

Modelo não linear Gompertz com erro assimétrico para descrição do acúmulo da matéria seca do feijoeiro cultivar bolinha

Kelly Pereira de Lima^{1†}, Felipe Augusto Fernandes², Ricardo Andrade Lira Rabelo³, Augusto Ramalho de Moraes²

¹Programa de Pós-Graduação em Ciência da Computação, Departamento de Computação, Universidade Federal do Piauí (UFPI)

²Programa de Pós-Graduação em Estatística, Instituto de Matemática, Estatística e Computação Científica, Universidade Estadual de Campinas (Unicamp)

³Departamento de Computação, Centro de Ciências da Natureza, Universidade Federal do Piauí (UFPI)

⁴Departamento de Estatística, Instituto de Ciências Exatas e Tecnológicas, Centro de Ciências da Natureza, Universidade Federal de Lavras (UFLA)

Resumo: A análise do acúmulo de matéria seca no decorrer do crescimento do feijoeiro é de grande relevância, pois desempenha um papel crucial como instrumento informativo para orientar práticas de manejo apropriadas e identificar possíveis fatores que impactam o desenvolvimento da planta. A modelagem adequada desse tipo de crescimento pode fornecer subsídios para tal, possibilitando a otimização do manejo da cultura. Sendo assim, busca-se alternativas para a detecção desses tipos de valores atípicos, entre elas pode-se citar as distribuições assimétricas tais como: a normal e a *t*-Student. Desta maneira, este trabalho teve como objetivo ajustar modelos não lineares Gompertz com erro normal, erro normal assimétrico e erro *t*-Student assimétrico para a descrição do acúmulo de matéria seca do feijoeiro cv “Bolinha”. O experimento foi realizado na Universidade Federal de Lavras, no período da safra das águas nos anos de 2006/2007, em um delineamento em bloco causalizados, contendo três repetições; usou-se o esquema fatorial 5 x 8, sendo constituído de cinco densidade de semeadura (75, 145, 215, 285 e 355 mil plantas por hectare) e oito épocas de avaliação (13, 23, 33, 43, 53, 63, 73, 83 dias após emergência) e foi analisado o acúmulo de matéria seca no total. Como resultado da pesquisa os modelos não lineares Gompertz com erro assimétrico se mostraram adequados para descrição do acúmulo de massa seca.

Palavras-chave: Assimetria; curva de crescimento; feijão; outlier.

Gompertz nonlinear model with asymmetric error to describe the dry matter accumulation of the bean cultivar bolinha

Abstract: The analysis of dry matter accumulation during the growth of the common bean plant is of significant relevance as it serves as a critical tool for guiding appropriate management practices and identifying potential factors affecting plant development. Proper modeling of this growth can provide valuable insights, facilitating the optimization of crop management. This study explores alternatives for detecting outliers, including asymmetric distributions such as the normal and *t*-Student distributions. The objective was to fit Gompertz nonlinear models with normal error, asymmetric normal error, and asymmetric *t*-Student error to describe the dry matter accumulation of the “Bolinha” bean cultivar. The experiment was conducted at the Federal University of Lavras during the 2006/2007 rainy season in a randomized block design with three replications, using a 5 x 8 factorial scheme. This included five sowing densities (75, 145, 215, 285, and 355 thousand plants per hectare) and eight evaluation periods (13, 23, 33, 43, 53, 63, 73, and 83 days after emergence), and analyzed total dry matter accumulation. The results demonstrated that Gompertz nonlinear models with asymmetric error were suitable for describing dry matter accumulation.

Keywords: Asymmetry; growth curve; bean; outlier.

† Autor correspondente: kelly.lima.88@gmail.com

Manuscrito recebido em: 15/07/2024

Manuscrito revisado em: 18/09/2024

Manuscrito aceito em: 18/09/2024

Introdução

Os modelos não lineares são amplamente utilizados em várias pesquisas devido à sua estrutura que permite aplicações em diversas áreas do conhecimento científico, como agricultura, biologia, ciências médicas, economia, entre outras. Na análise do crescimento de curvas, esses modelos são empregados para estimar os determinantes do crescimento, uma vez que frequentemente possuem parâmetros com interpretações práticas e de interesse direto para os pesquisadores.

Na literatura, presume-se frequentemente que as observações sigam uma distribuição normal, o que implica que a inferência em modelos não lineares considere a normalidade dos resíduos. No entanto, em muitas situações práticas, os resíduos ou a variável resposta não seguem uma distribuição normal. Isso pode ocorrer, por exemplo, quando os dados são provenientes de uma distribuição com caudas mais pesadas ou menos pesadas do que a distribuição normal (LABRA *et al.*, 2012).

Na prática, frequentemente recorre-se à transformação de dados como método para alcançar a normalidade. Apesar de ter sido bem-sucedido em alguns casos, suas aplicações nem sempre são convincentes. Isso pode acarretar dificuldades na interpretação dos resultados, limitações na aplicabilidade a diferentes conjuntos de dados e variações na transformação necessária para alcançar a normalidade, especialmente em casos de falta de homogeneidade (AZZALINI; CAPITANIO, 1999).

Quando os dados apresentam uma distribuição com comportamento assimétrico, isso pode ser modelado através da classe das distribuições assimétricas. A classe das distribuições assimétricas está sendo amplamente aplicada em diversas áreas do conhecimento, tanto na teoria quanto na prática. Em aplicações práticas, é crucial obter resultados que não sejam influenciados por observações atípicas. A distribuição *t* assimétrica é reconhecida por sua robustez em comparação com a distribuição normal assimétrica, sendo uma alternativa para inferência robusta em diversos tipos de modelos (MAGHAMI; BAHRAMI; SAJADI, 2020), incluindo modelos não lineares.

A análise do crescimento em curvas utiliza modelos lineares e não lineares com erros normalmente distribuídos para estimar os determinantes do crescimento, sendo que os modelos não lineares têm se destacado devido à interpretação prática de seus parâmetros. De acordo com Vieira *et al.* (2008), essa análise, complementada pela absorção de nutrientes ao longo do crescimento das plantas, é fundamental para avaliar os efeitos dos fatores de manejo.

Na cultura do feijoeiro, há pouca ênfase na modelagem do acúmulo de matéria considerando diferentes distribuições para os erros. No entanto, essa cultura frequentemente apresenta valores atípicos ao longo do tempo, ligados a diversos processos fisiológicos das plantas. Tais observações atípicas (outliers) podem significativamente afetar a distribuição dos erros (HUBERT; ROUSSEEUW; AELST, 2008).

Este estudo visa ajustar modelos não lineares Gompertz, considerando diferentes distribuições de erros como normal, normal assimétrica e *t* assimétrica, para descrever o acúmulo de matéria seca na cultivar “Bolinha” de feijoeiro.

Material e Métodos

Os dados utilizados neste trabalho são provenientes de um experimento realizado na área experimental do Departamento de Agricultura da Universidade Federal de Lavras (UFLA), em Lavras - MG. O solo onde o experimento foi conduzido é classificado como Latossolo Vermelho Distrófico, originalmente sob cerrado, e foi manejado sob sistema de plantio convencional com uma aração e duas gradagens.

O experimento foi realizado em delineamento em blocos casualizados, com três repetições, sendo os tratamentos dispostos no esquema fatorial 5 x 8. O fatorial foi constituído por cinco densidades de semeadura: 75, 145, 215, 285 e 385 mil plantas por hectare (ha^{-1}) e oito períodos de avaliação: 13, 23, 33, 43, 53, 63, 73 e 83 dias após a emergência (DAE) das plantas. Foi utilizada a cultivar de feijão do tipo Bolinha, cujas características incluem grãos amarelos e arredondados, hábito de crescimento do tipo II, porte ereto, com cerca de 32 a 33 gramas por 100 grãos e ciclo médio, completando seu desenvolvimento entre 85 e 100 dias desde a germinação até a colheita.

Com a intenção de obter o acúmulo de massa seca da parte aérea ao longo do ciclo da cultura, foram realizadas coletas a cada 10 dias, sendo a primeira coleta realizada aos 13 dias após a emergência (DAE). Em cada coleta, as partes das plantas foram separadas em haste (H), folhas (F), flores e vagens (V), e grãos (G). Os materiais foram pesados em balança de precisão, determinando-se assim o peso da matéria seca de cada parte da planta. Posteriormente, esses dados foram transformados em acúmulo de matéria seca por hectare $kg ha^{-1}$.

Em seguida, foram obtidos os valores observados de acúmulo de matéria seca nas diferentes partes da planta, somando-se cada parte para obter o peso da matéria seca e dos nutrientes na haste (H), nas folhas (F), na combinação de haste e folhas (HF), na combinação de haste, folhas e vagens (HFV), e no total (HFVG). Assim, a variável de interesse para o acúmulo de matéria seca foi o total.

Com a finalidade de estudar o efeito das épocas de semeadura sobre o acúmulo total de matéria seca e descrever como esse acúmulo ocorre para cada densidade de semeadura, realizou-se o desdobramento da interação, analisando a época dentro de cada densidade de semeadura. As análises foram realizadas conforme sugestão de Gomes-Pimentel (2009) e Steel, Torrie e Dickey (1997).

Posteriormente, foi realizada uma análise de variância seguida de uma análise de resíduos com o objetivo de verificar as pressuposições de normalidade, utilizando o teste de Shapiro-Wilk (SHAPIRO; WILK, 1965), homogeneidade de variâncias, com o teste de Bartlett (BARTLETT, 1937), e independência dos dados, com o teste de Durbin-Watson (DURBIN; WATSON, 1950). Além disso, para análise de resíduos onde os pressupostos não foram atendidos, será utilizada a análise de variância com transformação de posto, um método não paramétrico (SAWILOWSKY, 1990).

Conforme Wobbrock (2011), a Análise de Variância com Transformação de Posto Alinhado (ART) foi utilizada, uma técnica estatística baseada no teste de Transformação de Posto (RT), que combina elementos da ANOVA com métodos não paramétricos. A ART é geralmente aplicada quando os pressupostos tradicionais da ANOVA, como normalidade e homocedasticidade, são violados.

O processo de alinhamento da ART remove os efeitos principais de outras variáveis, isolando o efeito da variável de interesse antes da transformação em postos. Isso corrige as interações antes da atribuição dos postos, garantindo que efeitos principais e interações não sejam confundidos nos resultados. A principal diferença está no pré-processamento dos dados: antes de realizar o teste, os dados são transformados em postos, garantindo maior robustez da análise para dados não normais e heterocedásticos.

Para o desdobramento dos dados, foi utilizado o modelo não linear de Gompertz, considerando diferentes distribuições de erro: erro normal, erro normal assimétrico e erro t assimétrico. O modelo não linear Gompertz utilizado para descrever o acúmulo de matéria seca do feijoeiro foi:

$$Y_i = \alpha e^{-e^{k(b-x_i)}} + \varepsilon_i$$

em que Y_i é o acúmulo da matéria seca no total, α é assíntota superior ou acúmulo máximo, b é abscissa relacionada ao ponto de inflexão, k está relacionado com crescimento ou indica o índice de maturidade e o x_i representa os dias após a emergência, com i corresponde às épocas de avaliação (13, 23, 33, 43, 53, 63, 73 e 83 DAE) e ε_i é o erro aleatório associado ao modelo. Foram consideradas as seguintes distribuições para o erro aleatório: normal, normal assimétrica (SN) e t assimétrica (ST).

De acordo com Fernandes *et al.* (2019), os parâmetros da regressão não linear clássica foram estimados utilizando o método dos mínimos quadrados, ajustados por meio do processo iterativo de Gauss-Newton. Quando os resíduos não seguem uma distribuição normal, como no caso de erros normais assimétricos ou distribuições t assimétricas, o método da Máxima Verossimilhança (MV) se apresenta como uma alternativa mais apropriada. Esse método visa estimar os parâmetros que maximizam a função de verossimilhança, ou seja, os parâmetros que conferem a maior probabilidade aos dados observados, dado o modelo ajustado (MYUNG,2003).

Da mesma forma que foi realizada uma análise de resíduos para a análise de variância, também foi empregada na modelagem dos dados. Assim, para verificar as pressuposições de normalidade, independência e homoscedasticidade dos resíduos, foram utilizados os testes de Shapiro-Wilk, Breusch-Pagan e Durbin-Watson, respectivamente.

Com relação à seleção e comparação dos modelos, foi utilizada a abordagem do critério de informação de Akaike (AIC).

Resultados e Discussão

O acúmulo de matéria seca total, composto pela soma da haste, folha, vagem e grão, foi submetido a uma análise descritiva e gráfica para visualizar o comportamento da variável ao longo dos dias após emergência (DAE) em cinco densidades de semeadura (Dens). Para o acúmulo de matéria seca no total, em que consiste na soma da haste, folha, vagem e grão, foi realizada uma análise descritiva e gráfica para visualizar o comportamento da variável em estudo, ao longo dos dias após emergência (DAE) em cinco densidade de semeadura (Dens).

Table 1: Descriptive statistics for total dry matter accumulation in Bolinha bean plants.

	Mínimo	1º Quantil	Mediana	Média	3º Quantil	Máximo	Assimetria	Curtose
Total	27,90	246,50	621,74	979,69	1377,15	4361,28	1,40	4,37

Source: from authors (2024).

A distribuição pode ser simétrica ou assimétrica, sendo classificada como assimétrica positiva ou negativa dentro desta última categoria. Uma distribuição é considerada simétrica quando o coeficiente de assimetria é igual a 0, indicando que média, mediana e moda coincidem. Se a distribuição é assimétrica, pode-se classificar a assimetria como fraca (entre 0 e 0,15), moderada (entre 0,15 e 1) ou forte (acima de 1) (PETRUCCI; OLIVEIRA, 2017)

Neste estudo, a variável em análise apresentou uma assimetria considerada forte, com um coeficiente de assimetria de 1,04. Esse valor indica uma distribuição assimétrica positiva, ou à esquerda.

Neste estudo, foi empregada uma abordagem não paramétrica com transformação de posto, conforme evidenciado na Tabela 2. Verifica-se que a interação entre densidade e dias após emergência é significativa, assim como outras fontes de variação. Isso indica a existência de um efeito de dependência entre densidade de semeadura e dias após emergência.

Table 2: Analysis of variance with rank transformation was carried out to evaluate the sources of variation (FV), including the numbers of degrees of freedom (GL), F values and their respective significance of the F test, for total dry matter in the Bolinha bean cultivar.

FV	GL	Total
Bloco	2	5,89*
Dens	4	56,69**
DAE	7	56,17**
Dens x DAE	28	4,16**

Source: from authors (2024).

Legend: (ns) not significant, (*) and (**) significant by the test performed.

A pressuposições dos modelos foram verificada para cada densidade de semeadura. Ou seja, normalidade residual, homogeneidade de variância e independência. Na Tabela 3, tem-se os respectivos valores-p dos testes de normalidade (Shapiro -Wilk), de homoscedasticidade (Breusch-Pagan) e independência (Durbin-Watson).

Table 3: p-values from the Shapiro-Wilk (SW), Breusch-Pagan (BP) and Durbin-Watson (DW) tests.

Densidade	SW	BP	DW
75	0,028*	0,069(ns)	0,410(ns)
145	0,002**	0,120(ns)	0,890(ns)
215	0,024*	0,094(ns)	0,460(ns)
285	0,230 (ns)	0,093 (ns)	0,850(ns)
355	0,697 (ns)	0,072(ns)	0,790(ns)

Source: from authors (2024).

Legend: (ns) not significant, (*) and (**) significant by the test performed.

As densidade 75, 145, 215 apresentaram problema com a falta de normalidade residual, pois apresentaram valores-p $> 0,05$ no teste de Shapiro-Wilk. Não foi identificado nenhum problema com homogeneidade de variância ou independência, conforme observado nos testes de Breusch-Pagan (valor-p $> 0,05$) e Durbin-Watson (valor-p $> 0,05$), respectivamente.

Assim, foram ajustados modelos não lineares Gompertz com erros normais, assimétricos normais (SN) e assimétricos t (ST) utilizando 4 graus de liberdade, valor baseado na estimativa da curtose do acúmulo de matéria seca total para cada densidade de semeadura. As estimativas e critérios de qualidade de ajuste para os parâmetros dos modelo Gompertz para a distribuição normal(N), normal assimétrica (SN), t assimétrica (ST) podem ser vista na Tabela 4.

Com base na estimativa do parâmetro α conforme a Tabela 4, observa-se que o aumento da densidade está associado a um aumento no acúmulo assintótico.

De acordo com Lima *et al.* (2019), as estimativas dos parâmetros indicaram que, na maioria dos casos, o aumento da densidade resultou em um acúmulo gradativo de biomassa ao longo do ciclo de vida do feijoeiro para a cultivar Jalo. Além disso, foi observado um incremento gradual na matéria seca, com um comportamento semelhante para a cultivar Bolinha. Assim, os resultados observados estão em concordância com os encontrados por Vieira *et al.* (2008)

Nas densidade mais baixas como 75 e 145 mil planta ha⁻¹ analisa-se que o acúmulo da matéria seca é menor considerando as demais densidade. Além disso, houve uma superestimação na estimativa do acúmulo máximo de matéria seca ao considerar a densidade de 145 plantas por hectare.

Table 4: Estimates and quality of fit criteria for the parameters of the Gompertz model for the normal distribution (N), asymmetric normal (SN), asymmetric t (ST).

Dist-Dens	Parâmetros				
	α	β	k	λ	AIC
N-75	748,782	38,387	0,086	0	338,228
N-145	1382,104	51,692	0,032	0	347,762
N-215	9834,875	122,059	0,014	0	353,993
N-285	2210,495	32,515	0,066	0	379,877
N-355	3814,359	36,648	0,048	0	378,441
SN-75	694,277	29,856	0,055	2,965	335,985
SN-145	1382,086	51,691	0,032	0,009	347,762
SN-215	6333,197	113,886	0,012	4,702	348,859
SN-285	2952,570	35,610	0,022	16,081	371,878
SN-355	3738,305	34,479	0,042	2,229	377,421
ST ₍₄₎ -75	652,167	34,346	0,084	0,335	334,898
ST ₍₄₎ -145	38701,459	417,454	0,004	12,163	325,555
ST ₍₄₎ -215	4376,834	93,307	0,013	2,986	346,160
ST ₍₄₎ -285	2994,576	39,075	0,021	7,453	373,632
ST ₍₄₎ -355	3702,188	34,760	0,044	0,800	378,657

Source: from the authors (2024).

Com relação α pode-se observar que o acúmulo máximo da matéria seca variou entre 630 kg ha⁻¹ a 9850 kg ha⁻¹. Vale ressaltar que as densidade 215 e 285 apresentaram os maiores valores de acúmulo de matéria seca. Além disso, para as densidade de semeadura 75, 145, 215 e 355 mil plantas ha⁻¹ verificou-se uma redução na estimativa de α quando comparado as distribuições do erro.

Com relação ao ponto de abscissa associado ao ponto de inflexão pode-se avaliar que 29 a 77,5 DAE. Esse ponto ocorrer antes para o modelo Gompertz. Já a taxa de maturidade k teve pouca variação considerando as diferentes distribuições no dois modelos e variando de 0,01 a 0,10.

O parâmetro λ , associado à assimetria da distribuição, revelou que os modelos para cada densidade de semeadura exibiram valores de assimetria positiva, indicando que os resíduos dos modelos apresentaram uma leve assimetria à direita. Os valores variaram de 0 a 17, ressaltando que o valor zero indica uma distribuição simétrica.

Considerações Finais

O modelo Gompertz foi adequado para descrever o acúmulo de matéria seca nas cinco densidades de semeadura (75, 145, 215, 255, 285 e 355 mil plantas por hectare). Isso demonstra que este modelo pode ser utilizado para ajustar adequadamente o crescimento das plantas em diferentes condições de densidade, possibilitando a avaliação precisa do desenvolvimento das culturas.

Além disso, as distribuições normal assimétrica e t assimétrica conseguiram captar os valores atípicos encontrados e detectar a assimetria positiva nos dados. Este fato reforça a relevância de utilizar distribuições que melhor representem a natureza dos dados agrônômicos, especialmente em experimentos onde a variabilidade e os outliers são frequentes.

Foi observada uma superestimação em alguns dos parâmetros dos modelos avaliados, o que sugere que ajustes mais precisos são necessários para evitar estimativas exageradas que possam comprometer a interpretação dos resultados. No entanto, o uso das distribuições assimétricas normal e t contribuiu para mitigar esses efeitos, melhorando a precisão dos ajustes.

Com o ajuste do modelo, foi possível verificar que as distribuições normal e t assimétrica apresentaram um comportamento igual ou superior ao da distribuição normal. Isso reforça a necessidade de conhecer bem o comportamento dos dados e, assim, aplicar distribuições mais apropriadas, especialmente em contextos de assimetria

O modelo Gompertz estimou o acúmulo máximo de matéria seca em 3814 kg ha⁻¹, 3738,305 kg ha⁻¹ e 3702,188 kg ha⁻¹ na maior densidade de semeadura (355 mil plantas por hectare) para as distribuições normal, normal assimétrica e t-assimétrica, respectivamente. Observou-se que as distribuições normal assimétrica e t-assimétrica resultaram em estimativas ligeiramente mais conservadoras em comparação à distribuição normal. Isso pode ser atribuído à capacidade dessas distribuições de capturar a assimetria presente nos dados, além de serem mais robustas à presença de outliers. Assim, a escolha da distribuição adequada mostrou-se fundamental para obter estimativas mais precisas do acúmulo de matéria seca.

Agradecimento

Este trabalho foi realizado com o apoio financeiro da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), sob o processo 2024/04547-9.

Referências

- AZZALINI, A.; CAPITANIO, A. Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Wiley, v. 61, n. 3, p. 579–602, aug 1999.
- BARTLETT, M. S. Properties of sufficiency and statistical tests. *Proc. R. Soc. Lond. A*, v. 160, n. 901, p. 268–282, 1937.
- DURBIN, J.; WATSON, G. S. Testing for serial correlation in least squares regression: I. *Biometrika*, v. 37, n. 3/4, p. 409–428, 1950.
- FERNANDES, Felipe Augusto *et al.* Ajuste de modelos de crescimento para frangos machos. *Sigmae*, v. 8, n. 2, p. 788-795, 2019.
- GOMES-PIMENTEL, F. *Curso de estatística experimental*. [S.l.: s.n.], 2009.
- HUBERT, M.; ROUSSEEUW, P. J.; AELST, S. V. High-breakdown robust multivariate methods. *Statistical Science, Institute of Mathematical Statistics*, v. 23, n. 1, p. 92–119, feb. 2008.
- LABRA, F. V. *et al.* Estimation and diagnostics for heteroscedastic nonlinear regression models based on scale mixtures of skew-normal distributions. [S.l.]: *Elsevier BV*, 2012. 2149–2165 p.
- DE LIMA, K. P. *et al.* Modelagem não linear da biomassa seca do feijoeiro cv. Jalo. *Sigmae*, v. 8, n. 2, p. 359-369, 2019.
- MAGHAMI, M. M.; BAHRAMI, M.; SAJADI, F. A. On bias reduction estimators of skew-normal and skew-t distributions. *Journal of Applied Statistics*, Informa UK Limited, p. 1–23, jan 2020.

- MYUNG, I. J. Tutorial on maximum likelihood estimation. *Journal of mathematical Psychology*, v. 47, n. 1, p. 90-100, 2003.
- SAWILOWSKY, S. S. Nonparametric tests of interaction in experimental design. *Review of Educational Research*, v. 60, n. 1, p. 91-126, 1990.
- SHAPIRO, S.; WILK, M. An analysis of variance test for normality. *Biometrika*, v. 52, n. 3, p. 591-611, 1965.
- STEEL, R. G.; TORRIE, J. H.; DICKEY, D. A. *Principles and procedures of statistics: A biological approach*. [S.l.: s.n.], 1997.
- VIEIRA, N. M. B. et al. Altura de planta e acúmulo de matéria seca do feijoeiro cvs. brs mg talismã e ouro negro em plantio direto e convencional. *Ciência e Agrotecnologia, SciELO Brasil*, v. 32, n. 6, p. 1687-1693, 2008.
- WOBBROCK, J. O. et al. The aligned rank transform for nonparametric factorial analyses using only anova procedures. *In: Proceedings of the SIGCHI conference on human factors in computing systems*. 2011. p. 143-146.