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

O modelo logístico com erros assimétricos e heterocedásticos
aplicado a dados de altura do milho

Rick Anderson Freire Mangueira

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

Piracicaba
2015

Rick Anderson Freire Manguiera
Estatístico

**O modelo logístico com erros assimétricos e heterocedásticos aplicado a dados
de altura do milho**

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 Mestre
em Ciências. Área de concentração: Estatística e Experi-
mentação Agronômica

**Piracicaba
2015**

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

Mangueira, Rick Anderson Freire

O modelo logístico com erros assimétricos e heterocedásticos aplicado a dados de altura do milho / Rick Anderson Freire Mangueira. - - versão revisada de acordo com a resolução CoPGr 6018 de 2011. - - Piracicaba, 2015.

79 p. : il.

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

1. Modelos não lineares 2. Curvas de crescimento 3. Assimetria dos erros I. Título

CDD 633.15
M277m

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

DEDICATÓRIA

*Aos Meus pais Sérgio Manguiera e Mauricéia Freire,
aos meus avós Socorro Manguiera e Dinaldo Bezerra, e Aderval Chargas e Dorinha,
e a toda minha família.*

AGRADECIMENTOS

A Deus pela vida, pela saúde e pela força para concluir esta etapa de minha vida.

A minha família pelo incentivo e investimento em minha educação desde os primeiros anos colegiais.

A Maria Joseane, que foi meu primeiro incentivo para iniciar esta jornada junto a ESALQ/USP.

Aos professores Doutores Ana Patrícia Bastos Peixoto e Tiago Almeida de Oliveira, que me receberam com bastante carinho na cidade de Piracicaba-SP nos primeiros dias de mestrado.

Aos colegas Elias Medeiros e Djair Durand pelo acolhimento também nos primeiros dias na cidade.

Aos colegas de mestrado e de disciplinas Andreia, Bruna Wendpap, Daniel Miquelluti, Douglas, Erasnilson Vieira, Fernando Mayer, Gislaine Vieira, Otávio Barros, Patrícia Araripe, Tatiana Assis e Valiana Teodoro, pelos estudos coletivos e trocas de conhecimento ao longo do ano.

Aos professores do Programa de Pós Graduação em Estatística e Experimentação Agrônômica, ESALQ/USP, pelo conhecimento compartilhado, convívio e por fazerem parte de minha formação profissional.

A minha orientadora, Dra. Taciana Villela Savian, pelos esclarecimentos, tardes de estudo compartilhadas e condução deste trabalho.

Aos funcionários do LCE/ESALQ/USP pela competência e eficiência em seus trabalhos, que facilitam nosso dia-a-dia.

Aos colegas Jacob Crosariol Netto e Daniela Viana da FCAV/UNESP, pela amizade e disponibilidade dos dados utilizados nesse trabalho.

Aos colegas de pós-graduação José Nilton e Altemir Braga, pelas tardes de consultorias e discussões a cerca da estatística.

Aos colegas de pós graduação Adriale Biase, Alessandra Santos, Ana Júlia, Cristian Villegas, Davi Butturi, Everton Batista, Elisângela Oliveira, Elizabeth Hashimoto, Iuri Emmanuel, Lucas Santana, Lúcio de Araújo, Luiz Nakamura, Marcello Neiva, Marcus Maverick, Maria Cristina, Mírian Araújo, Natália Martins, Pedro Cerqueira, Pedro Amодо, Rafael Maia, Rafael Moral, Reginaldo Hilário, Renan Pinto, Ricardo Klein, Rodrigo

Pescim, Simone Grego, Thiago Gentil, Thiago Oliveira e Tiago Santana, pelas conversas descontraídas de salinha e cozinha, pelas noites de counter strike e pizzas, pelas sextas de futebol, noitadas de bar e baladas, churrascos, almoços e jantas.

A Gabriela Oioli pelo amor e carinho.

A Coordenação de Aperfeiçoamento de Pessoal de nível Superior - CAPES pelo apoio financeiro concedido por meu da bolsa de estudos.

SUMÁRIO

RESUMO	9
ABSTRACT	11
LISTA DE FIGURAS	13
LISTA DE TABELAS	15
1 INTRODUÇÃO	17
2 REVISÃO BIBLIOGRÁFICA	19
2.1 A cultura do milho no Brasil	19
2.2 Estudo do crescimento	21
2.3 Modelos de regressão	23
2.4 Modelos não lineares	24
2.5 Modelo Empírico	24
2.6 Resíduos heterocedásticos	25
2.7 Resíduos autocorrelacionados	26
2.8 Estimação de parâmetros de modelos não lineares	27
2.8.1 Método de mínimos quadrados ordinários	27
2.8.2 Método de mínimos quadrados ponderados	29
2.8.3 Método de mínimos quadrados generalizados	31
2.8.4 Estimação por máxima verossimilhança	32
2.9 Método Iterativo	34
2.9.1 Método Iterativo de Gauss-Newton	34
2.10 Distribuições assimétricas	35
2.10.1 Distribuição normal assimétrica(SN)	36
2.10.2 Distribuição normal assimétrica com parâmetros de locação e escala	37
2.10.3 Modelo não linear normal assimétrico	38
2.10.4 Distribuição t-Student assimétrica (ST)	39
2.10.5 Distribuição t-student assimétrico com parâmetros de locação e escala	41
3 METODOLOGIA	43
3.1 Material	43
3.2 Métodos	43

3.3	Análise da homogeneidade de variância dos erros - Teste de Hartley	44
3.4	Análise da normalidade	44
3.5	Valores iniciais	45
3.6	Seleção do modelo	46
3.7	Software estatístico	47
4	RESULTADOS E DISCUSSÃO	49
4.1	Análise exploratória	49
4.2	Ajuste modelo logístico considerando erros simétricos normais	52
4.3	Ajuste do modelo logístico com erros com distribuições assimétricas	56
4.3.1	Ajuste modelo logístico considerando erros normais assimétricos (Skew-normal)	57
4.3.2	Ajuste modelo logístico considerando erros t-student assimétricos (Skew-t) . . .	61
4.4	Escolha do modelo	64
5	CONCLUSÃO	67
	REFERÊNCIAS	69
	ANEXOS	74

RESUMO

O modelo logístico com erros assimétricos e heterocedásticos aplicado a dados de altura do milho

A produção de milho tem uma grande importância para o país. Ter o conhecimento sobre o crescimento da planta é de fundamental importância para seu manejo. Pode-se obter esse conhecimento fazendo um estudo por meio de modelos de crescimento, para se obter informações por meio de parâmetros com interpretações biológicas que trazem consigo um resumo sobre a curva característica do crescimento da planta, que podem ajudar no planejamento da cultura e principalmente conhecer qual período a planta mais cresce, a época mais adequada para adubação e realização do controle de pragas. Considerar características não comuns de suposições do modelo pode dar maior confiabilidade nos resultados do ajuste, como por exemplo a consideração da heterocedasticidade e não normalidade residual. Sendo assim, esse trabalho teve o objetivo de ajustar o modelo logístico considerando a heterocedasticidade e diferentes distribuições para o erro como normalidade, assimetria normal e assimetria t-student, a dados da altura da planta do milho do híbrido transgênico 30F35 Y (Yieldgard), observados ao longo do tempo. O experimento foi realizado no município de Votuporanga-SP, em área experimental do Pólo Regional Noroeste Paulista da APTA (Agência Paulista de Tecnologia dos Agro-negócios), no ano agrícola 2011/2012. A primeira medição da altura da planta de milho foi realizada 15 dias após a semeadura. As medições seguintes ocorreram com 30, 40, 50 e 122 dias, respectivamente, após a semeadura. Em cada dia de avaliação foi medido a altura das plantas em centímetros, com auxílio de uma régua, sendo esta medida da base da planta (solo) até o ápice da última folha expandida do cartucho. Toda a análise foi realizada utilizando o software R. Todos os modelos considerados se ajustaram bem a curva de crescimento do híbrido transgênico 30F35 Y (Yieldgard), porém o modelo logístico considerando erros normais assimétricos foi escolhido como mais adequado para modelar a curva, com base nos avaliadores utilizados.

Palavras-chave: Modelos não lineares; Curvas de crescimento; Assimetria dos erros

ABSTRACT

Logistic model with skewed and heteroskedastic errors applied to maize height data

Maize production is of great importance for the country. Knowing the plant development is of major importance to its management. Such knowledge may be attained through growth curves studies, to obtain information through parameters with biological interpretation which are able to synthesize the plant's growth curve. This synthesis may help in management issues, such as information on the period of most rapid growth, best time to apply fertilizers and control pests. Considering uncommon features of the model assumptions may provide greater reliability on the results of the fitted model, such as residual heteroscedasticity and non-normality. In that sense, this work aimed to fit the logistic model considering variance heterogeneity and different error distributions such as normal, skew-normal and skew-t, to the transgenic hybrid 30F35Y maize height data through time. The experiment was conducted in the municipality of Votuporanga-SP, in an experimental station of the Pólo Regional Noroeste Paulista da Agência Paulista de Tecnologia dos Agro-Negócios (APTA) during the agronomic year of 2011/2012. The first height data measurement was taken 15 days after sewing. The following measurements were taken at 30, 40, 50 and 122 days after sewing. Each day the plant height was measured in centimeters using a ruler, measuring from the plant base (soil) until the edge of the last expanded leaf. All analyses were carried out using software R. All considered models fitted the data reasonably well, however the logistic model considering skew-normal errors was chosen as most adequate to model the data, basing on the goodness-of-fit tests.

Keywords: 1. Nonlinear models; Growth curves; Asymmetrical errors

LISTA DE FIGURAS

Figura 1 - Plântula (Fonte: (RITCHIE; HANWAY; BENSON, 2003))	20
Figura 2 - Gráfico do comportamento da distribuição normal assimétrica	38
Figura 3 - Gráfico do comportamento da distribuição t-student assimétrica	41
Figura 4 - Gráfico de dispersão dos dados de altura de plantas de milho ao longo do tempo	49
Figura 5 - Gráfico boxplot da altura de pés de milho (cm) em função do tempo (dias)	50
Figura 6 - Gráfico da altura de pés de milho (cm) de cada planta em função do tempo (dias)	52
Figura 7 - Taxa de crescimento instantânea (TCI) em cm/dia - modelo logístico normal	54
Figura 8 - Gráfico de probabilidades normais com envelope simulado com 95% de confiança	55
Figura 9 - Gráfico dos resíduos padronizados	55
Figura 10 - (a) Histograma dos resíduos ordinários do modelo (b) Histograma de uma distribuição normal com mesma média e variância dos resíduos ordinários do modelo	56
Figura 11 - Taxa de crescimento instantânea (TCI) em cm/dia - modelo logístico nor- mal assimétrico	59
Figura 12 - Histograma dos resíduos do modelo logístico com erros normais assimétricos	60
Figura 13 - Taxa de crescimento instantânea (TCI) em cm/dia - modelo logístico t- student assimétrico	62
Figura 14 - Histograma dos resíduos do modelo logístico com erros t-student assimétricos	63
Figura 15 - Ajuste do modelo logístico	64
Figura 16 - Ajuste do modelo logístico normal assimétrico à altura da planta do milho	66

LISTA DE TABELAS

Tabela 1 - Variância e assimetria em cada dia de observação	50
Tabela 2 - Estimativas dos parâmetros do modelo logístico e seus respectivos erros padrão	53
Tabela 3 - Estimativas dos parâmetros do modelo logístico e seus respectivos erros padrão considerando erros com distribuição normal assimétrica	58
Tabela 4 - Estimativas dos parâmetros do modelo logístico e seus respectivos erros padrão considerando erros com distribuição t-student assimétrica	62
Tabela 5 - Critério de informação de Akaike - AIC dos modelos logísticos ajustados .	65

1 INTRODUÇÃO

A produção da cultura do milho no Brasil tem bastante destaque no mundo. A produção nacional é a terceira maior do mundo, e no ano agrícola 2013/2014 produziu aproximadamente 78.554,0 mil toneladas do grão, em uma área plantada de 15.769,1 mil ha (CONAB, 2014). Esse produto é utilizado para alimentação humana e animal, e também vem sendo usado como fonte de bioenergia.

Segundo Duarte et al. (2011), o consumo humano desse grão provém em sua maioria do processo de industrialização dos processos de moagem úmida e seca. Desses processos resultam subprodutos como a farinha de milho, o fubá, a quirera, farelos, óleo e farinha integral desengordurada, que são utilizadas em diversas receitas do cotidiano. Além disso, também é utilizado em receitas típicas de festas culturais de algumas regiões do país.

Com relação ao consumo animal, o milho é o insumo que na cadeia produtiva de suínos e aves são consumidos aproximadamente 70% do milho produzido no mundo e entre 70 e 80% do milho produzido no Brasil (DUARTE et al., 2011).

Tendo conhecimento destas condições e da importância desse grão, é importante obter o máximo de conhecimento do desenvolvimento e crescimento dessa cultura, o que propicia um planejamento de qualidade para o plantio, e assim, um controle na qualidade do produto final pós colheita.

Uma ferramenta importante que pode ser utilizada para obter conhecimento do crescimento da planta e de seus estádios fenológicos, é o estudo de curvas de crescimento. Esse estudo é feito por meio de modelos não lineares estatísticos que apresentam vantagens como interpretação biológicas para seus parâmetros.

O estudo de curvas de crescimento podem gerar estratégias e planejamento para os plantios futuros da cultura, também possibilita o manejo mais adequado da cultura, como por exemplo, conhecer qual período a planta mais cresce, a época mais adequada para adubação e controle de pragas. Além disso, conforme citado por Fernandes et al. (2014), o estudo de curvas de crescimento por meio de modelos não lineares apresenta uma grande capacidade de sintetização das informações presentes no conjunto de dados, resumindo-as em apenas alguns parâmetros com interpretações práticas.

Atualmente muitos estudos de curvas de crescimentos por meio de modelos não lineares são realizados, tanto na área animal quanto vegetal. Uma informação im-

portante que deve ser levada em consideração nos estudos são as suposições envolvidas na modelagem estatística, suposições estas nos dados e principalmente nos resíduos do modelo a ser ajustado.

O cumprimento das suposições em relação aos resíduos como homogeneidade de variância e normalidade, são importantes para utilização de algumas técnicas e testes estatísticos. Porém, nem sempre essas suposições são alcançadas. Em estudo com medidas ao longo do tempo, por exemplo, os resíduos carregam uma dependência e heterogeneidade de variâncias vindas das medições em seus dias distintos. A normalidade dos resíduos também nem sempre ocorre. Em casos com valores discrepantes podem levar a quebra dessa suposição.

Para contornar esses problemas, de cumprimento de suposição, várias alternativas são descritas na literatura, como por exemplo a transformação de variável, porém essa alternativa acarreta em outras dificuldades, como por exemplo a mudança na interpretação dos parâmetros do modelo e na distribuição residual.

Outra alternativa que pode-se considerar é atribuir diferentes distribuições para os erros. A utilização de outras distribuições já é realidade no estudo de curvas de crescimento. As distribuições assimétricas, como normal assimétrica e t-student assimétrica que são geradas a partir das distribuições normal e t-student respectivamente, podem ser utilizadas para este fim, pois, estas distribuições dão suporte tanto para assimetria quanto para caudas pesadas geradas por valores discrepantes.

Sendo assim, este trabalho foi realizado com o objetivo de descrever a curva de crescimento dos dados da altura da planta do milho do híbrido transgênico 30F35 Y (Yieldgard) ao longo do tempo, por meio do modelo de crescimento logístico, levando em consideração a heterogeneidade de variância e a assimetria residual.

2 REVISÃO BIBLIOGRÁFICA

2.1 A cultura do milho no Brasil

A cultura do milho tem grande importância no cenário mundial, pois, é uma fonte alimentar de carboidratos, proteínas e óleos de baixo custo comercial, sendo utilizado de forma direta na alimentação humana e animal, e atualmente está sendo utilizado como fonte de bioenergia (NETTO, 2013).

A produção no Brasil é dividida em duas épocas de semeadura. A primeira safra, chamada de safra de verão, é realizada durante período chuvoso, outubro/novembro na região Sudeste do país. Já a segunda safra, ou safrinha como é conhecida, é realizada nos meses de fevereiro/março, predominantemente na região Centro-Oeste e nos estados do Paraná e São Paulo.

O Brasil é o terceiro maior produtor de milho do mundo, e juntamente com os Estados Unidos, China e Argentina, representam 70% da produção mundial. A produção nacional, considerando a primeira e segunda safra do ano agrícola 2013/2014, foi de aproximadamente 78.554,0 mil toneladas, com uma área plantada de 15.769,1 mil ha (CONAB, 2014).

O sistema de estádios do desenvolvimento da planta do milho pode ser dividido em vegetativo e reprodutivo. O estágio vegetativo apresenta uma subdivisão que vai da emergência da planta até o pendoamento. Já os estádios reprodutivos acontece do florescimento até a maturidade fisiológica da planta. Mais detalhes a cerca dos estádios da planta são explicados por Ritchie, Hanway e Benson (2003).

O crescimento da planta do milho é dividido em diversas fases distintas de desenvolvimento, desde a etapa inicial de plantio, semeadura, até a maturação e colheita dos grãos.

Conforme descrito por Fancelli e Neto (2000), o ciclo da cultura do milho é caracterizado em cinco etapas de desenvolvimento:

- i) Germinação e emergência: ocorre entre a semeadura e o aparecimento da plântula, variando entre 4 e 12 dias, em função da temperatura e umidade do solo;
- ii) Crescimento vegetativo: iniciado a partir da emissão da segunda folha, até o início do florescimento, com extensão variando em função do genótipo e de fatores climáticos;

- iii) Florescimento: estabelecido entre o início da polinização e o início da frutificação;
- iv) Frutificação: período que acontece entre a fecundação e o enchimento completo dos grãos, e sua duração varia entre 40 e 60 dias;
- v) Maturidade: período compreendido entre o final da frutificação e o aparecimento da “camada preta” no ponto de inserção dos grãos com o sabugo.

A Figura 1 ilustra de forma geral as partes de uma planta de milho (RITCHIE; HANWAY; BENSON, 2003).

Figura 1: Plântula (Fonte: (RITCHIE; HANWAY; BENSON, 2003))

Segundo Ritchie, Hanway e Benson (2003) cada planta de milho gera aproximadamente 20 folhas, floresce cerca de 65 dias após a emergência e atinge a maturidade fisiológica cerca de 125 dias após a emergência.

O plantio da cultura do milho deve ser muito bem planejado, pois determina o início de um processo de cerca de 120 dias que afetará todas as operações envolvidas, além de determinar as possibilidades de sucesso ou insucesso da lavoura (CRUZ et al., 2010).

O interesse de se estudar o crescimento das plantas e entender os processos envolvidos, é de fundamental importância para o planejamento do plantio. Tal estudo pode ser realizado por meio de curvas de crescimento ou modelos não lineares que produzem

uma interpretação biológica em seus parâmetros, e com isso traz conhecimento prévio para futuros plantios.

2.2 Estudo do crescimento

Para Seber e Wild (1989) a análise de dados de crescimento é importante em muitos campos de estudo, como por exemplo na biologia, na agricultura, na química, na medicina, ciências sociais, etc.

Segundo Kshirsagar e Smith (1995), os modelos de crescimento são bastante utilizados na análise de experimentos projetados de modo que as mesmas unidades experimentais são observadas repetidamente.

O estudo da trajetória de crescimento das plantas é de grande importância para se fazer o manejo adequado destas. Esse conhecimento auxilia na preparação de técnicas de cultivo, colheita, conservação e também para a detecção de problemas no desenvolvimento de culturas (PUIATTI et al., 2013). Sendo assim, conhecer o desenvolvimento e crescimento da planta, pode facilitar o manejo da cultura de uma forma geral.

O estudo de modelos de crescimento é uma das alternativas mais comuns de estudar o desenvolvimento de espécies, seja ela animal ou vegetal. São importantes para explicar relações de crescimento e produtividade de plantas, frutos, animais, etc. É possível modelá-lo, pois, há uma relação funcional entre o que se mede e o tempo em que é medido, e ainda, é possível resumir em poucos parâmetros as informações dos dados, pois, os parâmetros apresentam interpretações práticas.

Modelar os processos de crescimento e desenvolvimento vegetal significa realizar uma síntese dos mecanismos físicos e bioquímicos entre planta e o ambiente, representando-os por meio de funções matemáticas (LYRA et al., 2008).

Os estudos de curvas de crescimento são importantes uma vez que podem ser realizados em situações em que a informação contida numa sequência de pontos “tamanho tempo”, do crescimento de uma espécie seja reduzida e sintetizada num conjunto de parâmetros que tenham interpretação prática e preditiva (PRADO; SAVIAN; MUNIZ, 2013).

Lyra et al. (2008) avaliaram o ajuste do modelo logístico às variáveis de crescimento, altura da planta e matéria seca e do modelo exponencial ao índice de área foliar em diferentes épocas de plantio e manejo da cultura do milho, utilizando o coeficiente

de determinação (R^2) para verificar a explicação da variabilidade das variáveis. Já Souza et al. (2007) ajustaram o modelo de crescimento expolinear para analisar o crescimento da cultura do milho. Modelos de crescimento expolinear, logístico e Gompertz também foram utilizados por Moura et al. (2008) em dados do acúmulo de fitomassa seca das culturas do feijão-caupi e do milho.

Calegario et al. (2005) utilizaram os modelos não lineares logístico, Richards, Gompertz e Weibull na descrição do crescimento de *Eucalyptus*, considerando que os dados apresentavam heterocedasticidade e autocorrelação residual. Filho et al. (2008) utilizou o modelo logístico para ajustar curvas de crescimento para os dados de altura de plantas de dois cultivares de feijão, porém utilizando uma abordagem bayesiana na estimação dos parâmetros do modelo.

Os modelos logístico e Gompertz foram utilizados por Terra, Muniz e Savian (2010) para estudar o comprimento e diâmetro do fruto da tamareira-anã (*Phoenix roebelenii* Ot'Brien). Foram utilizados critérios como coeficiente de determinação ajustado (R_{aj}^2), desvio padrão residual (DPR), teste de Durbin-Watson (DW) e critério de informação de Akaike (AIC) para selecionar o melhor modelo. O modelo logístico apresentou o melhor ajuste para os dados de comprimento e diâmetro do fruto. Porém, no estudo não foi abordado sobre heterocedasticidade e normalidade residual.

Fernandes et al. (2014) também compararam os modelos logístico e Gompertz na descrição das curvas de crescimento do fruto do cafeeiro. Porém neste caso, foi considerado também a heterogeneidade de variâncias e realizado teste de normalidade dos resíduos. Os autores utilizaram o coeficiente de determinação ajustado (R_{aj}^2), critério de informação de Akaike (AIC) e medidas de curvatura de Bates e Watts como critérios de seleção do modelo mais adequado. O modelo Gompertz foi considerado o mais adequado, considerando a heterocedasticidade residual.

Em outro estudo com os modelos logístico e Gompertz, desenvolvido por Prado, Savian e Muniz (2013), o modelo logístico foi o mais adequado na descrição do crescimento de frutos de coqueiro da variedade anão verde. Os autores ainda consideraram uma estrutura de termos autoregressivos de ordem 1 e 2 no modelo.

2.3 Modelos de regressão

A análise de regressão tem como objetivo principal verificar como duas ou mais variáveis estão relacionadas em uma dada população. Suas aplicações se estendem nas mais diversas áreas do conhecimento, tais como: Agricultura, economia, administração, biologia, engenharia, etc.

Conforme citado por Montgomery, Peck e Vining (2006), modelos de regressão são usados para diversos fins, dentre eles: Descrição de dados, estimativa de parâmetros, previsão e estimativas e controle. Para Andrade e Ogliari (2013), o principal objetivo dos modelos de regressão é modelar o relacionamento entre diversas variáveis preditoras e uma variável resposta, e esse relacionamento pode ser feito por uma função linear ou não linear.

Os modelos de regressão podem ser classificados em lineares, linearizáveis e não lineares (DRAPER; SMITH, 1998). Para Ratkowsky (1983), o termo modelo de regressão linear é usado principalmente para duas diferentes abordagens. A primeira refere-se à linha reta de relação entre duas variáveis, e a segunda refere-se ao modelo em que os parâmetros a serem estimados aparecem linearmente.

Outros modelos que não são lineares nos parâmetros e que podem ser linearizados por meio de transformações matemáticas são os modelos linearizáveis ou intrinsecamente lineares. Segundo Mazucheli e Achcar (2002) um modelo não linear é considerado 'intrinsecamente linear' se este pode ser reduzido a um modelo linear por meio de uma reparametrização apropriada.

Para Seber e Wild (1989), por causa da relativa simplicidade da utilização de métodos de regressão linear, trabalhar com o modelo linear é muito atraente. Porém algumas linearizações tornam-se complexas e até impossíveis considerando os recursos existentes. Ainda para os autores, existem três principais razões para querer transformar os dados, para alcançar a linearidade, para a obtenção de erros que são aproximadamente normais distribuídos, e para conseguir uma variância dos erros constante.

Tomando-se como exemplo o modelo $Y = e^{\beta_0 + \beta_1 X}(1 + \varepsilon_0)$ citado por Seber e Wild (1989), onde $E[\varepsilon_0] = 0$ e $\text{var}[\varepsilon_0] = \sigma_0^2$ independentemente de X , sua linearização dar-se por uma transformação logarítmica:

$$\begin{aligned} \ln(Y) &= \beta_0 + \beta_1 X + \ln(1 + \varepsilon_0) \\ Y^* &= \beta_0 + \beta_1 X + \varepsilon_0^* \end{aligned}$$

onde $E[\varepsilon_0^*] \approx E[\varepsilon_0] = 0$ para ε_0 pequenos, e $\text{var}[\varepsilon_0^*]$ independente de X .

Para Bates e Watts (1988) é importante entender que a linearização do modelo também afeta os pressupostos do mesmo, ou seja, a interpretação do modelo transformado nem sempre equivale a do modelo não transformado.

2.4 Modelos não lineares

Segundo Souza (1998) a equação que representa a regressão não linear, de forma geral, pode ser escrita na forma:

$$\mathbf{Y} = \mathbf{f}(\mathbf{X}, \boldsymbol{\beta}) + \boldsymbol{\varepsilon} \quad (1)$$

sendo $\mathbf{Y}$ um vetor $(n \times 1)$ de observações, $\mathbf{f}(\mathbf{X}, \boldsymbol{\beta})$ é um vetor $(n \times 1)$ de funções não lineares em $\boldsymbol{\beta}$ e $\boldsymbol{\varepsilon}$ é um vetor $(n \times 1)$ de resíduos ou erros.

Sendo assim, uma função é considerada não linear em β se, e somente se, $\frac{\partial f(x_i, \beta)}{\partial \beta_p}$ (derivada parcial em relação aos parâmetros) resultar em uma função que dependa de pelo menos um dos parâmetros.

2.5 Modelo Empírico

O modelo logístico, é uma das funções mais utilizadas em estudos de crescimento. Apresenta um formato sigmoidal, simétrico em relação ao seu ponto de inflexão, que é característica desse tipo de estudo. Uma das parametrizações do modelo logístico é dada por:

$$y_i = \frac{\beta_1}{1 + \exp(\beta_2 - \beta_3 x_i)} + \varepsilon_i, \text{ com } i = 1, \dots, n$$

em que, y_i é a altura medida em centímetros (cm) da planta de milho; x_i é o tempo em dias após a semeadura; β_1 representa a assíntota horizontal, ou seja, o ponto de estabilização do crescimento da planta de milho em centímetros (cm); β_3 a taxa de expansão ou velocidade com que a planta atingia seu valor assintótico em cm/dia/cm; e ε_i é o erro experimental. Com relação ao parâmetro β_2 , ele não apresenta uma interpretação biológica direta, sendo ele um parâmetro de locação e está relacionado ao ponto de inflexão da curva.

Outras interpretações são obtidas a partir da função estudada, Freitas (2005)

discute algumas dessas interpretações. A taxa de crescimento absoluta instantânea (TCI) estima o incremento de crescimento para cada unidade de tempo. TCI é dada pela derivada do modelo em relação ao tempo, ou seja, $TCI = \frac{\partial y}{\partial x}$. A taxa de crescimento instantânea relativa (TCIR) no tempo é dada por $TCIR = y_x^{-1} \left(\frac{\partial y}{\partial x} \right)$. A taxa de crescimento instantânea absoluta é visualizada por meio gráfico sendo $\frac{\partial y}{\partial x}$ na ordenada e y na abscissa, sendo y aproximando-se de β_1 assintoticamente. A taxa de maturidade absoluta (TMA) é encontrada pela razão de $\frac{\partial y}{\partial x}$ em relação ao crescimento assintótico β_1 . O valor obtido na TMA indica a proporção de crescimento diário relativa ao valor assintótico β_1 .

2.6 Resíduos heterocedásticos

A suposição de variância constante ao longo do tempo, ou seja, homogeneidade, é um requisito básico da análise de modelos de regressão. Porém, essa suposição nem sempre é satisfeita.

A heterogeneidade de variâncias é característica de estudos realizados com medidas repetidas ao longo do tempo. Segundo Rawlings, Pantula e Dickey (1998), variâncias heterogêneas implicam que algumas observações contêm mais informações do que outras, ou seja, observações com maior variância influenciam as estimativas de forma mais intensa do que as que possuem menor variância. O método de estimação ou até mesmo transformações de variáveis, podem resolver tal problema na modelagem.

Outra possível técnica para contornar a heterocedasticidade e obter melhores resultados no ajuste de modelos é utilizando a ponderação pelo inverso da variância amostral, conforme utilizada por Mazzini et al. (2003). Tal melhora pode ser verificada por exemplo nos erros padrões das estimativas, indicando que a estimativa com menor erro padrão é mais confiável.

A heterocedasticidade é comumente avaliada por meios gráficos ou testes estatísticos. Um dos testes utilizados para tal confirmação, é o teste de Bartlett que considera a hipótese nula (H_0) sendo a igualdade de variâncias ao longo do tempo, porém esse teste é bastante sensível a normalidade das observações e também só é adequado em casos que $n > 5$, considerando n o número de dias de observação. Outro teste utilizado é o de Hartley que tem a mesma hipótese do teste de Bartlett, tal teste é realizado comparando-se sua estatística de teste $F_{max} = \frac{s_{max}^2}{s_{min}^2}$, em que s_{max}^2 e s_{min}^2 são respectivamente a maior e menor variâncias obtidas em relação aos dias de observação, com o valor tabelado de Hartley.

Vários estudos já foram realizados considerando resíduos heterocedásticos, um exemplo desse estudo é o de Mazucheli, Souza e Philippsen (2011), que ajustaram o modelo Gompertz em dados de peso-idade de codornas considerando as suposições de homogeneidade e heterogeneidade de variância, sob o enfoque bayesiano, e constataram que o modelo considerando a heterocedasticidade melhor se ajusta, pois, não superestima os parâmetros do peso médio assintótico das aves. Outro estudo considerando a heterogeneidade de variância foi realizado por Riazoshams e Midi (2009), os autores analisaram dados de crescimento de frangos por um modelo não linear com erros heterocedásticos.

2.7 Resíduos autocorrelacionados

Em estudos com medidas repetidas ao longo do tempo, os resíduos dos modelos herdam também a autocorrelação. Conforme citado por Seber e Wild (1989), em situações que os dados são recolhidos em sequência ao longo do tempo, pode-se dar origem a autocorrelação nos erros. Segundo Morettin e Toloi (2004), a característica geral de 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.

De acordo com Prado, Savian e Muniz (2013), em estudos de modelos não lineares que descrevem o crescimento, é bastante razoável incorporar a autocorrelação, tendo em vista que as medidas de crescimento são tomadas em uma mesma unidade experimental, planta, fruto ou animal, ao longo do tempo, estando, portanto, provavelmente correlacionadas.

Tal autocorrelação é introduzida admitindo-se que os erros do modelo são autocorrelacionados na forma de um processo autoregressivo estacionário de ordem p , $AR(p)$, em que $\varepsilon_i = \phi_1\varepsilon_{i-1} + \phi_2\varepsilon_{i-2} + \dots + \phi_p\varepsilon_{i-p} + u_i$, sendo $\phi_1, \phi_2, \dots, \phi_p$ parâmetros de autocorrelação e $u_i \sim N(0, \sigma^2)$ é o ruído branco, pode-se incluir tais estruturas no modelo para melhorar o ajuste do mesmo. Segundo Savian (2005), tal relação mostra que o erro da observação relativa a um período está relacionada com o erro da observação anterior. E ainda para a autora, se $\phi > 0$, os erros são autocorrelacionados positivamente, se $\phi < 0$ diz-se que há uma autocorrelação negativa e caso $\phi = 0$ os erros não são correlacionados..

O caso mais simples de um modelo autoregressivo é o de ordem $p = 1$, $AR(1)$: $\varepsilon_i = \phi_1\varepsilon_{i-1} + u_i$, sendo que ε_i depende apenas de ε_{i-1} e do ruído no instante i .

Em estudos com animais, Bergamasco, Aquino e Muniz (2001) utilizaram

modelos não lineares incorporando a autocorrelação residual de curvas de crescimento de fêmeas da raça holandesa, e obtiveram estimativas mais confiáveis. Já Pereira, Muniz e Silva (2005) compararam o ajuste de vários modelos não lineares no estudo da mineralização de nitrogênio em Latossolo, inserindo termos autoregressivos de ordem 2, obtendo assim, melhores ajustes. Em estudo de curvas de crescimento do coco, Prado, Savian e Muniz (2013) obtiveram melhores ajustes considerando uma estrutura com erros autoregressivos de primeira ordem.

2.8 Estimação de parâmetros de modelos não lineares

Para a estimação dos parâmetros de modelos de regressão não linear, deve-se levar em consideração algumas características dos dados que de alguma forma interfiram na qualidade do ajuste. A heterocedasticidade e a autocorrelação dos erros são exemplos destas características.

O método de mínimos quadrados é o um dos meios de estimação mais comumente utilizados e consiste em minimizar a soma de quadrados dos erros ou resíduos. Outro método utilizado é o da máxima verossimilhança que considera no processo estimação a maximização da função de verossimilhança.

Os métodos de estimação abordados na literatura são o método dos momentos, método de mínimos quadrados, método da máxima verossimilhança e métodos bayesianos. Neste trabalho, será descrito os métodos de mínimos quadrados e da máxima verossimilhança.

2.8.1 Método de mínimos quadrados ordinários

O método de estimação por mínimos quadrados ordinários considera que os erros do modelo são independentes, normalmente distribuídos, com média 0 e variâncias constantes, ou seja, $\varepsilon \sim N(\mathbf{0}; I\sigma^2)$. Segundo Ratkowsky (1990), esta suposição é atribuída quando refere-se a erros aditivos.

Considerando a Equação 1, a soma de quadrados dos erros é dada por:

$$S(\beta) = \varepsilon' \varepsilon = \{\mathbf{Y}' - [f(\mathbf{X}, \beta)]'\}[\mathbf{Y} - f(\mathbf{X}, \beta)]$$

desenvolvendo o produto, obtém-se:

$$= \mathbf{Y}'\mathbf{Y} - \mathbf{Y}'f(\mathbf{X}, \beta) - [f(\mathbf{X}, \beta)]'\mathbf{Y} + [f(\mathbf{X}, \beta)]'f(\mathbf{X}, \beta) \quad (2)$$

$$= \mathbf{Y}'\mathbf{Y} - [f(\mathbf{X}, \beta)]'\mathbf{Y} - [f(\mathbf{X}, \beta)]'\mathbf{Y} + [f(\mathbf{X}, \beta)]'f(\mathbf{X}, \beta) \quad (3)$$

$$= \mathbf{Y}'\mathbf{Y} - [f(\mathbf{X}, \beta)]'\mathbf{Y} - [f(\mathbf{X}, \beta)]'\mathbf{Y} + [f(\mathbf{X}, \beta)]'f(\mathbf{X}, \beta) \quad (4)$$

$$S(\beta) = \mathbf{Y}'\mathbf{Y} - 2[f(\mathbf{X}, \beta)]'\mathbf{Y} + [f(\mathbf{X}, \beta)]'f(\mathbf{X}, \beta). \quad (5)$$

Para a minimização da soma de quadrados residual é necessário derivar a Equação 5 em relação aos seus parâmetros e igualar as expressões resultantes à zero montando o sistema de equações normais não lineares como se segue:

$$\begin{aligned} \frac{\partial S(\beta)}{\partial \beta} &= -2 \left[\frac{\partial f(\mathbf{X}, \beta)}{\partial \beta} \right]' \mathbf{Y} + \left[\frac{\partial f(\mathbf{X}, \beta)}{\partial \beta} \right]' f(\mathbf{X}, \beta) + [f(\mathbf{X}, \beta)]' \left[\frac{\partial f(\mathbf{X}, \beta)}{\partial \beta} \right] \\ &= -2 \left[\frac{\partial f(\mathbf{X}, \beta)}{\partial \beta} \right]' \mathbf{Y} + \left[\frac{\partial f(\mathbf{X}, \beta)}{\partial \beta} \right]' f(\mathbf{X}, \beta) + \left[\frac{\partial f(\mathbf{X}, \beta)}{\partial \beta} \right]' f(\mathbf{X}, \beta) \\ &= -2 \left[\frac{\partial f(\mathbf{X}, \beta)}{\partial \beta} \right]' \mathbf{Y} + 2 \left[\frac{\partial f(\mathbf{X}, \beta)}{\partial \beta} \right]' f(\mathbf{X}, \beta) \\ &= -2 \left[\frac{\partial f(\mathbf{X}, \hat{\beta})}{\partial \beta} \right]' [\mathbf{Y} - f(\mathbf{X}, \hat{\beta})] = 0 \end{aligned}$$

sendo que $\frac{\partial f(\mathbf{X}, \beta)}{\partial \beta}$ é a matriz de derivadas parciais ($n \times k$):

$$\frac{\partial f(\mathbf{X}, \beta)}{\partial \beta} = \begin{bmatrix} \frac{\partial f(\mathbf{X}_1, \beta)}{\partial \beta_1} & \frac{\partial f(\mathbf{X}_1, \beta)}{\partial \beta_2} & \dots & \frac{\partial f(\mathbf{X}_1, \beta)}{\partial \beta_k} \\ \frac{\partial f(\mathbf{X}_2, \beta)}{\partial \beta_1} & \frac{\partial f(\mathbf{X}_2, \beta)}{\partial \beta_2} & \dots & \frac{\partial f(\mathbf{X}_2, \beta)}{\partial \beta_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f(\mathbf{X}_n, \beta)}{\partial \beta_1} & \frac{\partial f(\mathbf{X}_n, \beta)}{\partial \beta_2} & \dots & \frac{\partial f(\mathbf{X}_n, \beta)}{\partial \beta_k} \end{bmatrix}$$

Na maioria dos problemas com modelos não lineares, é mais prático encontrar as estimativas de mínimos quadrados por procedimentos de procura numérica direta do que, inicialmente obter as equações normais, $\left[\frac{\partial f(\mathbf{X}, \beta)}{\partial \beta} \right]' [\mathbf{Y} - f(\mathbf{X}, \beta)] = 0$, e, então, utilizar métodos numéricos iterativos para encontrar a solução dessas equações.

Segundo Ratkowsky (1990) a estimação por mínimos quadrados para os parâmetros de modelos de regressão não linear, ao contrário daqueles modelos de regressão linear, não podem ser determinados a partir de uma expressão matemática explícita. Em vez disso, deve-se obter os estimadores por outros meios, tais como um método iterativo,

começando com um conjunto de estimativas de parâmetros iniciais.

Para Mazucheli e Achcar (2002), em modelos multiparamétricos, as soluções das equações normais podem ser extremamente difíceis de serem obtidas e algum método iterativo de resolução de equações normais deve ser utilizado. Alguns métodos iterativos utilizados são o de Gauss-Newton, Davidon-Fletcher-Powell, Steepest descent, procura pelo gradiente conjugado, procura de Nelder-Mead e algoritmo de Golub-Pereyra.

Segundo Martins (2013), a solução do sistema de equações normais 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.

A solução do sistema de equações após o uso dos métodos iterativos leva ao estimador:

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}.$$

O método de estimação mínimos quadrados ordinários não fornece estimativas mínimas das variâncias quando $\text{Var}(\varepsilon) \neq I\sigma^2$ (RAWLINGS; PANTULA; DICKEY, 1998). Neste caso, o método que seria mais adequado é o de mínimos quadrados ponderados.

2.8.2 Método de mínimos quadrados ponderados

O método de mínimos quadrados ponderados é uma alternativa de estimação que pode ser utilizado quando a suposição de homogeneidade de variância dos erros não é verdadeira, pois, segundo Hoffmann e Vieira (1998) fornecem estimadores não tendenciosos e de mínima variância.

Considerando-se agora que $\varepsilon \sim N(\mathbf{0}; \mathbf{V}\sigma^2)$, sendo $\mathbf{V}$ uma matriz diagonal e positiva definida, que representa a variância de cada ε_i , tem-se:

$$\mathbf{V}\sigma^2 = \begin{bmatrix} V_1 & 0 & \cdots & 0 \\ 0 & V_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & V_n \end{bmatrix} \sigma^2$$

O fato de serem nulos os elementos fora da diagonal principal da matriz $\mathbf{V}$ significa que é válida a pressuposição de independência das observações, isto é, $E(\varepsilon_i, \varepsilon_j) = 0$

para $i \neq j$.

Ainda, tem-se a matriz diagonal dos pesos $\mathbf{W}$, cujos elementos são dados por

$$w_j = \frac{1}{\sqrt{V_j}}:$$

$$\mathbf{W} = \begin{bmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_n \end{bmatrix}$$

sendo assim, tem-se que $\mathbf{W}\mathbf{W} = \mathbf{V}^{-1}$ e $\mathbf{V} = \mathbf{W}^{-1}\mathbf{W}^{-1}$.

Pré-multiplicando a matriz de pesos $\mathbf{W}$ em ambos os lados da equação $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, tem-se:

$$\mathbf{W}\mathbf{Y} = \mathbf{W}\mathbf{X}\boldsymbol{\beta} + \mathbf{W}\boldsymbol{\varepsilon}$$

ou

$$\mathbf{Y}^* = \mathbf{X}^*\boldsymbol{\beta} + \boldsymbol{\varepsilon}^*$$

agora considera-se o vetor de erros $\boldsymbol{\varepsilon}^* = \mathbf{W}\boldsymbol{\varepsilon}$, sendo assim, tem-se que $E(\boldsymbol{\varepsilon}^*) = 0$ e:

$$\begin{aligned} \text{Var}(\boldsymbol{\varepsilon}^*) = \text{Var}(\mathbf{W}\boldsymbol{\varepsilon}) &= \mathbf{W}\text{Var}[\boldsymbol{\varepsilon}]\mathbf{W}' \\ &= \mathbf{W}\mathbf{V}\mathbf{W}'\sigma^2 \\ &= \mathbf{W}\mathbf{W}^{-1}\mathbf{W}^{-1}\mathbf{W} \\ &= \mathbf{I}\sigma^2 \end{aligned}$$

Com os resultados obtidos acima, pode-se encontrar as estimativas de mínimos quadrados ponderados, por meio dos mínimos quadrados ordinários (RAWLINGS; PANTULA; DICKEY, 1998), ou seja,

$$\begin{aligned} \hat{\boldsymbol{\beta}}_P &= (\mathbf{X}^{*'}\mathbf{X}^*)^{-1}\mathbf{X}^{*'}\mathbf{Y}^* \\ &= (\mathbf{X}'\mathbf{W}'\mathbf{W}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{W}'\mathbf{W}\mathbf{Y}) \\ &= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}) \end{aligned}$$

e a variância de $\hat{\beta}_P$ é dada por:

$$\begin{aligned}\text{Var}(\hat{\beta}_P) &= (\mathbf{X}^{*'} \mathbf{X}^*)^{-1} \sigma^2 \\ &= (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} \sigma^2\end{aligned}$$

2.8.3 Método de mínimos quadrados generalizados

O método de estimação de mínimos quadrados generalizados é utilizado quando além da suposição de homogeneidade de variâncias residuais é violada, há também autocorrelação dos erros.

O método de mínimos quadrados generalizados para estimação de parâmetros é mais adequado que os demais citados para dados que apresentam heterocedasticidade e autocorrelação dos erros. Segundo Calegario et al. (2005), tais características estão presentes em muitas bases de dados em diversas áreas do conhecimento.

A estrutura do erro é dada por $\varepsilon \sim N(\mathbf{0}; \mathbf{G}\sigma^2)$ em que $\mathbf{G}$ é uma matriz simétrica, positiva definida, que representa as variâncias e covariâncias dos erros (SAVIAN, 2005).

Conforme citado por Rawlings, Pantula e Dickey (1998), uma transformação linear é feita em $\mathbf{Y}$ de modo que o modelo transformado irá satisfazer as condições de $\text{Var}(\varepsilon) = \mathbf{G}\sigma^2$, ou seja, considera-se $\mathbf{T}\mathbf{T}' = \mathbf{G}$. Sabendo que $\mathbf{T}$ é não singular, admitindo inversa $\mathbf{T}^{-1}$ e pre-multiplicando no modelo $\mathbf{Y} = \mathbf{X}\beta + \varepsilon$, tem-se:

$$\mathbf{T}^{-1}\mathbf{Y} = \mathbf{T}^{-1}\mathbf{X}\beta + \mathbf{T}^{-1}\varepsilon$$

ou ainda:

$$\mathbf{Y}^* = \mathbf{X}^*\beta + \varepsilon^* \tag{6}$$

sendo $\mathbf{Y}^* = \mathbf{T}^{-1}\mathbf{Y}$, $\mathbf{X}^*\boldsymbol{\beta} = \mathbf{T}^{-1}\mathbf{X}\boldsymbol{\beta}$ e $\boldsymbol{\varepsilon}^* = \mathbf{T}^{-1}\boldsymbol{\varepsilon}$. Agora a $\text{Var}(\boldsymbol{\varepsilon}^*)$ será:

$$\begin{aligned}
 \text{Var}(\boldsymbol{\varepsilon}^*) &= \text{Var}(\mathbf{T}^{-1}\boldsymbol{\varepsilon}) \\
 &= \mathbf{T}^{-1}\text{Var}(\boldsymbol{\varepsilon})(\mathbf{T}^{-1})' \\
 &= \mathbf{T}^{-1}\mathbf{G}(\mathbf{T}^{-1})'\sigma^2 \\
 &= \mathbf{T}^{-1}\mathbf{T}\mathbf{T}'(\mathbf{T}^{-1})'\sigma^2 \\
 &= \mathbf{I}\sigma^2
 \end{aligned}$$

Na Equação 6, pode-se utilizar apropriadamente o conceito de mínimos quadrados ordinários, sendo assim, tem-se:

$$\begin{aligned}
 \hat{\boldsymbol{\beta}}_G &= (\mathbf{X}^{*'}\mathbf{X}^*)^{-1}\mathbf{X}^{*'}\mathbf{Y}^* \\
 &= (\mathbf{X}'(\mathbf{T}^{-1})'\mathbf{T}^{-1}\mathbf{X})^{-1}\mathbf{X}'(\mathbf{T}^{-1})'\mathbf{T}^{-1}\mathbf{Y} \\
 &= (\mathbf{X}'(\mathbf{T}\mathbf{T}')^{-1}\mathbf{X})^{-1}\mathbf{X}'(\mathbf{T}\mathbf{T}')^{-1}\mathbf{Y} \\
 &= (\mathbf{X}'\mathbf{G}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{G}^{-1}\mathbf{Y}.
 \end{aligned}$$

sendo $\mathbf{G}$ a matriz:

$$\mathbf{G} = \frac{\sigma_\varepsilon^2}{1 - \phi_1^2} \begin{bmatrix} 1 & \phi_1 & \phi_1^2 & \cdots & \phi_1^{n-1} \\ \phi_1 & 1 & \phi_1 & \cdots & \phi_1^{n-2} \\ \phi_1^2 & \phi_1 & 1 & \cdots & \phi_1^{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \phi_1^{n-1} & \phi_1^{n-2} & \phi_1^{n-3} & \cdots & 1 \end{bmatrix}$$

2.8.4 Estimação por máxima verossimilhança

Outro método de estimação de parâmetros em modelos não lineares é o método da máxima verossimilhança, que diferentemente da minimização no método de mínimos quadrados, consiste em maximizar a função de verossimilhança para obter os estimadores dos parâmetros. Bolfarine e Sandoval (2010) consideram que sejam $X_1, \dots, X_n$ uma amostra aleatória de tamanho n da variável aleatória X com função de densidade (ou probabilidade) $f(x|\theta)$, com $\theta \in \Theta$, onde Θ é o espaço paramétrico, então a função de verossimilhança de θ

correspondente a amostra aleatória observada é dada por:

$$L(\theta; x) = \prod_{i=1}^n f(x_i; \theta).$$

O valor que maximiza a função $L(\theta; x)$ é $\hat{\theta} \in \Theta$ é o estimador de máxima verossimilhança de θ .

No caso em que considera-se um modelo com erros normalmente distribuídos, ou seja, $\varepsilon \sim N(\mathbf{0}; I\sigma^2)$, sua função de verossimilhança é dada por:

$$L(\theta; \sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n \frac{[y_i - f(x_i, \theta)]^2}{\sigma^2} \right\}$$

Na maioria dos casos, trabalha-se com o logaritmo natural da função de verossimilhança, pois, maximizar o logaritmo natural de uma função é em geral mais simples e chega-se nos mesmos resultados da maximização da função original.

O logaritmo natural da função de verossimilhança de θ é denotado por:

$$\ell(\theta; x) = \log L(\theta; x)$$

Considerando novamente o modelo com erros normais e ignorando as constantes, $\ell(\theta; x)$ será dados por:

$$\begin{aligned} \ell(\theta; \sigma^2) &= -\frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n [y_i - f(x_i, \theta)]^2 \\ &= -\frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} SQRes(\theta) \end{aligned}$$

O estimador de verossimilhança pode ser encontrado calculando-se a raiz da derivada da função logaritmo de verossimilhança para função deriváveis, ou seja:

$$\ell'(\theta; x) = \frac{\partial \ell(\theta; x)}{\partial \theta} = 0$$

Segundo Macerau (2012), o logaritmo da função de verossimilhança é aplicado como um artifício para facilitar o desenvolvimento analítico, visando determinar os estimadores, pois, o máximo da função de verossimilhança e o máximo do seu logaritmo ocorrem no mesmo ponto.

Para Montgomery, Peck e Vining (2006), a estimação por máxima verossimilhança é uma alternativa quando a distribuição dos erros de um modelo é desconhecida.

2.9 Método Iterativo

Para esse trabalho, o método iterativo escolhido para cálculo das estimativas de parâmetros é o método de Gauss-Newton.

2.9.1 Método Iterativo de Gauss-Newton

O método de Gauss-Newton consiste em uma aproximação linear e utiliza valores iniciais atribuídos aos parâmetros no processo de iteração. Para tanto, é feita uma expansão em série de Taylor centrada em β^0 (valores iniciais atribuídos aos parâmetros) para a função não linear:

$$f(x_i, \beta) = f(x_i, \beta^0) + \frac{\partial f(x_i, \beta^0)}{\partial \beta_1}(\beta_1 - \beta_1^0) + \dots + \frac{\partial f(x_i, \beta^0)}{\partial \beta_p}(\beta_p - \beta_p^0) \quad (7)$$

$$i = 1, 2, \dots, n \quad (8)$$

ou ainda, considerando o modelo da Equação 7, tem-se:

$$f(X, \beta) \approx f(X, \beta^0) + F_{\beta^0}(\beta - \beta^0)$$

$$Y = f(X, \beta^0) + F_{\beta^0}(\beta - \beta^0) + \varepsilon$$

$$= Y - f(X, \beta^0) = F_{\beta^0}(\beta - \beta^0) + \varepsilon$$

$$y^{**} = F_{\beta^0}\theta + \varepsilon$$

sendo, $F_{\beta^0} = \frac{\partial f(X, \beta)}{\partial \beta}$ e $\theta = (\beta - \beta^0)$. Assim, por mínimos quadrados tem-se:

$$\hat{\theta} = (F'_{\beta^0} F_{\beta^0})^{-1} F'_{\beta^0} Y^{**}$$

Após este procedimento, estima-se a soma de quadrados com o $\hat{\beta}$ encontrado no passo anterior. O próximo passo consiste em repetir os cálculos considerando como valor inicial a estimativa do parâmetro obtido acima. Se a soma de quadrados obtida for menor, o processo continua até que se encontre uma soma de quadrados maior do que a anterior, convergindo preferencialmente até a sétima casa decimal.

É perceptível que há uma importância na escolha do valor inicial do parâmetro β^0 . Para Martins (2013) o sucesso na utilização do algoritmo de Gauss-Newton, como método iterativo, vai depender da escolha apropriada da função resposta e de bons valores iniciais. Embora existam algumas orientações gerais para determinação de valores iniciais, o processo de escolha é um procedimento decidido pelo pesquisador.

Ainda para a autora, 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.

Conforme citado por Calegario et al. (2005), uma das limitações do uso de modelos não lineares é a escolha correta dos valores iniciais dos parâmetros para o processo de iteração. Segundo Montgomery, Peck e Vining (2006), uma má escolha do valor inicial pode causar uma convergência para um mínimo local, e é inconsistente achar que uma ótima solução foi obtida.

2.10 Distribuições assimétricas

Em muitos casos quando não há o cumprimento da suposição de normalidade residual em modelos não lineares, o pesquisador utiliza transformações na variável resposta. Porém, transformar a variável resposta acarreta na mudança de interpretação dos resultados do estudo, como por exemplo a interpretação dos parâmetros. Uma alternativa para contornar este problema, utilizando outro enfoque para a falta de normalidade dos erros é considerar outras distribuições para os mesmos, como as distribuições assimétricas que dão suporte tanto para assimetria quanto para caudas pesadas.

De acordo com Campos (2011), sabe-se que muitos estudos nem sempre estão de acordo com o modelo normal devido a falta de simetria na distribuição ou presença de distribuições com caudas mais pesadas ou mais leves que a normal. Nesse sentido, conforme comentado por Cruz e Branco (2009), vários trabalhos na literatura vem propondo distribuições mais flexíveis para o termo de erro, que podem ser mais realistas, ver por exemplo Azzalini e Capitanio (1999), DiCiccio e Monti (2004) e Genton (2004).

Algumas distribuições assimétricas são úteis, pois preservam propriedades de uma distribuição da qual foi gerada. Por exemplo a distribuição normal assimétrica (ou skew-normal) e a distribuição assimétrica t-student (ou skew-t), ambas herdam algumas propriedades das distribuições normal e t-student, respectivamente.

Segundo Campos (2011), tanto a distribuição normal quanto a distribuição t-student pressupõem a simetria dos dados, assim, seria interessante obter distribuições assimétricas que tenham como caso particular distribuições simétricas conhecidas, como é o caso da distribuição skew-normal e skew-t.

Vários autores utilizaram distribuições assimétricas para estudos de modelagem em que ocorrem caudas mais pesadas ou leves que a distribuição normal. Campos (2011) ajustou modelos não lineares considerando erros normais assimétricos e t-student assimétricos com e sem heterocedasticidade, utilizando uma abordagem bayesiana, e notou que o modelo assimétrico normal obteve um melhor ajuste comparado aos demais.

Um estudo do modelo de crescimento de Gompertz aplicado em dados de pesos de codornas, machos e fêmeas, realizado por Rossi e Santos (2014), considerou distribuições assimétricas e simétricas para os erros, e foi constatado que o modelo com erros assimétricos normais e t-student, respectivamente, foram os que melhor se ajustaram aos dados.

Já Guedes et al. (2014) ajustaram modelos de regressão com erros normais e com erros normais assimétricos à dados de altura de plantas, e o modelo com erros normais assimétricos mostrou-se mais adequado em algumas situações.

Cruz e Branco (2009) fizeram um estudo de modelos de crescimento com enfoque bayesiano e considerando distribuições assimétricas dos erros em dados clínicos de gestantes, e concluíram que a distribuição skew-t adotada para os erros foi a que melhor se ajustou aos dados.

Como mostrado acima, estudos considerando distribuições assimétricas dos erros são frequentemente utilizadas levando em consideração diferentes enfoques.

2.10.1 Distribuição normal assimétrica(SN)

A distribuição normal assimétrica foi introduzida por Azzalini (1985) e é útil para estudos que apresentam assimetria nos erros de modelos não lineares, pois, preserva algumas propriedades da distribuição normal. Ela é construída a partir da função densidade da distribuição normal e da função de distribuição acumulada. Conforme citado por Campos (2011), a distribuição normal assimétrica representa a generalização da distribuição normal na qual tem um parâmetro de forma adicional que define a direção da assimetria da distribuição.

Conforme Azzalini (1985), a função densidade de uma variável aleatória com distribuição normal assimétrica é dada por:

$$\phi(z; \lambda) = 2\phi(z)\Phi(\lambda z), z \in \mathbb{R}$$

em que ϕ e Φ são a função densidade de probabilidade e a função de distribuição acumulada padrão, respectivamente, Z é uma variável aleatória com distribuição normal assimétrica com parâmetro λ , sendo $\lambda \in \mathbb{R}$.

Algumas propriedades da distribuição normal assimétrica podem ser vistas a seguir:

1. A densidade de $SN(0)$ é igual $N(0, 1)$;
2. Caso $Z \sim SN(\lambda)$, então quando $\lambda \rightarrow \infty$, $Z \rightarrow^D HN(0, 1)$;
3. Caso $Z \sim SN(\lambda)$, então $-Z \sim SN(-\lambda)$;
4. Caso $Z \sim SN(\lambda)$, então a f.d.p de Z é unimodal e $\log f(z)$ é uma função côncava de z .

A distribuição normal assimétrica apresenta a seguinte função de densidade acumulada (Φ):

$$\Phi(z; \lambda) = 2 \int_{-\infty}^z \int_{-\infty}^{\lambda t} \phi(t)\phi(u)du dt,$$

e ainda,

1. $1 - \Phi(-z; \lambda) = \Phi(z; -\lambda)$;
2. $\Phi(z; 1) = \{\Phi(z)\}^2$;
3. $\sup |\Phi(z) - \Phi(z; \lambda)| = \Pi^{-1} \arctan |\lambda|$;
4. Caso $Z \sim SN(\lambda)$ então $Z^2 \sim \chi_1^2$.

2.10.2 Distribuição normal assimétrica com parâmetros de locação e escala

Segundo Azzalini (1985), uma variável aleatória Z tem distribuição normal assimétrica com parâmetro μ , parâmetro de escala σ^2 e parâmetro de assimetria λ , denotada por $Z \sim SN(\mu, \sigma, \lambda)$, se sua função de densidade for dada por:

$$f(z) = 2\phi(z; \mu, \sigma^2)\Phi\left(\frac{\lambda(z - \mu)}{\sigma}\right),$$

em que $\phi(\cdot; \mu, \sigma^2)$ denota a densidade de uma distribuição normal univariada com média μ e variância $\sigma^2 > 0$ e $\Phi(\cdot)$ a função de distribuição acumulada univariada normal padrão.

O comportamento da distribuição pode ser vista por meio da Figura 2, sendo as curvas simuladas mantendo os parâmetros $\mu = 0$ e $\sigma = 1$ fixos e o parâmetro de assimetria λ variando em $-1,5$, 0 e $1,5$. Verifica-se no caso do parâmetro de assimetria ser $\lambda = 0$ a distribuição da variável passa a ser normal assimétrica.

Figura 2: Gráfico do comportamento da distribuição normal assimétrica

2.10.3 Modelo não linear normal assimétrico

O modelo não linear normal assimétrico é definido, segundo Cancho, Lachos e Ortega (2010), por:

$$Y_i = \eta(\beta, x_i) + \varepsilon_i, i = 1, \dots, n,$$

sendo Y_i a variável resposta, $\eta(\cdot)$ é uma função injetiva e duas vezes diferenciável em relação ao vetor de parâmetros β , x_i é um vetor de variáveis explicativas e o erro aleatório $\varepsilon \sim SN\left(-\sqrt{\frac{2}{\pi}}\lambda, \sigma^2, \lambda\right)$ que corresponde ao modelo de regressão onde a distribuição do erro

tem média zero. Tem-se ainda que:

$$E[Y_i] = \eta(\beta, x_i), \text{eVar}[Y_i] = \sigma^2 + \left(1 - \frac{2}{\pi}\right) \lambda^2$$

e ainda $Y_i \sim SN\left(\eta(\beta, x_i) - \sqrt{\frac{2}{\pi}}\lambda, \sigma^2, \lambda\right), i = 1, \dots, n$.

Sabendo-se que a distribuição normal assimétrica é unimodal, tem-se então que se $\lambda < 0$, a distribuição tem uma assimetria negativa, caso $\lambda > 0$ tem-se uma assimetria positiva e se $\lambda = 0$ volta-se a distribuição normal simétrica $N(\mu, \sigma^2)$.

Para estimar os parâmetros de um modelo não linear que tem seus dados distribuídos segundo uma normal assimétrica, pode-se considerar a função de verossimilhança. O logaritmo natural da função de verossimilhança, segundo Cancho, Lachos e Ortega (2010), considerando-se um vetor θ dado de uma amostra $y = (y_1, \dots, y_n)$ é dado por $\ell(\theta) = \sum_{i=1}^n \ell_i(\theta)$, onde:

$$\ell_i(\theta) = \log 2 - \frac{1}{2} \log 2\pi - \frac{1}{2} \log \psi - \frac{d_i}{2} + \log \Phi_1(K_i),$$

com $\psi = \sigma^2 + \lambda^2$, $d_i = \frac{(Y_i - \eta(\beta, x_i) + \sqrt{\frac{2}{\pi}}\lambda)^2}{\sigma^2 + \lambda^2}$ e $K_i = \frac{\lambda}{\sigma\psi^{1/2}} \left(Y_i - \eta(\beta, x_i) + \sqrt{\frac{2}{\pi}}\lambda\right)$.

2.10.4 Distribuição t-Student assimétrica (ST)

A distribuição t-student assimétrica é derivada da distribuição t-Student e é aplicável em casos que além de assimetria residual, também ocorre a presença de valores extremos, ou seja, a dados com estrutura assimétricas e com caudas pesadas. É mais eficaz do que a distribuição normal assimétrica, pois, é mais robusta nesses casos.

Conforme citado por Godoi e Branco (2007), uma variável aleatória Z é denominada t-student assimétrica padrão com parâmetros de assimetria λ e curtose ν , se sua função densidade de probabilidade for dada por:

$$f_z(z) = 2t_\nu(z)T_{\nu+1}\left(\lambda_z\sqrt{\frac{1+\nu}{\nu+z^2}}\right)$$

onde t_ν denota a função densidade de probabilidade de uma distribuição t-student padrão com ν graus de liberdade, $T_{\nu+1}$ a função de distribuição acumulada de uma distribuição t-student padrão com $\nu + 1$ graus de liberdade e $-\infty < z < \infty$.

Algumas propriedades acerca da distribuição podem ser consideradas. Segundo Godoi e Branco (2007) se $Z \sim ST(\nu, \lambda)$ e $Y \sim ST(\mu, \sigma^2, \nu, \lambda)$ então:

1. $|Z| \sim HT(\nu)$: $|Z|$ tem distribuição t-positiva ou half-t com ν graus de liberdade;
2. Se $\nu = 1$ então $Z \sim SC(\lambda)$: $SC(\lambda)$ é a distribuição Cauchy-assimétrica com parâmetro de assimetria λ ;
3. Quando $\lambda \rightarrow \infty$, Z converge a uma $HT(\nu)$;
4. Quando $\nu \rightarrow \infty$, Z converge a uma $SN(\lambda)$;
5. $-Z \sim ST(\nu, -\lambda)$;
6. $Z^2 \sim F - Snedecor(1, \nu)$;
7. $F_Z(z; \nu, -\lambda) = 1 - F_Z(-z; \nu, \lambda)$;
8. Se $Y \sim ST(\mu, \sigma^2, \nu, \lambda)$ e $Y_1 = a + bY$ então $Y_1 \sim ST(a + b\mu, b^2\sigma^2, \text{signal}(b)\lambda, \nu)$.

Ainda conforme citado pelos autores, por meio da propriedade 8, pode-se afirmar que a distribuição t-student assimétrica é fechada para transformações lineares, ou seja, qualquer combinação linear de uma variável aleatória com distribuição skew-t será também uma distribuição skew-t.

Da mesma forma que ocorre na distribuição skew-normal, o parâmetro λ caracteriza a forma da distribuição. Se λ apresentar valores negativos, isso indica uma assimetria negativa. Caso λ obtenha valores positivos, então há uma assimetria positiva. E no caso de $\lambda = 0$, a distribuição será simétrica e igual a distribuição de t-student.

O fato da distribuição normal assimétrica ser caso particular da distribuição t-Student assimétrica, implica que também a normal é um caso particular da mesma quando $\lambda = 0$. Essa informação é importante, pois, por serem modelos encaixados pode-se utilizar métodos como teste da razão de verossimilhança, além dos critérios de informação como AIC e BIC para realizar comparações dos modelos com relação ao qual é mais adequado para ajustes.

2.10.5 Distribuição t-student assimétrico com parâmetros de locação e escala

Considerando agora a distribuição skew-t com parâmetros de locação e escala, Azzalini e Capitanio (2003) descrevem que se $Y \sim ST(\xi, \omega, \alpha, \nu)$ então:

$$f(y; \xi, \omega, \alpha, \nu) = 2t(y; \xi, \omega, \nu)T(\zeta; \nu + 1),$$

sendo $\zeta = \alpha z \tau$, $z = \frac{y - \xi}{\omega}$, $\tau = \left(\frac{\nu + 1}{\nu + z^2}\right)^{\frac{1}{2}}$, $t(y; \xi, \omega, \nu) = \frac{\Gamma(\frac{\nu+1}{2})\left(1 + \frac{z^2}{\nu}\right)^{-\frac{\nu}{2} - \frac{1}{2}}}{\omega(\pi\nu)^{\frac{1}{2}}\Gamma(\frac{\nu}{2})}$ e $T(\zeta; \nu + 1) = \int_{-\infty}^{\zeta} t(u; 0, 1, \nu) du$.

Tal parametrização é utilizada por Monti (2008), e foi utilizada no ajuste do modelo deste trabalho.

Na Figura 3 é apresentado o comportamento da curva da distribuição em diferentes cenários. Os parâmetros de locação ($\xi = 0$) e escala ($\omega = 1$) foram mantidos fixos na simulação. Já os parâmetros de assimetria (λ) e curtose (ν) foram variados conforme a legenda do gráfico. No caso em que o parâmetro $\lambda = 0$, tem-se uma distribuição t-student com ν graus de liberdade.

Figura 3: Gráfico do comportamento da distribuição t-student assimétrica

A estimação de modelos que consideram erros com distribuição t-student assimétrica também pode ser realizada pelo método de verossimilhança. A função de ve-

rossimilhança é descrita por:

$$L(\nu, \lambda; z) = \prod_{i=1}^n 2t_{\nu}(z_i)T_{\nu+1} \left(\lambda_{z_i} \sqrt{\frac{1+\nu}{\nu+z_i^2}} \right).$$

Vários estudos da distribuição t assimétrica com diversas abordagens, como aplicações e inferenciais, podem ser encontrados na literatura, ver Azzalini e Capitanio (2003), Monti (2008), Azzalini e Genton (2008), Macerau (2012) e Cruz e Branco (2009).

3 METODOLOGIA

3.1 Material

Os dados da altura da planta de milho utilizados neste trabalhos foram do híbrido transgênico 30F35 Y (Yieldgard), e foram adaptados de Netto (2013).

O experimento foi conduzido no município de Votuporanga-SP, em área experimental do Pólo Regional Noroeste Paulista da APTA (Agência Paulista de Tecnologia dos Agronegócios), no ano agrícola 2011/2012. Foi utilizado o delineamento em blocos casualizados com 4 repetições (blocos), sendo cada parcela constituída de quatro linhas de quatro metros de comprimento espaçadas à 0,90m entre si, na densidade de 7 plantas por metro linear. Foram selecionadas dez plantas de cada linha central, as quais foram marcadas e utilizadas nas avaliações durante todo o ciclo cultural, totalizando 20 plantas avaliadas por parcela.

Em cada dia de avaliação foi medida a altura das plantas em centímetros, com auxílio de uma régua, sendo esta medida da base da planta (solo) até o ápice da última folha expandida do cartucho. A primeira medição da altura da planta de milho foi realizada 15 dias após a semeadura, e as medições seguintes ocorreram com 30, 40, 50 e 122 dias, respectivamente, após a semeadura.

Como foi mencionado, os dados foram adaptados do pesquisador Netto (2013), sendo assim, para este trabalho foi selecionado apenas um dos híbridos utilizados no experimento e um tratamento, que foi a testemunha, já que o objetivo do presente trabalho refere-se a uma das variáveis mensuradas no experimento, altura da planta. E ainda, devido a contraversões encontradas nos dados como o decrescimento de algumas plantas ao longo do tempo, só foram utilizadas 58 plantas para este trabalho.

3.2 Métodos

Nesta seção são apresentados a metodologia abordada no presente trabalho, tais como, os testes indicados para verificação das suposições em relação aos resíduos do modelo, descrição da obtenção de valores iniciais e métodos de seleção de modelos.

3.3 Análise da homogeneidade de variância dos erros - Teste de Hartley

A pressuposição de homogeneidade de variâncias associada ao resíduo do modelo foi verificada por meio do teste de Hartley (1950). Esse teste, também conhecido como F máximo, tem sua estatística de teste dada por:

$$F_{max} = \frac{s_{max}^2}{s_{min}^2}$$

em que s_{max}^2 é a maior variância obtida em relação aos dias, e s_{min}^2 = a menor variância também obtida em relação aos dias.

A hipótese testada é:

$$\begin{aligned} H_0 &: \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = \sigma_5^2 \\ H_1 &: \sigma_i^2 \neq \sigma_{i'}^2 \end{aligned}$$

O teste é realizado comparando-se o F_{max} com o valor da tabela Hartley, levando-se em consideração ν graus de liberdade, que é também o número de observações em um dia menos um ($r - 1$), k tratamentos ou dias de observações e α nível de significância ($F_{k,\nu,\alpha}$). Se $F_{max} > F_{(k,\nu)\alpha}$, rejeita-se a hipótese nula de que há homogeneidade de variância, caso $F_{max} < F_{(k,\nu)\alpha}$, considera-se as variâncias homogêneas ao longo do tempo.

3.4 Análise da normalidade

Na análise de diagnóstico para avaliar a qualidade de ajuste do modelo é necessário investigar características como normalidade dos erros. Tal investigação pode se basear em termos gráficos ou em testes realizados para este fim.

Por meio gráfico, tem-se o gráfico quantil-quantil com envelope simulado que apresenta em seu eixo das ordenadas os resíduos do modelo o qual deseja-se verificar a normalidade e no eixo das abscissas os quantis normais. Os pontos, que desejamos verificar normalidade (dados, resíduos), plotados no gráfico precisam estar em torno de uma reta que apresenta um intervalo de confiança em torno da mesma. Caso os pontos estejam dentro deste intervalo, os resíduos são considerados normais.

Outra forma de verificar-se a normalidade dos resíduos dos modelos é por meio

do teste de Shapiro e Wilk (1965). O teste é baseado em uma estatística de teste, dada por $W_{cal} = \frac{b^2}{n \sum_{i=1} (x_i - \bar{x})^2}$ com b sendo uma constante determinada a partir dos dados e de uma tabela, que pode ser usada para tomar-se a decisão de rejeição da hipótese nula testada, ou tal hipótese também pode ser investigada por meio do valor descritivo comparado com o nível de significância estabelecido.

As hipóteses testadas são:

H_0 : Os resíduos seguem uma distribuição normal;

H_1 : Não há normalidade residual.

Caso um nível de significância α seja maior do que o valor descritivo do teste, então rejeita-se a hipótese testada H_0 , caso contrário, α menor que o nível descritivo do teste, então não rejeita-se a hipótese H_0 de que há normalidade residual.

3.5 Valores iniciais

Os valores iniciais para os parâmetros, necessários para o algoritmo iterativo, foram obtidos por meio da linearização do modelo logístico. Considerando que o modelo logístico é intrinsecamente não linear, e que tem-se um conhecimento a priori do parâmetro β_1 , ou seja, o valor máximo da altura da planta de milho, então, desconsiderando a estrutura aditiva dos erros, o modelo pode ser reparametrizado da seguinte forma:

$$\begin{aligned} y_i &= \frac{\beta_1}{1 + \exp(\beta_2 - \beta_3 x_i)} \\ \frac{\beta_1}{y_i} &= 1 + \exp(\beta_2 - \beta_3 x_i) \\ \frac{\beta_1}{y_i} - 1 &= \exp(\beta_2 - \beta_3 x_i) \\ \ln \left(\frac{\beta_1}{y_i} - 1 \right) &= \beta_2 - \beta_3 x_i \\ Y^* &= \beta_2 - \beta_3 x_i \end{aligned}$$

sendo, $Y^* = \ln \left(\frac{\beta_1}{y_i} - 1 \right)$.

Assim, as estimativas dos parâmetros β_2 e β_3 , considerando o ajuste de um modelo de regressão linear simples, constituíram os valores iniciais para estimação dos

parâmetros do modelo não linear, necessários no método iterativo.

Nos casos em que foram consideradas distribuições assimétricas para os resíduos do modelo logístico, inicialmente utilizou-se os mesmos valores iniciais obtidos na linearização, e para os demais parâmetros (de locação, escala e assimetria), foram inseridos valores iniciais de acordo com o conhecimento a priori desses parâmetros.

3.6 Seleção do modelo

Considerando o bom ajuste dos modelos estudados, pode-se realizar a escolha do melhor para o ajuste aos dados por meio de alguma medida para este fim. Pode-se utilizar diversos avaliadores para seleção dos modelos, calculados de diversas maneiras.

Os avaliadores da qualidade do ajuste dos modelos utilizados neste foram os seguintes:

a) Critério de Informação de Akaike (AIC): O melhor ajuste segundo o critério de informação de Akaike (AIC), (AKAIKE, 1974), é o modelo que apresenta menor valor do mesmo. O cálculo é realizado pela equação abaixo:

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

sendo L o máximo da função de verossimilhança e p é o número de parâmetros do modelo.

b) Teste da razão de verossimilhança: o teste tem por objetivo a comparação dos ajustes de dois modelos sendo o modelo sob H_0 um caso especial do modelo sob a hipótese alternativa, ou seja, os modelos a serem comparados devem ser encaixados.

A estatística de teste é dada por:

$$TRV = -2\log \left[\frac{L(\theta_0)}{L(\hat{\theta})} \right] = 2[\log L(\hat{\theta}) - \log L(\theta_0)],$$

em que, $\log L(\theta_0)$ e $\log L(\hat{\theta})$ são as verossimilhanças dos modelos com mais parâmetros e menos parâmetros respectivamente, que sobre $\mathbf{H}_0 : \theta = \theta_0$ segue uma distribuição qui-quadrado com p graus de liberdade, sendo H_0 rejeitada com um nível de significância de α , se $TRV > \chi_{p,1-\alpha}^2$.

3.7 Software estatístico

Toda parte de análise computacional foi realizada por meio do software estatístico R (R Core Team, 2013). No ajuste do modelo logístico considerando normalidade na distribuição dos erros, foi utilizado a função *gnls* do pacote *nlme* (PINHEIRO et al., 2013). Já para o ajuste do modelo logístico em que foi considerado distribuições assimétricas dos erros, foi usado a função *mle2* do pacote *bbmle* (BOLKER; TEAM, 2014).

4 RESULTADOS E DISCUSSÃO

Nesta seção serão apresentadas as análises realizadas, tais como gráficos, resultados de testes estatísticos, estimação de parâmetros dos modelos, e ajustes do modelo logístico aos dados de altura de plantas de milho considerando para o vetor de erros a distribuição normal, normal assimétrica e t de student assimétrica.

4.1 Análise exploratória

Os dados de altura de 58 plantas de milho, medidas em centímetros, ao longo de 122 dias de experimento estão representados no diagrama de dispersão a seguir (Figura 4). Pode-se observar na Figura 4, que apesar dos dias de observações não serem equidistantes, há uma leve percepção que o comportamento dos dados aproximam-se de uma curva sigmoidal. Observa-se também que o crescimento das plantas de milho estabiliza-se em aproximadamente 200cm, e que o crescimento mais acentuado ocorre até aproximadamente 50 dias após o plantio, depois desse período há uma desaceleração do crescimento.

Figura 4: Gráfico de dispersão dos dados de altura de plantas de milho ao longo do tempo

Ainda por meio da Figura 4 é possível verificar uma possível presença de heterogeneidade de variâncias, ou seja, a cada dia de observação a variabilidade na medida

da altura de plantas de milho aumenta. Este comportamento pode ser melhor visualizado no gráfico de box plot (Figura 5) em que claramente observa-se um aumento, ao longo do tempo, da distância interquartílica. Ainda no diagrama de caixas (box plot) é possível observar um deslocamento da mediana para as extremidades da caixa em alguns tempos evidenciando uma certa assimetria da distribuição.

Figura 5: Gráfico boxplot da altura de pés de milho (cm) em função do tempo (dias)

A heterogeneidade de variâncias verificada na análise gráfica pode ser comprovada por meio do teste de Hartley para homogeneidade de variâncias. Como já foi comentado, o teste de Hartley baseia-se na razão da maior variância obtida em relação aos dias e a menor variância também obtida em relação aos dias. Na Tabela 1, são apresentadas as variâncias referentes a cada dia de observação após a semeadura da planta do milho. Por meio desses valores é possível obter a estatística do teste de Hartley.

Tabela 1: Variância e assimetria em cada dia de observação

Dias após semeadura	15 dias	30 dias	40 dias	50 dias	122 dias
Variância	7,7547	33,4229	162,9558	245,9083	359,1664
Assimetria	-0,4121	-0,8136	-0,0349	-0,3037	0,0149

Percebe-se um claro aumento de variância ao longo dos dias após semeadura, conforme já foi comentado pelo gráfico de dispersão da Figura 4. Sendo assim, o teste de Hartley apresentou um valor calculado $F_{max} = 46,32$, e que comparado ao valor tabelado $F_{(5,57)5\%} \cong 2,40$, concluímos pela rejeição da hipótese nula, ou seja, as variâncias são heterogêneas ao longo dos dias que foram realizadas observações.

Os coeficientes de assimetria de cada dia de observação apresentados também na Tabela 1, indicam que até 50 dias após a semeadura, houve uma assimetria negativa, ou seja, os dados estão concentrados no lado direito da média. Já no 122 após a semeadura, ocorreu uma assimetria positiva, o que indica que os dados estão mais concentrados do lado esquerdo da média.

Considerando todas as observações de altura da planta do milho do experimento, obteve-se um coeficiente de assimetria de 0,2646.

Observando o crescimento das plantas individualmente, na Figura 6 nota-se que algumas curvas de plantas que estão representadas nos quadrantes de números 26, 27, 35 e 58 tiveram um crescimento mais acentuado do penúltimo dia de observação para o último. Já para as plantas representadas nos quadrantes 1, 31, 32 e 56 pode-se verificar que nesse período praticamente não houve crescimento, isso pode ser justificado pelo fato de que a planta pode ter sido atacada por percevejos que injetam toxinas e as mesmas deixam de crescer. O crescimento pode ser afetado por diversos fatores estressores, entre eles as pragas que podem causar danos às plantas de milho, devido ao processo de alimentação do percevejo, causando desde deformações, perfilhamentos intensos ou até morte das plantas (NETTO, 2013).

Figura 6: Gráfico da altura de pés de milho (cm) de cada planta em função do tempo (dias)

Segundo Ritchie, Hanway e Benson (2003), o crescimento da planta do milho atinge sua maturidade fisiológica cerca de 125 dias após a emergência, podendo variar de acordo com os diferentes híbridos, estações do ano, datas de plantio e locais.

4.2 Ajuste modelo logístico considerando erros simétricos normais

Na Tabela 2 são apresentadas as estimativas dos parâmetros do modelo logístico, com seus respectivos erros padrão, considerando os erros independentes e identicamente distribuídos de uma normal com média zero e variância constante.

Em relação às estimativas dos parâmetros é possível observar que, para o parâmetro β_1 foi obtida uma estimativa de aproximadamente 163,6cm representando a altura assintótica das plantas. Segundo Freitas (2005), estudando curvas de crescimento animal, o parâmetro A, equivalente ao β_1 deste trabalho, é uma estimativa do peso assintótico do animal porém, quando o peso adulto não é atingido o parâmetro reflete uma estimativa do peso às últimas pesagens. No contexto deste trabalho em que as plantas foram avaliadas até 122 dias após a semeadura e considerando que a maturidade fisiológica das plantas de milho

acontecem em aproximadamente 125 dias após a emergência da planta, é possível considerar que a estimativa do parâmetro β_1 reflita a altura assintótica das plantas de milho.

No estudo realizado por Fernandes et al. (2014), o parâmetro k é chamado de índice de maturidade ou precocidade e, quanto maior seu valor, menos tempo será necessário para o fruto atingir sua massa fresca assintótica. Considerando que neste trabalho k é representado pelo parâmetro β_3 , obteve-se uma estimativa de aproximadamente $0,11dia^{-1}$, sendo tal valor a velocidade que a planta atinge o valor assintótico (β_1).

Outra informação importante obtida com os parâmetros é o ponto de inflexão da curva. Segundo Martins (2013), é possível calcular o ponto de inflexão, que é quando há uma desaceleração no crescimento, que ocorre quando atingi-se metade do crescimento, ou seja, $\frac{\beta_1}{2} \cong 82cm$. A relação dos parâmetros $\frac{\hat{\beta}_2}{\hat{\beta}_3} \cong 40dias$, indica que até o dia 40 ocorreu a velocidade máxima do crescimento por dia, após esse dia houve uma desaceleração no crescimento da altura da planta.

Tabela 2: Estimativas dos parâmetros do modelo logístico e seus respectivos erros padrão

Parâmetro	Estimativa	Erro padrão	p-valor
β_1	163,5833	1,8202	$< 0,0001$
β_2	4,3616	0,1519	$< 0,0001$
β_3	0,1096	0,0039	$< 0,0001$

Com relação a taxa de crescimento instantânea absoluta (TCI), o modelo logístico com erros normais apresenta-se conforme a Figura 7. Considerando a interpretação de Freitas (2005) para o caso das plantas do milho, tem-se que a TCI estima o incremento no crescimento para cada unidade de tempo, ou seja, a cada dia. Verifica-se que da semeadura até 122 após a mesma, houve uma variação de até 4,48cm de crescimento ao dia. Pode-se observa na Figura 7 que tal crescimento foi obtido em 40 dias após a semeadura da planta do milho. Após esse dia, há uma diminuição diária do crescimento até o último dia de observação.

Figura 7: Taxa de crescimento instantânea (TCI) em cm/dia - modelo logístico normal

Outra informação obtida foi a taxa de maturidade absoluta (TMA) média. Para Freitas (2005), a TMA indica a proporção de crescimento diário relativo ao valor assintótico. A Taxa de maturidade absoluta do modelo foi $TMA = 1,49\%$, ou seja, esse valor indica que a cada dia houve um aumento em média de 1,49% de altura das plantas relativo ao valor assintótico estimado $\hat{\beta}_1 = 163,6\text{cm}$.

Para avaliar a qualidade do ajuste e obter a validação do modelo, faz-se necessário a realização de uma análise de diagnóstico. Inicialmente, para verificação da normalidade dos resíduos do modelo, observa-se no gráfico de probabilidades normal com envelope simulado (Figura 8) que muitos pontos encontram-se fora do envelope, ou seja, aparentemente os resíduos não seguem distribuição normal. Tal afirmativa pode ser comprovada por meio do teste de normalidade de Shapiro-Wilk em que obteve-se um nível descritivo de 0,00002497, ou seja, rejeita-se a hipótese de que há normalidade residual considerando o nível de 5% de significância.

Figura 8: Gráfico de probabilidades normais com envelope simulado com 95% de confiança

Na Figura 9 é possível verificar a distribuição dos resíduos padronizados em função dos dias de observação do crescimento da planta. Percebe-se que eles estão aumentando a variação das medidas a cada dia de observação. Tal característica é comum em dados que apresentam heterogeneidade de variâncias.

Figura 9: Gráfico dos resíduos padronizados

Na Figura 10, visualiza-se os histogramas dos resíduos ordinários (a) e o histograma de uma distribuição normal (b) com mesma média e variância dos resíduos. Percebe-se que o histograma (a) apresenta uma assimetria negativa em torno do 0 se comparado ao histograma (b).

Figura 10: (a) Histograma dos resíduos ordinários do modelo (b) Histograma de uma distribuição normal com mesma média e variância dos resíduos ordinários do modelo

Após o ajuste do modelo logístico com erros simétricos foi verificada a existência da autocorrelação residual, por meio da inserção de termos autoregressivos de ordem AR(1) e AR(2) no modelo. Porém os novos termos mostraram-se não significativos. Sendo assim, foi considerado que os resíduos não possuem autocorrelação residual.

4.3 Ajuste do modelo logístico com erros com distribuições assimétricas

Tendo como base os valores iniciais das estimativas dos parâmetros do modelo anterior, foi possível iniciar o processo de estimação considerando a distribuição dos erros skew-normal. Tal processo é bastante sensível aos valores iniciais, por isso, pode-se encontrar diferentes estimações e também a não convergência dos processos iterativos utilizados na estimação por máxima verossimilhança da distribuição.

As dificuldades de convergência e sensibilidade da função de estimação também pode ser encontrada em outros estudos como o de Santos, Scalón e Ozaki (2014) que

não conseguiram convergência para produtividade agrícola de todos os municípios estudados. Isso também está de acordo com o estudo de Macerau (2012) que encontrou algumas dificuldades de convergência utilizando funções do software R. Como comentado por Santos (2011), uma desvantagem da distribuição skew-normal em relação a normal é a possibilidade de não ocorrer convergência no algoritmo que estima os parâmetros do modelo.

4.3.1 Ajuste modelo logístico considerando erros normais assimétricos (Skew-normal)

Para o ajuste utilizou-se a função *mle2* do software R (R Core Team, 2013), considerando o método iterativo de Nelder-Mead que faz parte da função citada.

Na Tabela 3 são apresentadas as estimativas dos parâmetros do modelo logístico, com seus respectivos erros padrão, considerando erros normais assimétricos.

Para o parâmetro β_1 foi obtido uma estimativa de aproximadamente 168,9cm, que representa a altura assintótica das plantas. Comparando tal valor com o obtido no modelo logístico com erros normais, há um aumento expressivo de 5,3cm no valor da altura assintótica das plantas de milho.

O ajuste apresentou uma estimativa para o parâmetro da taxa de velocidade do crescimento β_3 de aproximadamente $0,12dia^{-1}$, ou seja, o ajuste do modelo logístico com erros normais assimétricos teve uma estimativa da taxa de velocidade do crescimento maior que o modelo logístico com erros normais. Sendo assim, esse parâmetro indica que o valor assintótico da altura da planta do milho é atingido mais rápido se considerado o modelo logístico com erros normais assimétricos.

O modelo logístico com erros normais assimétricos ajustado aos dados da altura da planta do milho apresentou um ponto de inflexão de $\frac{\hat{\beta}_2}{\beta_3} \cong 40$ dias, o que indica até 40 dias após semeadura da planta de milho houve uma velocidade de crescimento máximo, após esse dia ocorreu uma desaceleração do crescimento das plantas de milho. Esse resultado é semelhante ao resultado obtido quando foi considerado o modelo logístico com erros normais.

O parâmetro de assimetria, que mede o grau de desvio ou afastamento da simetria, apresentou um valor negativo $\hat{\lambda} = -22,3266$ o que indica uma assimetria negativa na distribuição residual.

Tabela 3: Estimativas dos parâmetros do modelo logístico e seus respectivos erros padrão considerando erros com distribuição normal assimétrica

Parâmetro	Estimativa	Erro padrão	p-valor
β_1	168,9313	1,9512	$< 0,0001$
β_2	5,0371	0,1649	$< 0,0001$
β_3	0,1243	0,0039	$< 0,0001$
λ	-22,3266	1,4351	$< 0,0001$
σ	5,4062	1,0287	$< 0,0001$

Considerando que a estimativa do incremento no crescimento para cada unidade de tempo é dada pela taxa de crescimento instantânea absoluta (TCI), para o modelo logístico com erros normais assimétricos os valores obtidos são apresentados na Figura 11.

Observa-se que da semeadura até 122 após a mesma, houve uma variação de até 5,25cm de crescimento ao dia. Esse crescimento máximo foi obtido em 40 dias após a semeadura da planta do milho. Verifica-se também que após o 40 dia, há uma diminuição diária do crescimento até o último dia de observação. Comparando com o que foi obtido no modelo normal, o modelo logístico com erros com distribuição normal assimétrica apresentou um aumento nessa taxa de 0,77cm, ou seja, considerando a estimação dos parâmetros deste modelo, a planta tem um incremento no crescimento maior do que o modelo logístico com erros normais.

Figura 11: Taxa de crescimento instantânea (TCI) em cm/dia - modelo logístico normal assimétrico

Com relação a taxa de maturidade absoluta (TMA) média, o modelo logístico com erros normais assimétricos também apresentou um aumento na proporção do crescimento diário relativo ao valor assintótico, se comparado ao modelo logístico com erros normais. Foi obtido uma taxa de maturidade absoluta de $TMA = 1,58\%$, o que indica um aumento de 1,58% no crescimento da altura da planta de milho ao dia, relativo ao valor assintótico de 168,9cm.

Com relação aos resíduos do modelo considerando erros normais assimétricos, a Figura 12 mostra o histograma dos resíduos ordinários. É possível identificar a assimetria à esquerda contida nos mesmos.

Figura 12: Histograma dos resíduos do modelo logístico com erros normais assimétricos

O modelo logístico considerando erros normais assimétricos apresentou um máximo do logaritmo natural da função de verossimilhança de $-1165,06$, tal valor é necessário para o cálculo do teste da razão de verossimilhança. O modelo ainda teve o critério de informação de Akaike de $AIC = 2340,114$.

Para realizar uma comparação dos modelos logístico considerando erros normais e normais assimétricos, foi feito o teste da razão de verossimilhança. Segundo Colosimo e Giolo (2006), este teste é baseado na função de verossimilhança e envolve a comparação dos valores do logaritmo da função de verossimilhança maximizada sem restrição e sob a hipótese nula.

Considerando que o modelo com erros normais apresentou um máximo do logaritmo natural de verossimilhança de $-1175,7520$ e o modelo em que considerou-se erros assimétricos apresentou o valor de $-1165,06$ para a mesma medida, então, obteve-se o valor do teste da razão de verossimilhança de $TRV = 21,3848$ que comparando-o com o valor tabelado $(\chi^2_{1;0,95}) = 3,8410$ rejeita-se a hipótese nula de que o parâmetro λ é nulo, ou seja, o modelo considerando os erros normais assimétricos é significativo comparado ao modelo com erros normais.

4.3.2 Ajuste modelo logístico considerando erros t-student assimétricos (Skew-t)

Na Tabela 5 são apresentadas as estimativas dos parâmetros do modelo logístico com erros t-student assimétricos, com seus respectivos erros padrão.

Com relação às estimativas dos parâmetros, observa-se que o modelo logístico com erros t-student assimétricos obteve estimativas mais próximas às estimativas obtidas no ajuste do modelo logístico com erros normais do que o modelo logístico com erros normais assimétricos. O parâmetro assintótico da altura da planta β_1 obteve uma estimativa de aproximadamente 167,6cm, ou seja, isso reflete a altura das últimas medições da altura da planta do milho. Já para o parâmetro que mede a velocidade do crescimento da planta (β_3), foi obtido uma estimativa de $0,1041dia^{-1}$, ou seja, comparando esse parâmetro com o mesmo obtido no modelo logístico onde foram considerados erros normais assimétricos, o ajuste do modelo logístico com erros t-student indica que a planta chega a maturidade mais lentamente.

Considerando o ponto em que há a desaceleração do crescimento, o ponto de inflexão da curva, o modelo logístico com erros t-student assimétricos atingiu essa desaceleração em $\frac{\hat{\beta}_2}{\hat{\beta}_3} \cong 40$ dias, ou seja, até o dia 40, a planta da cultura do milho do híbrido transgênico 30F35 Y (Yieldgard) atinge sua velocidade máxima de crescimento diário. Comparando com os demais ajustes, foi obtido o mesmo ponto em 40 dias após a semeadura para todos os modelos ajustados. O parâmetro (α) que modela a assimetria dos erros apresentou uma estimativa de $-0,016$, ou seja, também uma assimetria negativa conforme encontrada no modelo logístico com erros normais assimétricos. Porém, o mesmo apresentou um nível descritivo de 0,9553, ou seja, esse parâmetro não é significativo para o modelo, sendo assim esse modelo assimétrico não seria adequado no estudo. Já o parâmetro dos graus de liberdade de uma distribuição t-student apresentou uma estimativa de aproximadamente 7,6.

Tabela 4: Estimativas dos parâmetros do modelo logístico e seus respectivos erros padrão considerando erros com distribuição t-student assimétrica

Parâmetro	Estimativa	Erro padrão	p-valor
β_1	167,5888	3,4989	$< 0,0001$
β_2	4,1858	0,2066	$< 0,0001$
β_3	0,1041	0,0044	$< 0,0001$
α	-0,0160	0,2867	0,9553
ω	12,2076	0,9353	$< 0,0001$
ν	7,5967	2,7607	0,0059

A taxa de crescimento instantânea absoluta (TCI) obtida considerando a distribuição t-student assimétrica varia até 4,36cm conforme é apresentado na Figura 13, ou seja, do período da semeadura até 122 dias houve um crescimento diário de até 4,36cm. Esse valor máximo foi obtido em 40 dias após a semeadura da planta do milho. Verifica-se também que após o 40º dia há uma diminuição diária do crescimento até o último dia de observação. Esse comportamento também foi observado quando foi ajustado o modelo logístico com erros normais e erros normais assimétricos.

Figura 13: Taxa de crescimento instantânea (TCI) em cm/dia - modelo logístico t-student assimétrico

A taxa de maturidade absoluta média obtida no modelo logístico considerando erros t-student assimétricos foi de $TMA = 1,11\%$, o que indica um aumento de $1,11\%$ ao dia relativo ao valor assintótico de $167,6\text{cm}$. Comparando a taxa de maturidade absoluta obtida no modelo logístico com erros t-student assimétricos aos demais ajustes, houve uma diminuição no valor dessa taxa, ou seja, se considerado o modelo logístico com erros t-student assimétricos para o ajuste aos dados da altura da planta do milho, o crescimento diário percentual é menor que os obtidos nos demais ajustes.

Na Figura 14 é mostrado o histograma dos resíduos ordinários do modelo logístico com erros t-student assimétricos. Também apresentando uma leve assimetria negativa.

Figura 14: Histograma dos resíduos do modelo logístico com erros t-student assimétricos

Realizando o teste de razão de verossimilhança para comparar os modelos com erros normais assimétricos e com erros t-student assimétrico, e foi obtido em módulo o valor de $TRV = 22,2200$. Conclui-se com isso que o modelo logístico com erros t-student também é significativo para se ajustar aos dados da altura das plantas de milho.

4.4 Escolha do modelo

Os ajustes dos modelos logístico considerando erros normais, erros normais assimétricos e erros t-student assimétricos são mostrados na Figura 15. Todos os modelos apresentaram um ajuste satisfatório aos dados de altura, em cm, de plantas de milho e descrevem bem a curva de crescimento.

Figura 15: Ajuste do modelo logístico

Apesar dos três modelos apresentarem um ajuste satisfatório, faz-se necessário escolher o que melhor se ajuste comparado aos demais.

Segundo Burnham e Anderson (2004), a seleção do modelo deve ser baseada em um critério bem justificado e deve ser baseado em uma filosofia sobre modelos e inferência estatística baseada nos modelos.

Pelo critério de informação de Akaike, que segue na Tabela 5, o modelo que melhor se ajusta é o que apresenta seu menor valor. Sendo assim, o modelo logístico considerando erros normais assimétricos, apresentou o melhor ajuste, comparado aos demais, obtendo $AIC = 2340,1140$.

Tabela 5: Critério de informação de Akaike - AIC dos modelos logísticos ajustados

Modelo	AIC
Logístico normal	2359,5050
Logístico skew-normal	2340,1140
Logístico skew-t	2364,3320

Os ajustes obtidos apresentaram um ponto de inflexão da curva em 40 dias após a semeadura, indicando que a planta obteve sua velocidade máxima de crescimento até os 40 dias, após esse dia, ocorre uma desaceleração no crescimento. Porém, os ajustes apresentaram valores assintóticos diferentes. Sendo aproximadamente 163,6cm para o modelo logístico que considerou-se erros normais, 168,9cm para o modelo logístico com erros normais assimétricos e 167,6cm para o modelo logístico com erros t-student.

Com relação a taxa de velocidade do crescimento, os modelos logístico normal e logístico t-student apresentaram valores aproximados de $0,11dia^{-1}$ e $0,10dia^{-1}$, respectivamente. Já o modelo logístico normal assimétrico apresentou uma taxa de velocidade do crescimento de $0,12dia^{-1}$.

O modelo selecionado apresentou o seguinte ajuste mostrado na Figura 16:

Figura 16: Ajuste do modelo logístico normal assimétrico à altura da planta do milho

5 CONCLUSÃO

Os ajustes dos modelos logísticos considerando normalidade nos erros, assimetria normal e assimetria t-student, mostraram se ajustar bem aos dados da altura da planta do milho do híbrido transgênico 30F35 Y (Yieldgard).

O modelo logístico considerando erros normais assimétricos foi o que apresentou melhor ajuste aos dados de altura da planta do milho do híbrido transgênico 30F35 Y (Yieldgard), segundo o critério de informação de Akaike (AIC).

O ajuste apresentou um ponto de inflexão da curva em 40 dias após a semeadura, ou seja, até esse dia foi obtido a velocidade máxima do crescimento. O valor assintótico obtido foi de aproximadamente 168,9cm, ou seja, a planta do milho atinge sua maturidade quando obtém 168,9cm. Já a taxa de velocidade do crescimento obtida no ajuste foi de $0,12dia^{-1}$. O parâmetro de assimetria apresentou valor negativo de aproximadamente $-22,33$. A taxa de crescimento diário na altura da planta do milho foi de até 5,25cm obtida em 40 dias após a semeadura.

REFERÊNCIAS

- AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control**, Piscataway, v. 19, p. 716–723, 1974.
- ANDRADE, D. F.; OGLIARI, P. J. **Estatística para as Ciências Agrárias e Biológicas com noções de experimentação**. 3. ed. Florianópolis, SC: ufsc, 2013. 475 p.
- AZZALINI, A. A class of distributions which includes the normal ones. **Scandinavian Journal of Statistics**, Padova, v. 12, p. 171–178, 1985.
- AZZALINI, A.; CAPITANIO, A. Statistical application of the multivariate skew normal distribution. **Journal of the Royal Statistical Society**, Padova, v. 61, n. 3, p. 579–602, 1999.
- . Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. **Journal of the Royal Statistical Society: Series B**, Padova, v. 65, p. 367–389, May 2003.
- AZZALINI, A.; GENTON, M. Robust likelihood methods based on the skew-t and related distributions. **International Statistical Review**, Padova, v. 76, n. 1, p. 106–129, 2008.
- BATES, D. M.; WATTS, D. G. **Nonlinear Regression Analysis and Its Applications**. New York: John Wiley & Sons, 1988. 365 p.
- BERGAMASCO, A. F.; AQUINO, L. H. de; MUNIZ, J. A. Ajuste de modelos não-lineares a dados de crescimento de fêmeas da raça holandesa. **Ciência e Agrotecnologia**, Lavras, v. 25, n. 2, p. 235–241, mar./abr. 2001.
- BOLFARINE, H.; SANDOVAL, M. C. **Introdução à Inferência Estatística**. Rio de Janeiro: SBM, 2010. 159 p.
- BOLKER, B.; TEAM, R. D. C. **bbmle: Tools for general maximum likelihood estimation**. [S.l.], 2014. R package version 1.0.17. Disponível em: <<http://CRAN.R-project.org/package=bbmle>>.
- BURNHAM, K. P.; ANDERSON, D. R. Multimodel inference: Understanding aic and bic in model selection. **Sociological Methods & Research**, Washington, v. 33, n. 2, p. 261–304, Nov. 2004.
- CALEGARIO, N.; CALEGARIO, C. L. L.; MAESTRI, R.; DANIELS, R. Melhoria da qualidade de ajuste de modelos biométricos florestais pelo emprego da teoria dos modelos não lineares generalizados. **Scientia Forestalis**, Piracicaba, v. 69, p. 38–50, 2005.
- CAMPOS, A. M. **Uma abordagem bayesiana para alguns modelos de crescimento na presença de assimetria e heterocedasticidade**. 2011. 63 p. Dissertação (Mestrado) — Instituto de Ciências Matemáticas e de Computação - USP, São Carlos, 2011.
- CANCHO, V. G.; LACHOS, V. H.; ORTEGA, E. M. M. A nonlinear regression model with skew-normal errors. **Statistical Papers of Springer-Verlag**, New York, v. 51, p. 547–558, 2010.

COLOSIMO, E. A.; GIOLO, S. R. **Análise de sobrevivência aplicada**. São Paulo: Edgard Blucher, 2006. 370 p.

Companhia Nacional de Abastecimento, CONAB. **Acompanhamento da Safra Brasileira**. 2014. Disponível em: <http://www.conab.gov.br/OlalaCMS/uploads/arquivos/140807085954boletimgraosa_gosto2014.pdf>. Acesso em: 14 nov. 2014.

CRUZ, J. C.; FILHO, I. A. P.; ALVARENGA, R. C.; NETO, M. M. G.; VIANA, J. H. M.; OLIVEIRA, M. F.; MATRANGOLO, W. J. R.; FILHO, M. R. de A. **Sistema de Produção**. 6. ed. Sete Lagoas: EMBRAPA, 2010. 65 p.

CRUZ, R. De la; BRANCO, M. D. Bayesian analysis for nonlinear regression model under skewed errors, with application in growth curves. **Biometrical Journal**, New York, v. 51, n. 4, p. 588–609, 2009.

DICICCIO, T. J.; MONTI, A. C. Inferential aspects of the skew exponential power distribution. **Journal of the American Statistical Association**, Washington, v. 99, n. 466, p. 439–450, Apr 2004.

DRAPER, N. R.; SMITH, H. **Applied Regression Analysis**. 3. ed. New York: John Wiley & Sons, 1998. 706 p.

DUARTE, J. O.; CRUZ, J. C.; GARCIA, J. C.; MATTOSO, M. J. **Sistema de Produção**. 7. ed. [S.l.], 2011.

FANCELLI, A. L.; NETO, D. D. **Produção de milho**. Guaíba: Agropecuária, 2000. 360 p.

FERNANDES, T. .; PEREIRA, A. A.; MUNIZ, J. A.; SAVIAN, T. V. Seleção de modelos não lineares para a descrição da curva de crescimento do fruto do cafeeiro. **Coffee Science**, Lavras, v. 9, n. 2, p. 207–215, abr./jun. 2014.

FILHO, S. M.; SILVA, F. F.; CARNEIRO, A. P. S.; MUNIZ, J. A. Abordagem bayesiana das curva de crescimento de duas cultivares de feijoeiro. **Ciência Rural**, Santa Maria, v. 38, n. 6, p. 1516–1521, Set. 2008.

FREITAS, A. R. Curva de crescimento na produção animal. **Revista Brasileira de Zootecnia**, Viçosa, v. 34, n. 3, p. 786–795, 2005.

GENTON, M. G. **Skew-Elliptical distributions and their applications: A journey Beyond Normality**. Boca Raton: CHAPMAN & HALL, 2004. 396 p.

GODOI, L. G.; BRANCO, M. D. **A distribuição t-assimétrica univariada: propriedades e inferência**. Brasil: [s.n.], 2007. Disponível em: <<http://www.ime.usp.br/cpq/main/arquivos/outros/Luciana20Graziela20de20Godoi.pdf>>. Acesso em: 29 out. 2014.

GUEDES, T. A.; ROSSI, R. M.; MARTINS, A. B. T.; JANEIRO, V.; CARNEIRO, J. W. P. Applying regression models with skew-normal errors to the height of bedding plants of *stevia rebaudiana* (bert) bertonii. **Acta Scientiarum**, Maringá, v. 36, n. 3, p. 463–468, julio-septiembre 2014.

- HARTLEY, H. O. The maximum f-ratio as a short-cut test for heterogeneity of variance. **Biometrika**, Michigan, v. 37, n. 3/4, p. 308–312, 1950.
- HOFFMANN, R.; VIEIRA, S. **Análise de Regressão: Uma Introdução à Econometria**. 3. ed. São Paulo: HUCITEC, 1998. 379 p.
- KSHIRSAGAR, A. M.; SMITH, W. B. **Growth Curves**. New York: Marcel Dekker, 1995. 359 p.
- LYRA, G. B.; SOUZA, J. L.; LYRA, G. B.; TEODORO, I.; FILHO, G. M. Modelo de crescimento logístico e exponencial para o milho br 106, em três épocas de plantio. **Revista Brasileira de Milho e Sorgo**, Sete Lagoas, v. 7, n. 3, p. 211–230, 2008.
- MACERAU, W. M. de O. **Comparação das distribuições α -estável, normal, t de student e Laplace assimétricas**. Abril 2012. 95 p. Dissertação (Mestrado) — Universidade Federal de São Carlos, São Carlos, 2012.
- MARTINS, M. C. **Análise de Resíduos Projetados em Modelos Não-Linear de Regressão: Uma Aplicação em Crescimento de Frutos**. 2013. 90 p. Dissertação (Mestrado) — Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2013.
- MAZUCHELI, J.; ACHCAR, J. A. Algumas considerações em regressão não linear. **Acta Scientiarum**, Maringá, v. 24, n. 6, p. 1761–1770, 2002.
- MAZUCHELI, J.; SOUZA, R. M. de; PHILIPPSEN, A. S. Modelo de crescimento de gompertz na presença de erros normais heterocedásticos: Um estudo de caso. **Revista Brasileira de Biometria**, Jaboticabal, v. 29, n. 1, p. 102–121, jan./mar. 2011.
- MAZZINI, A. R. de A.; MUNIZ, J. A.; AQUINO, L. H. de; SILVA, F. F. E. Análise da curva de crescimento de machos hereford. **Ciência e Agrotecnologia**, Lavras, v. 27, n. 5, p. 1105–1112, set./out. 2003.
- MONTGOMERY, D. C.; PECK, E. A.; VINING, G. G. **Introduction to Linear Regression Analysis**. 4. ed. New York: John Wiley & Sons, 2006. 621 p.
- MONTI, A. C. **Inference in the Skew-t and related distributions**. 2008. Disponível em: <<http://www.eief.it/files/2008/11/monti-8-maggio.pdf>>. Acesso em: 14 nov. 2014.
- MORETTIN, P. A.; TOLOI, C. M. C. **Análise de Séries Temporais**. São Paulo: Edgard Blucher, 2004. 535 p.
- MOURA, M.; SOUZA, L.; SILVA, T.; SOARES, J.; BRANDÃO, J. do C. E. Modelos de crescimento para feijão-caupi e o milho, sob sistemas de plantio exclusivo e consorciado, no semiárido brasileiro. **Revista Brasileira de Agrometeorologia**, Santa Maria, v. 16, n. 3, p. 275–284, 2008.
- NETTO, J. C. **Infestação e dados de *Dichelops melacanthus* (Dallas, 1851) (Heteroptera: Pentatomidae) em híbridos transgênicos e convencionais de milho, submetidos ao controle químico**. 2013. 56 p. Dissertação (Mestrado) — Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista “Júlio de Mesquita Filho”, Jaboticabal, 2013.

PEREIRA, J. M.; MUNIZ, J. A.; SILVA, C. A. Nonlinear models to predict nitrogen mineralization in an oxisol. **Scientiae Agricola**, Piracicaba, v. 62, n. 4, p. 395–400, 2005.

PINHEIRO, J.; BATES, D.; DEBROY, S.; SARKAR, D.; R Core Team. **nlme: Linear and Nonlinear Mixed Effects Models**. [S.l.], 2013. R package version 3.1-113.

PRADO, T. K. L. do; SAVIAN, T. V.; MUNIZ, J. A. Ajuste dos modelos gompertz e logístico aos dados de crescimento de frutos de coqueiro anão verde. **Ciência Rural**, Santa Maria, v. 43, n. 5, p. 803–809, maio 2013.

PUIATTI, G. A.; CECON, P. R.; NASCIMENTO, M.; PUIATTI, M.; FINGER, F. L.; SILVA, A. R. da; NASCIMENTO, A. C. C.. Análise de agrupamento em seleção de modelos de regressão não lineares para descrever o acúmulo de matéria seca em plantas de alho. **Revista Brasileira de Biometria**, Jaboticabal, v. 31, n. 3, p. 337–351, 2013.

R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2013. Disponível em: <<http://www.R-project.org/>>. Acesso: 17 jan. 2014

RATKOWSKY, D. A. **Nonlinear Regression Modeling, a unified Practical Approach**. New York: Marcel Dekker, 1983. 276 p.

_____. **Handbook of Nonlinear Regression Models**. New York: Marcel Dekker, 1990. 241 p.

RAWLINGS, J. O.; PANTULA, S. G.; DICKEY, D. A. **Applied Regression Analysis: A Research Tool**. 2. ed. New York: Springer texts in statistics, 1998. 657 p.

RIAZOSHAMS, H.; MIDI, H. A nonlinear regression model for chickens growth data. **European Journal of Scientific Research**, França, v. 35, n. 3, p. 393–404, Aug 2009.

RITCHIE, S. W.; HANWAY, J. J.; BENSON, G. O. **Como a planta de milho se desenvolve**. 2003. (Arquivo do Agrônomo, 103). Disponível em: <<http://brasil.ipni.net/ipniweb/region/brasil.nsf/0/DE02520B8765B8D683257AA0003C46A6/FILE/Encarte103.pdf>>. Acesso em: 14 nov. 2014.

ROSSI, R. M.; SANTOS, L. A. dos. Modelo bayesiana para curva de crescimento de codornas assumindo assimetria nos erros. **Semina: Ciências Agrárias**, Londrina, v. 35, n. 3, p. 1637–1648, maio-jun 2014.

SANTOS, C. O. **A distribuição skew-normal como modelo para a produtividade de milho aplicada ao seguro agrícola**. 2011. 92 p. Dissertação (Mestrado) — Universidade Federal de Lavras, Lavras, 2011.

SANTOS, C. O.; SCALON, J. D.; OZAKI, V. A. A distribuição skew-normal aplicada ao seguro agrícola. **2nd Risk Management and Commodity Trading Conference**, Piracicaba, p. 1–26, 2014.

SAVIAN, T. V. **Estimação dos Parâmetros no Modelo para Degradabilidade IN SITU de Mestens e Loftens**. 2005. 55 p. Dissertação (Mestrado) — Universidade Federal de Lavras, Lavras, 2005.

SEBER, G. A. F.; WILD, C. J. **Nonlinear Regression**. New York: John Wiley & Sons, 1989. 768 p.

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

SOUZA, G. da S. e. **Introdução aos Modelos de Regressão Linear e Não-Linear**. Brasília: Embrapa, 1998. 489 p.

SOUZA, L. S. B.; MOURA, M. S. B.; SILVA, T. G. F.; SOARES, J. M.; SANTOS, W. S. Modelo de crescimento para cultura do milho sob diferentes sistemas de plantio nas condições do semi-árido nordestino. **CONGRESSO BRASILEIRO DE AGROMETEOROLOGIA**, 15., 2007, Aracaju: UFSE, 2007. 5 p.

TERRA, M. F.; MUNIZ, J. A.; SAVIAN, T. V. Ajuste de modelos não-lineares aos dados de crescimento de frutos da tamareira-anã (*phoenix roebelenii o'brien*). **Magistra**, Cruz das Almas, v. 22, n. 1, p. 01–07, jan./mar. 2010.

ANEXOS

Programação do software R

```
##### Dados #####
data=read.csv("C:/Users/Rick/Documents/Arquivos Rick/ESALQ/Dissertação/
             dados/milho4acorr_semout_mais_m2.csv",h=T)
a=data$X # dias de observação
b=data$y # variável resposta

##### Análise exploratória #####
with(data,tapply(b,X,var)) # Variância em cada dia de observação
require(moments) # pacote da função skewness
with(data,tapply(b,X,skewness)) # Assimetria em cada dia de observação
boxplot(b~a) # boxplot
plot(a,b,xlab="Tempo (dias)", ylab="Altura (cm)") # dispersão dos dados

##### plot do crescimento de cada planta #####
rept=as.factor(data$rept)
library(lattice)
xyplot(b~a|rept, type="l",cex=.8)

##### Valores iniciais por meio da linearização do modelo logístico #####
in_log <- 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)
}
A1=in_log(a,b)[1];A1
B1=in_log(a,b)[2];B1
K1=in_log(a,b)[3];K1

##### Ajuste do Modelo logístico com erros simétricos #####
require(nlme)
```

```

kkk=gnlsControl(nlsTol=0.1,opt = c("nlminb", "optim"))
regr_log<-gnls(b~((A)/(1+exp(B-K*a))),start=c(A=A1,B=B1,K=K1),
              data=data,control=kkk)

##### Ponto de inflexao #####
regr_log$coefficients[2]/regr_log$coefficients[3]

##### Obtenção dos resíduos #####
d_log=residuals(regr_log) # resíduos ordinários
ri_pad=d_log/sqrt(var(d_log)) # resíduo padronizado
plot(a,ri_pad, ylab="Resíduos padronizados",xlab="Dias de observação")

##### Teste de Normalidade residual #####
shapiro.test(d_log)

##### Envelope simulado #####
require(car)
qqPlot(d_log,xlab="Quantis normais",ylab="Resíduos",main="")

##### Verificação da existência de auto-correlação residual #####
require(nlme)
regr_log1=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=data,control=kkk)
regr_log2=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=data,control=kkk)
anova(regr_log,regr_log1)
anova(regr_log,regr_log2)

##### Histograma comparando a distribuição dos residuos #####
##### com uma normal de mesma média e mesmo desvio padrão #####
par(mfrow=c(1,2),pty="s")
hist(d_log,freq=FALSE,col=gray(.3),xlab='', ylab='densidade',main="a")
hist(rnorm(290, 0.8447074, 13.94708),freq=FALSE,col=gray(.7),xlab='',
      ylab='densidade',main="b")

##### Ajuste do modelo com erros normais assimétricos #####
n=length(a)
vero2=rep(0,n)

```

```

lvero2<-function(a,b,beta1,beta2,beta3,lam,sigma){
  eta=beta1/(1+exp(beta2-beta3*a))
  psi=sigma^2+lam^2
  ki=lam/(sigma*sqrt(psi))*(b-eta+sqrt(2/pi)*lam)
  di=((b-eta+sqrt(2/pi)*lam)^2)/psi
  vero2=log(2)-0.5*log(2*pi)-0.5*log(psi)-di/2+log(pnorm(ki))
  soma=-sum(vero2)
  return(soma)
}
require(bbmle)
fit_skn<-mle2(lvero2,start=list(beta1=168.9313,beta2=5.0371,beta3=0.1243,
  lam=-22.37,sigma=5.46),data=list(a=a,b=b),
  method = "Nelder-Mead",hessian=TRUE)

##### ponto inflexao #####
coef(fit_skn)[2]/coef(fit_skn)[3]

##### Estimativas #####
beta1=coef(fit_skn)[1];beta1
beta2=coef(fit_skn)[2];beta2
beta3=coef(fit_skn)[3];beta3
lam=coef(fit_skn)[4];lam
sigma=coef(fit_skn)[5];sigma

##### Valores ajustados #####
yaj_skn=beta1/(1+exp(beta2-beta3*a))

##### Resíduos ordinários #####
ri_skn=(b-yaj_skn)
hist(ri_skn,freq=FALSE,main="",xlab="",ylab="Densidade",col=gray(.7))

##### Teste da razão da verossimilhança #####
maxloglik=regr_log$logLik
logL=-1165.06
trv=-2*(maxloglik-logL);trv

##### Ajuste do modelo com erros t-student assimétricos #####
n=length(a)
vero2=rep(0,n)

```

```

lvero2<-function(a,b,beta1,beta2,beta3,alpha,omega,nu){
  mu=beta1/(1+exp(beta2-beta3*a))
  z=(b-mu)/omega
  tau=sqrt((nu+1)/(nu+z^2))
  ze=alpha*z*tau
  f=((gamma((nu+1)/2))*((1+(z^2/nu))^((-nu-1)/2)))/(omega*sqrt(pi*nu)
    *gamma(nu/2))
  F=pt(ze,nu+1)
  vero2=log(2*f*F)
  soma=-sum(vero2)
  return(soma)
}

require(bbmle)
fit_skt<-mle2(lvero2,start=list(beta1=166.84,beta2=4.4089,beta3=0.1091,
  alpha=-1.50,omega= 17.1251,nu=5.67),
  data=list(a=a,b=b), method = "Nelder-Mead",hessian=TRUE)

##### Ponto de inflexão #####
coef(fit_skt)[2]/coef(fit_skt)[3]

##### Estimativas #####
beta1=coef(fit_skt)[1];beta1
beta2=coef(fit_skt)[2];beta2
beta3=coef(fit_skt)[3];beta3
alpha=coef(fit_skt)[4];alpha
omega=coef(fit_skt)[5];omega
nu=coef(fit_skt)[6];nu

##### Valores ajustados #####
yaj_skt=beta1/(1+exp(beta2-beta3*a))

##### Resíduos ordinários #####
ri_skt=(b-yaj_skt)
hist(ri_skt, freq=FALSE,main="",xlab="",ylab="Densidade",col=gray(.5))

##### Teste da razão da verossimilhança #####
maxloglik=-1165.06
logL=-1176.17
trv=-2*(maxloglik-logL);trv

```

```
##### Plot modelos ajustados #####
plot(b ~ a, xlab = "Tempo (dias)", ylab = "Altura (cm)", main = "")
at=seq(1,122,0.5)
b1=163.5832998/(1+exp(4.3616241-0.1096413*at))
b2=168.9289469/(1+exp(5.0370324-0.1243004*at))
b3=167.58881092/(1+exp(4.18580475-0.10414570*at))
lines(at,b1,col="red",lwd=2)
lines(at,b2, col="blue", lwd=2)
lines(at,b3, col="green", lwd=2)
legend('bottomright',c('Modelo logístico normal',
    'Modelo logístico skew-normal','Modelo logístico skew-t'),
    col=c('red','blue','green'),inset=0.01,bty='n',lwd=2)

##### AIC #####
AIC(regr_log,fit_skn,fit_skt)
```