,2];varks
varls<-var(dados)[3,3];varls
exvarv<-((varvs-varvf)/(varvf))*100;exvarv
exvark<-((varks-varkf)/(varkf))*100;exvark
exvarl<-((varls-varlf)/(varlf))*100;exvarl

## Porcentagem de viés
vs<-coef(regrlm)[1];vs
ks<-coef(regrlm)[2];ks
ls<-coef(regrlm)[3];ls
medvs<-mean(dados$vf1);medvs
medks<-mean(dados$kf1);medks
medls<-mean(dados$lf1);medls
exviv<-((medvs-vs)/vs)*100;exviv
exvik<-((medks-ks)/ks)*100;exvik
exvil<-((medls-ls)/ls)*100;exvil

## teste do viés
zv<-((medvs-vs)/sqrt(varvs/nexp));zv
zk<-((medks-ks)/sqrt(varks/nexp));zk
zl<-((medls-ls)/sqrt(varls/nexp));zl
pnorm(zv,0,1,lower.tail=T)
pnorm(zk,0,1,lower.tail=T)
pnorm(zl,0,1,lower.tail=T)

## teste da var
quiv<-((nexp-1)*varvs)/varvf;quiv
quik<-((nexp-1)*varks)/varkf;quik
quilt<-((nexp-1)*varls)/varlf;quilt

## aproximação normal
zqv<-sqrt(2*quiv)-sqrt(2*(nexp-1)-1);zqv
zqk<-sqrt(2*quik)-sqrt((2*(nexp-1))-1);zqk
zql<-sqrt(2*quilt)-sqrt((2*(nexp-1))-1);zql
pnorm(zqv,0,1,lower.tail=F)
pnorm(zqk,0,1,lower.tail=F)
pnorm(zql,0,1,lower.tail=F)

## Histogramas do vetor de parâmetros ajustados a cada amostra simulada
hist(vf1,freq=F,main="",nclass=20,col="lightblue",border='lightblue',
xlab='alpha',ylab='densidade')
hist(kf1,freq=F,main="",nclass=18,col="lightblue",border='lightblue',
xlab='beta',ylab='densidade')
hist(lf1,main="",prob=T,nclass=20,col="lightblue",border='lightblue',
xlab='gamma',ylab='densidade')

## Teste de normalidade para cada um dos vetores de parâmetros
shapiro.test(vf1)
shapiro.test(kf1)
shapiro.test(lf1)

```

```

## Obtenção dos resíduos projetados para cada amostra simulada
nrpl<-list()
for(k in 1:nexp) {
  ul<- expression(Apl/(1+exp(Bpl-(Kpl*x))))
  ul<-deriv(ul, c("Apl","Bpl","Kpl"), hessian = TRUE)
  Apl <- vfl[k]; Bpl <- kfl[k]; Kpl <- lfl[k]
  Xl<-attr(eval(ul),"gradient")
  Xlt=t(Xl)
  XltXl<- crossprod(Xl)
  Wl<-attr(eval(ul),"hessian") # p matrizes n x p
  Wnl<-matrix(Wl,n*p,p) # matriz W com dimensão (n*p) x p
  ## Matriz de projeção
  H<-Xl%%solve(XltXl)%%Xlt
  I=diag(n)
  xij<-matrix(0,n,1)
  T<-matrix(0,n,p*p)
  for (j in 1:(p*p))
  {
    l=(1+((j-1)*n)):(n+((j-1)*n))
    xij<-Wl[l]
    T[,j]= as.vector(xij)
  }
  Tf=cbind(T[,5],T[,6],T[,9])
  S<-(I-H)%%Tf
  St=t(S)
  StS<- crossprod(S)
  H1<-S%%solve(StS)%%St
  V<-cbind(Xl,S)
  Vt=t(V)
  VtV<- crossprod(V)
  H2<-V%%solve(VtV)%%Vt
  trH2<-sum(diag(I-H2))
  sigma2=(t(r1)%%(I-H2)%%r1)/trH2
  sigma=sqrt(sigma2)
  rpl<-((I-H2)%%r1)/(sigma*diag(sqrt(I-H2)))
  nrpl[[k]]<-rpl
}

## Data frame com todos os 1000 resíduos projetados
rpl_sim <- nrpl[[1]]
for(i in 2:nexp) rpl_sim <- cbind(rpl_sim, nrpl[[i]])

## Ordenação da matriz de resíduos projetados em cada uma das amostras simuladas
ordrpl=apply(rpl_sim,2,sort)
minresl=apply(ordrpl,1,min)
meanresl=apply(ordrpl,1,mean)
maxresl=apply(ordrpl,1,max)
q025=apply(ordrpl,1,quantile,prob=0.025,na.rm=TRUE)
q50=apply(ordrpl,1,quantile,prob=0.5,na.rm=TRUE)
q975=apply(ordrpl,1,quantile,prob=0.975,na.rm=TRUE)

```

```
## Data frame com as estatísticas
mmrpl=cbind(minresl,q025,meanresl,q50, q975, maxresl);mmrpl

## Envelope simulado
i=seq(1,11,1);i
n=length(i);n
ppi=(i-0.375)/(n+0.25);ppi
zi=qnorm(ppi,0,1)

## Grafico de envelope simulado
plot(zi,sort(rprl),ylim=c(-2,2),pch=19,xlab='',ylab='resíduos ordenados')
lines(zi,meanresl,col="green",lwd=1)
lines(zi,q50, col="blue",lwd=1,lty=2)
lines(zi,q025,lwd=1)
lines(zi,q975,lwd=1)
lines(zi,minresl,col="red",lwd=1,lty=2)
lines(zi,maxresl,col="red",lwd=1,lty=2)
legend("bottomright",legend=c('média','quantil 50','quantis 25 e 75',
'mínimo e máximo'),col=c('green','blue','black','red'),lty=c(1,2))
```