

## O MODELO LOGÍSTICO CONSIDERANDO DIFERENTES DISTRIBUIÇÕES PARA OS ERROS APLICADO A DADOS DE ALTURA DO MILHO

Rick Anderson Freire MANGUEIRA<sup>1</sup>  
Taciana Villela SAVIAN<sup>1</sup>  
Joel Augusto MUNIZ<sup>2</sup>  
Renata Alcarde SERMARINI<sup>1</sup>  
Jacob CROSARIOL NETTO<sup>3</sup>

- RESUMO: O milho é um dos principais cereais produzidos no mundo devido a sua grande utilização na alimentação humana e animal. 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. Esse trabalho teve o objetivo de ajustar o modelo logístico considerando a heterocedasticidade e diferentes distribuições para o erro, a saber, normal, assimétrica normal e assimétrica *t*-student, aplicado a dados de altura (cm) da planta do milho do híbrido transgênico 30F35 Y (Yieldgard), observados ao longo do tempo (dias). Os modelos considerados se ajustaram bem a curva de crescimento da cultura, 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.

<sup>1</sup>Universidade de São Paulo - USP, Escola Superior de Agricultura Luiz de Queiroz, Departamento de Ciências Exatas - LCE, CEP: 13418-900, Piracicaba, SP, Brasil. E-mail: [rickanderson0310@yahoo.com.br](mailto:rickanderson0310@yahoo.com.br); [tvsavian@usp.br](mailto:tvsavian@usp.br); [ralcarde@usp.br](mailto:ralcarde@usp.br)

<sup>2</sup>Universidade Federal de Lavras - UFLA, Departamento de Ciências Exatas, CP: 3037 - CEP: 37200-000, Lavras, MG, Brasil. E-mail: [joamuniz@prpg.ufla.br](mailto:joamuniz@prpg.ufla.br)

<sup>3</sup>Universidade Estadual Paulista Júlio de Mesquita Filho - UNESP, Faculdade de Ciências Agrárias e Veterinárias - FCAV, CEP:14884-900, Jaboticabal, SP, Brasil. E-mail: [jacob.netto@hotmail.com](mailto:jacob.netto@hotmail.com)

## 1 Introdução

O Brasil é um dos principais produtores de milho no mundo, sendo que na safra 2013/2014 foi o terceiro maior produtor mundial deste cereal, produzindo aproximadamente 78.554,0 mil toneladas do grão, em uma área plantada de 15.769,1 mil ha (CONAB, 2014). Segundo Crosariol Netto (2013), 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.

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 mais consumido 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).

A produção no Brasil é dividida em duas épocas de semeadura, sendo a primeira safra chamada de safra de verão, realizada durante período chuvoso, outubro/novembro na região Sudeste, Sul, Norte, Centro-Oeste e também no Oeste da Bahia 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. 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 a ser utilizada para obter conhecimento do crescimento da planta e de seus estádios fenológicos é o estudo de curvas de crescimento. 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 em 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.

Uma informação importante que deve ser levada em consideração nos estudos das curvas de crescimento são as suposições envolvidas na modelagem estatística, dos dados e principalmente dos 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. Em estudo com medidas ao longo do tempo, por exemplo, os resíduos carregam uma dependência e heterogeneidade de variância vindas das medições nas mesmas unidades experimentais, em ocasiões distintos. A normalidade dos resíduos também nem sempre ocorre, sendo que a ocorrência de valores discrepantes podem levar à quebra dessa suposição.

Para contornar esses problemas, várias alternativas são descritas na literatura, como por exemplo a transformação de variável ou atribuir diferentes distribuições para os erros. A utilização de outras distribuições já é realidade no estudo de curvas de crescimento. As distribuições assimétricas, como a 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 de 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 Material

Os dados utilizados neste estudo, adaptados de Crosariol Netto (2013), são referentes a altura da planta de milho do híbrido transgênico 30F35 Y (Yieldgard).

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.

Para este trabalho foi selecionado apenas um dos híbridos utilizados no experimento original 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.

Alguns fatores físicos e não controlados podem afetar o crescimento da planta e também na coleta de dados. Casos como ataques de pragas que impedem a translocação de nutrientes tornando as plantas suscetíveis ao vento, podem causar o decréscimo da planta ao longo do tempo, a derrubada ou quebra da planta, entre outros. Devido a esse fato de condições do experimento, de um total de

altura de 80 plantas mensuradas, optou-se pela utilização de apenas 58 plantas consideradas sadias e que não sofreram interferências de fatores físicos.

### 3 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. A parametrização utilizada no trabalho foi:

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

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;  $\beta_1$  representa a assíntota horizontal, ou seja, o ponto de estabilização do crescimento da planta de milho em centímetros (cm);  $\beta_3$  a taxa de expansão ou velocidade com que a planta atingia seu valor assintótico em cm/dia/cm; e  $\varepsilon_i$  é o erro experimental. Com relação ao parâmetro  $\beta_2$ , ele não apresenta uma interpretação biológica direta, sendo ele um parâmetro de locação e está relacionado ao ponto de inflexão da curva.

Outras interpretações podem ser obtidas a partir da função estudada e Freitas (2005) discute algumas dessas interpretações. A taxa de crescimento absoluta instantânea (TCI) estima o incremento de crescimento para cada unidade de tempo e é dada pela derivada do modelo em relação ao tempo, ou seja,  $TCI = \frac{\partial y}{\partial x}$ . A taxa de maturidade absoluta (TMA) é encontrada pela razão da TCI em relação ao crescimento assintótico  $\beta_1$ , indicando a proporção de crescimento diário relativa ao valor assintótico  $\beta_1$ .

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.

A autocorrelação residual, também característica de estudos com medidas repetidas ao longo do tempo, pode ser 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. 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  $\varepsilon_i$  depende apenas de  $\varepsilon_{i-1}$  e do ruído no instante  $i$ .

A estimação dos parâmetros dos modelos não lineares pode ser realizada pelo método dos mínimos quadrados generalizados (MQG). Nesse método 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 residuais em sua diagonal principal e covariâncias fora da diagonal:

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

O estimador de MQG é dado pela equação  $\hat{\beta}_G = (\mathbf{X}'\mathbf{G}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{G}^{-1}\mathbf{Y}$ . Na maioria dos problemas com modelos não lineares, é mais prático encontrar as estimativas de mínimos quadrados por procedimento de procura numérica direta do que inicialmente obter equações normais, e então utilizar métodos numéricos iterativos para encontrar a solução dessas equações. Segundo Mazucheli e Achcar (2002), as soluções das equações normais podem ser extremamente difíceis de serem obtidas e algum método deve ser utilizado. Deve-se obter estimadores por métodos numéricos iterativos, começando com um conjunto de valores iniciais atribuídos aos parâmetros para que se inicie o método (RATKOWSKY, 1990).

Outro método que pode ser utilizado é o de máxima verossimilhança, que 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  $\Theta$  é o espaço paramétrico, então a função de verossimilhança de  $\theta$  correspondente a amostra aleatória observada é dada por:

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

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

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.

## 4 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 (CRUZ; BRANCO, 2009). 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, conforme

apresentado em Cruz e Branco (2009), por meio de uma abordagem bayesiana para modelos não lineares com erros não normais.

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

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.

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 bayesiano 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, para machos e fêmeas, foram os que melhor se ajustaram aos dados.

Já Guedes *et al.* (2014) ajustaram modelos de regressão com erros normais e normais assimétricos aos dados de altura de plantas, e o modelo bayesiano 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  $t$ -student assimétrica adotada para os erros foi a que melhor se ajustou aos dados.

Estudos considerando distribuições assimétricas dos erros são frequentemente utilizadas levando em consideração diferentes enfoques. Neste trabalho, as distribuições assimétricas abordadas foram da normal e  $t$ -student, sob o enfoque frequentista.

A distribuição normal assimétrica foi introduzida por Azzalini (1985) e é útil para estudos que apresentam assimetria nos dados ou nos erros de modelos, 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.

Segundo Azzalini (1985), uma variável aleatória  $Z$  tem distribuição normal assimétrica com parâmetro  $\mu$ , parâmetro de escala  $\sigma^2$  e parâmetro de assimetria  $\lambda$ ,

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  $\phi(\cdot)$  denota a função de densidade de uma distribuição normal univariada com média  $\mu$  e variância  $\sigma^2 > 0$  e  $\Phi(\cdot)$  a função de distribuição acumulada univariada normal padrão. As propriedades da distribuição podem ser encontradas em Azzalini (1985).

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, aos dados com estrutura assimétricas e com caudas pesadas, sendo assim, mais eficaz do que a distribuição normal assimétrica, neste caso.

Tal distribuição com parâmetros de locação e escala é abordada em Azzalini e Capitanio (2003), onde também são descritas suas propriedades. Considerando a notação em que  $Y \sim ST(\xi, \omega, \alpha, \nu)$ , a função de distribuição da  $t$ -student assimétrica pode ser escrita pela parametrização utilizada por Azzalini e Genton (2008), como sendo:

$$f(y; \xi, \omega, \alpha, \nu) = 2\omega^{-1}t(z; \nu)T\left(\alpha z\sqrt{\frac{\nu + 1}{\nu + z^2}}; \nu + 1\right),$$

com  $y \in \mathbb{R}$ ,  $\xi \in \mathbb{R}$  o parâmetro de locação,  $\omega > 0$  parâmetro de escala,  $\alpha \in \mathbb{R}$  parâmetro de assimetria,  $\nu$  graus de liberdade e ainda,  $z = \frac{y - \xi}{\omega}$ ,  $t$  e  $T$  denotam a função densidade  $t$ -student padrão e função de distribuição acumulada, respectivamente.

## 5 Diagnóstico do modelo

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 pela razão da maior e menor variância obtida em relação aos dias de observação, que é comparada com o valor crítico da tabela de Hartley, com nível de significância  $\alpha$  e  $\nu$  graus de liberdade, e caso a estatística do teste seja maior que o quantil da distribuição, a decisão é de que há evidências da presença de heterogeneidade de variâncias.

Para análise de normalidade foi utilizado o gráfico quantil-quantil com envelope simulado que apresenta em seu eixo das ordenadas os resíduos do modelo o qual deseja-se verificar o pressuposto e no eixo das abscissas os quantis da distribuição normal. Outra forma para verificar tal informação é por meio do teste de Shapiro e Wilk (1965) que tem a hipótese investigada por meio do valor descritivo comparado com o nível de significância. Caso o nível de significância seja maior do que o valor descritivo do teste, então rejeita-se a hipótese de normalidade, caso contrário, então há evidências de que os resíduos procedem de uma distribuição normal.

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. Dentre os diversos avaliadores de qualidade de ajuste dos modelos existentes na literatura foram utilizados o Critério de Informação de Akaike (AIC), (AKAIKE, 1974), e o teste da razão de verossimilhanças. No AIC, o melhor ajuste segundo o critério é o modelo que apresenta menor valor do mesmo. O cálculo é realizado pela equação abaixo:

$$AIC = -2\log L + 2p \quad (2)$$

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

Já o teste da razão de verossimilhanças 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. 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 é realizada a comparação dos modelos. A estatística de teste é dada por:

$$TRV = -2\log \left[ \frac{L(\theta_0)}{L(\hat{\theta})} \right] = 2[\log L(\hat{\theta}) - \log L(\theta_0)], \quad (3)$$

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  $H_0 : \theta = \theta_0$  segue uma distribuição Qui-quadrado com  $p$  graus de liberdade, sendo  $H_0$  rejeitada se com um nível de significância  $\alpha$ ,  $TRV > \chi_{p,1-\alpha}^2$ .

No presente trabalho, considerou-se um nível de significância  $\alpha = 5\%$ .

## 6 Resultados

Para o ajuste do modelo logístico considerando normalidade na distribuição dos erros, utilizou-se a função *gnls* do pacote *nlme* (PINHEIRO *et al.*, 2013), do programa R Core Team (2013), que considera a estimação por mínimos quadrados generalizados. Já para o ajuste do modelo logístico em que foram consideradas distribuições assimétricas dos erros, foi usada a função *mle2* do pacote *bbmle* (BOLKER; TEAM, 2014), que utiliza o método da máxima verossimilhança em sua estimação.

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 1). Apesar dos dias de observações não serem equidistantes, o comportamento dos dados aproxima-se de uma curva sigmoideal, sendo que a altura 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 1).

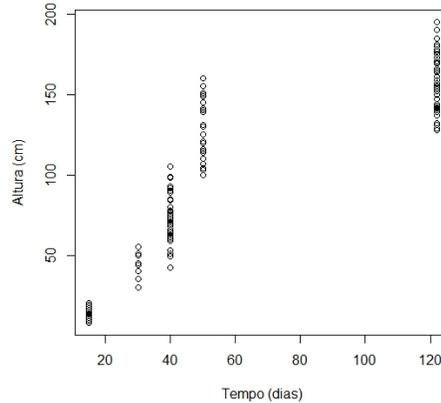


Figura 1 - Gráfico de dispersão dos dados de altura de plantas de milho, em centímetros, ao longo do tempo, em dias após o plantio.

Verifica-se 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. Esse aumento da variabilidade verificada na análise gráfica, pode ser comprovada pelo teste de Hartley para homogeneidade de variâncias.

Percebe-se um claro aumento de variância ao longo dos dias após semeadura. Sendo assim, o teste de Hartley apresentou uma estatística de teste  $F_{max} = 46,32$ , e que comparado ao valor crítico  $F_{(5,57)5\%} \cong 2,04$ , conclui-se pela rejeição da hipótese nula, ou seja, há evidências de que as variâncias são heterogêneas ao longo dos dias que foram realizadas observações (Tabela 1).

Tabela 1 - Variâncias estimadas em cada dia de observação

Dias após semeadura	15	30	40	50	122
Variância	7,75	33,42	162,96	245,91	359,17

Na Tabela 2 são apresentadas as estimativas dos parâmetros do modelo logístico, com seus respectivos erros padrão e níveis descritivos para significância dos parâmetros, considerando os erros normais, normais assimétricos e *t*-student assimétrico.

O processo de estimação do modelos considerando erros assimétricos é bastante sensível aos valores iniciais, por isso, pode-se encontrar diferentes estimativas e também a não convergência dos processos iterativos utilizados na estimação por máxima verossimilhança da distribuição. Como comentado por Santos (2011), uma desvantagem da distribuição normal assimétrica em relação a normal é a possibilidade de não ocorrer convergência no algoritmo que estima os parâmetros

do modelo. As dificuldades de convergência e sensibilidade de estimação dos parâmetros também foram relatadas em outros estudos, como o de Santos, Scalon e Ozaki (2014), que tinham o objetivo de modelar a distribuição de produtividade de milho em 30 municípios do Paraná, sendo que não conseguiram convergência dos estimadores de máxima verossimilhança para todos os municípios estudados. Isso também está de acordo com o estudo de Macerau (2012) que encontrou algumas dificuldades de convergência utilizando funções do R.

Tabela 2 - Estimativas dos parâmetros do modelo, seus respectivos erros padrão, níveis descritivos e critério de informação de Akaike (AIC), de acordo com a distribuição consideradas para os erros

Distribuição residual	Parâmetro	Estimativa	Erro padrão	p-valor	AIC
Normal	$\beta_1$	163,5833	1,8202	<0,0001	2359,51
	$\beta_2$	4,3616	0,1519	<0,0001	
	$\beta_3$	0,1096	0,0039	<0,0001	
Normal assimétrica	$\beta_1$	168,9313	1,9512	<0,0001	2340,11
	$\beta_2$	5,0371	0,1649	<0,0001	
	$\beta_3$	0,1243	0,0039	<0,0001	
	$\lambda$	-22,3266	1,4351	<0,0001	
<i>t</i> -student assimétrica	$\beta_1$	167,5888	3,4989	<0,0001	2364,33
	$\beta_2$	4,1858	0,2066	<0,0001	
	$\beta_3$	0,1041	0,0044	<0,0001	
	$\alpha$	-0,0160	0,2867	0,9553	
	$\nu$	7,5967	2,7607	0,0059	

Os ajustes apresentaram diferentes valores assintóticos. 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 assimétricos. Segundo Freitas (2005), estudando curvas de crescimento animal, o parâmetro A, equivalente ao  $\beta_1$  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  $\beta_1$  reflita a altura assintótica das plantas de milho.

Fernandes *et al.* (2014) estudando os modelos logístico e Gompertz para ajuste da curva de crescimento de massa fresca do fruto do cafeeiro, consideram o parâmetro k sendo o 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  $\beta_3$ , e tal valor representa a velocidade que a planta atinge o valor assintótico ( $\beta_1$ ), os modelos considerando distribuição dos erros normal e *t*-student assimétrico apresentaram

valores aproximados de  $0,11\text{dia}^{-1}$  e  $0,10\text{dia}^{-1}$ , respectivamente. Já o modelo considerando erros com distribuição normal assimétrico apresentou uma taxa de velocidade do crescimento um pouco superior aos demais de  $0,12\text{dia}^{-1}$ .

Outra informação importante obtida com os parâmetros é o ponto de inflexão da curva. Segundo Mischan e Pinho (2014), o ponto de inflexão, é um ponto de transição no crescimento da curva, ou seja, o ponto no qual a TCI passa de crescente para decrescente. Como o modelo logístico é simétrico no ponto de inflexão, tal ponto ocorre quando atingi-se metade do crescimento assintótico, ou seja,  $\frac{\beta_1}{2}$ . A relação dos parâmetros  $\frac{\hat{\beta}_2}{\hat{\beta}_3} \cong 40$  dias para todos os ajustes, indica que até o 40º dia após a semeadura ocorreu a velocidade máxima do crescimento por dia, após esse dia houve uma desaceleração no crescimento da altura da planta.

O parâmetro de assimetria, que mede o grau de desvio ou afastamento da simetria, apresentou um valor estimado negativo  $\hat{\lambda} = -22,3266$  no modelo considerando distribuição dos erros normais assimétricos, e  $\hat{\alpha} = -0,0160$  no modelo em que considerou-se erros *t*-student assimétricos, o que indica uma assimetria negativa na distribuição residual. Porém, o nível descritivo associado a  $\hat{\alpha}$  foi de 0,9553, ou seja, esse parâmetro não é significativo para o modelo, sendo assim esse modelo assimétrico não seria adequado no estudo.

Com relação a taxa de crescimento instantânea (TCI), tem-se que a mesma estima o incremento no crescimento para cada unidade de tempo, ou seja, a cada dia, os modelos ajustados apresentaram o máximo de crescimento em 40 dias após a semeadura, conforme a Figura 2, após esse dia há uma diminuição no crescimento diário da planta do milho até o último dia de observação. Verifica-se que da semeadura até 122 dias após a mesma, houve uma variação de até 4,48cm no crescimento ao dia considerando o modelo logístico com erros normais, uma variação de até 5,25cm considerando o modelo logístico com erros normais assimétricos e para o modelo em que considerou-se erros *t*-student assimétricos obteve-se variação de até 4,36cm ao dia.

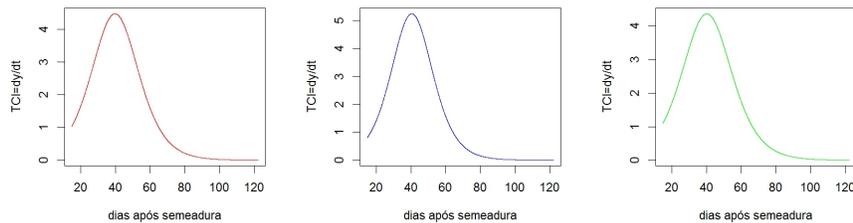


Figura 2 - Taxa de crescimento instantânea (TCI) em cm/dia - modelo logístico normal, assimétrico normal e assimétrico *t*-student, respectivamente.

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 considerando erros com distribuição normal foi 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  $\hat{\beta}_1 = 163,6\text{cm}$ . Já o modelo no qual considerou-se erros com distribuição normal assimétrico apresentou uma TMA de 1,58% relativo ao valor assintótico de 168,9cm. Para o modelo com distribuição dos erros *t*-student assimétrico obteve-se uma TMA de 1,11%, o que indica um aumento de 1,11% ao dia relativo ao valor assintótico de 167,6cm. 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.

Após o ajuste do modelo logístico 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.

Para avaliar a qualidade do ajuste e obter a validação do modelo considerando distribuição dos erros normal, 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 3) 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.

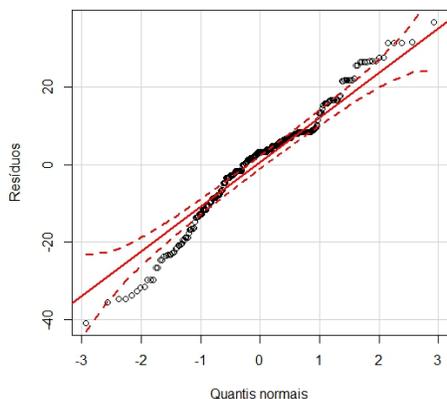


Figura 3 - Gráfico de probabilidades normais com envelope simulado com 95% de confiança.

Na Figura 4, visualiza-se os histogramas dos resíduos ordinários dos modelos considerando distribuição dos erros normais simétricos (a), normais assimétricos (b) e *t*-student assimétricos (c). Percebe-se que os histogramas apresentam uma assimetria negativa em torno do valor zero.

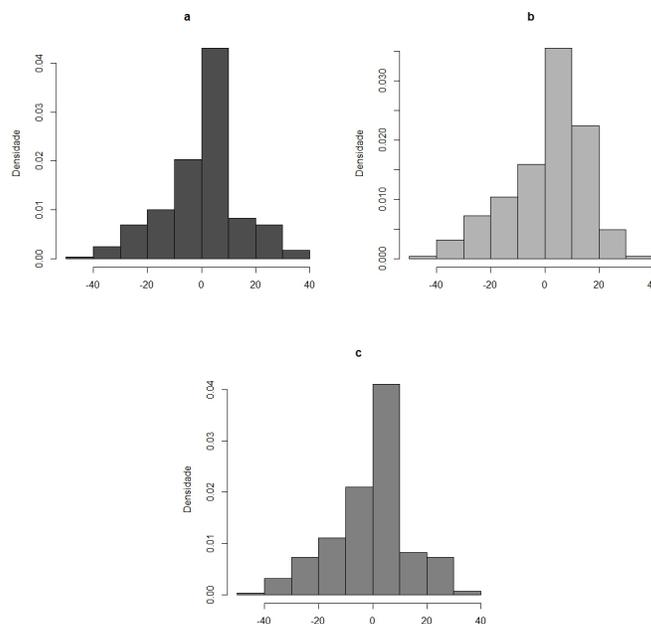


Figura 4 - (a) Histograma dos resíduos do modelo normal, (b) normal assimétricos (c) e *t*-student assimétricos.

Para realizar uma comparação dos modelos logístico considerando erros normais e normais assimétricos, foi feito o teste da razão de verossimilhança. Considerando que o modelo com erros normais apresentou um máximo do logaritmo natural de verossimilhança de  $-1175,75$ , 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 verossimilhanças,  $TRV = 21,38$ , que comparando-o com o quantil da distribuição Qui-quadrado  $(\chi^2_{1;0,95}) = 3,84$ , rejeita-se a hipótese nula de que o parâmetro  $\lambda$  é nulo, ou seja, o modelo considerando os erros normais assimétricos é significativo comparado ao modelo com erros normais. Realizando o teste de razão de verossimilhança para comparar os modelos com erros normais assimétricos e com erros *t*-student assimétrico, foi obtido em valor absoluto  $TRV = 22,22$ . Conclui-se com isso que o modelo logístico com erros *t*-student assimétricos também é significativo para se ajustar aos dados da altura das plantas de milho.

Os ajustes dos modelos logístico considerando erros normais, erros normais assimétricos e erros *t*-student assimétricos são mostrados na Figura 5. Todos os modelos apresentaram um ajuste satisfatório aos dados de altura, em cm, de plantas de milho e descrevem bem a curva de crescimento.

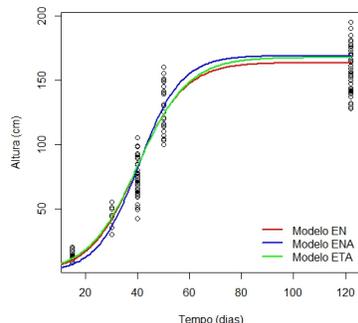


Figura 5 - Ajuste do modelo logístico aos dados de altura (cm) de plantas de milho em função do tempo (dias após a semeadura), considerando os erros com distribuição normal (EN), distribuição normal assimétrica (ENA) e distribuição *t*-student assimétrica (ETA).

Graficamente, os três modelos apresentam um ajuste satisfatório, portanto faz-se necessário escolher o que melhor se ajusta aos dados por meio de algum critério de seleção. Segundo Burnham e Anderson (2004), a seleção do modelo deve ser baseada em um critério bem justificado e também em uma filosofia e inferência estatística baseada nos modelos. Pelo critério de informação de Akaike (Tabela 2), 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  $AIC = 2340,11$ .

## Conclusões

Os modelos logísticos considerando normalidade dos erros, assimetria normal e assimetria *t*-student mostraram um ajuste satisfatório 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, segundo o critério de informação de Akaike (AIC) e pelo teste da razão de verossimilhanças. O ajuste apresentou um ponto de inflexão da curva em 40 dias após a semeadura, o valor assintótico obtido foi de aproximadamente 168,9cm e a taxa de velocidade do crescimento no ajuste foi de  $0,12\text{dia}^{-1}$ . 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.

## Agradecimentos

Ao Programa de Pós-Graduação em Agronomia (Estatística e Experimentação Agrônômica) da ESALQ/USP pela oportunidade e à CAPES pelo apoio financeiro.

MANGUEIRA, R. A. F.; SAVIAN, T. V.; MUNIZ, J. A.; SERMARINI, R. A.; CROSARIOL NETTO, J. Logistic model considering different distributions for errors applied in maize height data. *Rev. Bras. Biom.*, Lavras, v.34, n.2, p.317-333, 2016.

■ **ABSTRACT:** *Maize is one of the main cereals produced in the world due to its wide use in food and feed. Knowledge on plant growth is extremely important to its management. One can obtain this knowledge by the use of growth models to provide information through parameters with biological interpretations summarize the characteristic curve of plant growth. This work aimed to fit the logistic model considering heteroscedasticity and different distributions for error, namely, normal, skew normal and skew t-student applied to plant height data (cm) of the maize transgenic hybrid 30F35 Y (YieldGard ) observed over time (days). The models considered had a good fit to the growth curve of the culture, but the logistic model considering skew normal error was selected as most appropriate for modeling the curve, based on the evaluators used.*

■ **KEYWORDS:** *Nonlinear models; growth curves; skew errors.*

## Referências

AKAIKE, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, Piscataway, v.19, p.716–723, 1974.

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.

AZZALINI, A.; CAPITANIO, A. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society: Series B*, v.65, p.367–389, 2003.

AZZALINI, A.; GENTON, M. Robust likelihood methods based on the skew-t and related distributions. *International Statistical Review*, v.76, n.1, p.106–129, 2008.

BOLFARINE, H.; SANDOVAL, M. C. *Introdução à Inferência Estatística*. Rio de Janeiro: SBM, 2010. 159p.

- BOLKER, B.; TEAM, R Development Core. *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*, v.33, n.2, p.261–304, 2004.
- CAMPOS, A. M. *Uma abordagem bayesiana para alguns modelos de crescimento na presença de assimetria e heterocedasticidade*. Dissertação (Mestrado) — Instituto de Ciências Matemáticas e de Computação - ICMC-USP, 2011.
- COLOSIMO, E. A.; GIOLO, S. R. *Análise de sobrevivência aplicada*. 1. ed. São Paulo: Edgard Blucher, 2006. 370p.
- CONAB, Companhia Nacional de Abastecimento. *Acompanhamento da Safra Brasileira*. 2014. Disponível em: (<http://www.conab.gov.br/OlalaCMS/uploads/arquivos/140807085954boletimgraosagosto2014.pdf>).
- CROSARIOL NETTO, J. *Infestação e dados de Dichelops melacanthus (Dallas, 1851) (Heteroptera: Pentatomidae) em híbridos transgênicos e convencionais de milho, submetidos ao controle químico*. Dissertação (Mestrado) — Universidade Estadual Paulista "Júlio de Mesquita Filho" - Faculdade de ciências agrárias e veterinárias, Jaboticabal, São Paulo, 2013.
- CRUZ, R. De la; BRANCO, M. D. Bayesian analysis for nonlinear regression model under skewed errors, with application in growth curves. *Biometrical Journal*, 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*, v.99, n.466, p.439–450, Apr 2004.
- DUARTE, J. O.; CRUZ, J. C.; GARCIA, J. C.; MATTOSO, M. J. *Sistema de Produção*. 7.ed. [S.l.], 2011.
- FERNANDES, T. J.; PEREIRA, A. A.; MUNIZ, J. A.; SAVIAN, T. V. Seleção de modelos não lineares para a descrição das curva de crescimento do fruto do cafeeiro. *Coffee Science*, v.9, n.2, p.207–215, 2014.
- FREITAS, A. R. Curva de crescimento na produção animal. *Revista Brasileira de Zootecnia*, v.34, n.3, p.786–795, 2005.
- GENTON, M. G. *Skew-Elliptical distributions and their applications: A journey Beyond Normality*. Boca Raton, FL, 2004.
- 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*, v.36, n.3, p.463–468, 2014.

HARTLEY, H. O. The maximum f-ratio as a short-cut test for heterogeneity of variance. *Biometrika*, v.37, n.3/4, p.308–312, 1950.

MACERAU, W. M. O. *Comparação das distribuições  $\alpha$ -estável, normal, t de student e Laplace assimétricas*. Dissertação (Mestrado) — Universidade Federal de São Carlos, São Carlos - SP, Abril 2012.

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

MISCHAN, M. M.; PINHO, S. Z. *Modelos não lineares: funções assintóticas de crescimento*. São Paulo: Cultura Acadêmica, 2014.

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.

R CORE TEAM. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2013. Disponível em: <http://www.R-project.org/>.

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

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

ROSSI, R. M.; SANTOS, L. A. Modelo bayesiana para curva de crescimento de codornas assumindo assimetria nos erros. *Semina: Ciências Agrárias*, 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*. Dissertação (Mestrado) — Universidade Federal de 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*, p.1–26, 2014.

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

Recebido em 13.10.2015.

Aprovado após revisão em 28.01.2016.