



KELLY PEREIRA DE LIMA

**MODELOS NÃO LINEARES COM ERRO ASSIMÉTRICO
PARA DESCRIÇÃO DO ACÚMULO DE MATÉRIA SECA DO
FEIJOEIRO CULTIVAR BOLINHA**

LAVRAS – MG

2020

KELLY PEREIRA DE LIMA

**MODELOS NÃO LINEARES COM ERRO ASSIMÉTRICO PARA DESCRIÇÃO DO
ACÚMULO DE MATÉRIA SECA DO FEIJOEIRO CULTIVAR BOLINHA**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária para a obtenção do título de Doutora.

Dr. Augusto Ramalho de Moraes

Orientador

LAVRAS – MG

2020

**Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca
Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).**

Lima, Kelly Pereira de.

Modelos não lineares com erro assimétrico para descrição do
acúmulo de matéria seca do feijoeiro cultivar bolinha / Kelly
Pereira de Lima. - 2020.

86 p.

Orientador(a): Augusto Ramalho de Moraes.

Tese (doutorado) - Universidade Federal de Lavras, 2020.
Bibliografia.

1. Assimetria. 2. Modelo de crescimento. 3. Feijão. I. Moraes,
Augusto Ramalho de. II. Título.

KELLY PEREIRA DE LIMA

**MODELOS NÃO LINEARES COM ERRO ASSIMÉTRICO PARA DESCRIÇÃO DO
ACÚMULO DE MATÉRIA SECA DO FEIJOEIRO CULTIVAR BOLINHA
NONLINEAR MODELS WITH ASYMMETRIC ERROR FOR DESCRIPTION OF
ACCUMULATION OF DRY MATTER OF THE BEAN CULTIVAR BOLINHA**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária para a obtenção do título de Doutora.

APROVADA em 10 de junho de 2020.

Elias Silva de Medeiros	UFGD
Paulo Henrique Sales Guimarães	UFLA
Tales Jesus Fernandes	UFLA
Joel Augusto Muniz	UFLA

Dr. Augusto Ramalho de Moraes
Orientador

**LAVRAS – MG
2020**

*Especialmente a Deus acima de tudo, pois sem ele nada é possível.
Aos meus pais Maria Lúcia Avelino e Ozarias Marque, pelo apoio incondicional durante toda
essa jornada de estudo.
À todos meus familiares e amigos que, de um modo especial, contribuíram forma direta para a
finalização desse trabalho.
DEDICO*

AGRADECIMENTOS

Agradeço primeiramente a Deus que me concedeu e permitiu o gozo da saúde e da capacidade de aprender.

Aos meus pais, Maria Lúcia e Ozarias, pelo amor incondicional, pela formação do meu caráter e pelo exemplo que sempre me deram de homem e mulher, amo vocês.

Aos meus familiares, em especial ao Tio Raimundo (In memoriam), pela compreensão e por aprender comigo o que é uma pós-graduação. Ao Tio Francisco e o grande mestre Adaílton (In memoriam) que sempre me guiaram e orientaram com suas sábias palavras e conselhos espirituais.

Ao professor Augusto Ramalho pela atenção, orientação e ensinamento transmitido durante a realização desse trabalho. Desde já tenho uma profunda gratidão por ter me aceitado como sua orientada.

Aos professores do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, pelo conhecimento compartilhado durante todo esse percurso. Ao professor Joel Muniz, por sempre apoiar, motivar e incentivar os alunos.

Ao grupo de estudo em Regressão Não Linear Aplicada (NLIN), pela ajuda no desenvolvimento de minha pesquisa. Em especial a Ariana, Lucivânia, Edílson, Felipe, Walleff, Mirian, Vitor, Édipo e ao professor Tales.

Ao meu grande irmão, que vou levar da UFLA para a vida, Elias Medeiros, que sempre me incentivou, aconselhou e apoiou.

Ao Felipe, ao qual ensinei “abrir o coração para Lavras”, agradeço demais nossas conversas do dia-a-dia, as nossas risadas e os estudos na UFLA.

Um agradecimento mais que especial, à minha querida amiga Raquel Araújo, pela ajuda incondicional na minha vida pessoal e acadêmica.

Felizmente nessa jornada de mestrado e doutorado, não estive sozinha em minha empreitada. Conte com auxílio de colegas de valor, alguns dos quais tenho hoje a honra de considerar grandes amigos. Menção especial para Tia Régia, Lílian Maria, Cris Nogueira, Janaína, Laura, Leilão, Laís, Carol Bicalho, Jackelya, Jaqueline, Felipe, Laerte, Lucivânia, Miriazona, Walleff, Marcela, Ariana, Haiany, Victor, Denize Palmito, Paula Santos, Denise Paiva, Hebert, Ernandes (Xizes), Ossani, Henrique, Cristian, Nádia. Muito obrigada por me fazer sentir em casa, tenho uma GRATIDÃO ETERNA. Desculpa se esqueci de alguém, mas agradeço cada pessoa que contribuiu, de forma direta e indireta, na construção desta tese.

À Universidade Federal de Lavras, por me proporcionar a oportunidade de realizar a pós-graduação e por disponibilizar um ensino de qualidade e gratuito.

Aos funcionários do DEX e DES, em especial uma secretária mais que especial, atualmente uma grande amiga Nádia, por sempre nos receber de forma carinhosa com um sorriso contagiante. Além disso, prontamente disposta a resolver os grandes e pequenos problemas de cunho acadêmicos e pessoais.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001 e Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

*Feliz aquele que transfere o que sabe e aprende o que ensina.
(Cora Coralina)*

RESUMO

O estudo sobre o crescimento do feijoeiro através do acúmulo de matéria seca ao longo do tempo é de suma importância para a cultura, pois se torna uma ferramenta informativa no auxílio do manejo adequado e na detecção de fatores que afetam o desenvolvimento da planta. Sabe-se que a modelagem desse tipo de crescimento pode subsidiar um melhor conhecimento e aplicação eficiente das práticas agrícolas. Entretanto, o feijão tem um crescimento bem atípico e assim pode gerar valores que algumas distribuições estatísticas simétricas não conseguem captar, como por exemplo a distribuição normal. Sendo assim, busca-se alternativas para a incorporação desses tipos de valores atípicos, entre elas pode-se citar as distribuições assimétricas tais como: a normal e a t. Desta maneira, este trabalho teve como objetivo comparar os modelos não lineares Logístico e Gompertz diferentes distribuições de probabilidade para os erros incorporando assimetria como normal, normal assimétrico e t assimétrico para a descrição do acúmulo de matéria seca do feijoeiro cultivar “Bolinha”. Além disso, apresenta uma interface gráfica para tal análise. O experimento foi realizado na Universidade Federal de Lavras, no período da safra das águas nos anos de 2006/2007, em um delineamento em blocos causalizados, contendo três repetições; usou-se o esquema fatorial 5 x 8, sendo constituído de cinco densidades de semeadura (75, 145, 215, 285 e 355 mil plantas por hectare), 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 Logístico e Gompertz se mostraram adequados para descrição do acúmulo de massa seca. O modelo Logístico com erros normais assimétricos e erros t assimétricos obtiveram melhor resultado. O modelo Logístico com erros normais assimétricos destaca-se as densidades 145, 285, 355 mil plantas por hectare e para erros t assimétricos as densidades 75 e 215 mil plantas por hectare.

Palavras-chave: Assimetria. Modelo de crescimento. Feijão.

ABSTRACT

The study on the growth of common beans through the accumulation of dry matter along the time is of paramount importance for culture, as it becomes an informative tool in assistance in the proper management and in the detection of factors that affect the development of the plant. In this way, this study aimed to compare the nonlinear Logistic and Gompertz different probability distributions for errors incorporating asymmetry as normal, normal asymmetric and asymmetric t for the description of dry matter accumulation of common bean cultivar Bolinha. In addition, it presents an interface graphical parametric analysis. The experiment was carried out at the Federal University of Lavras, at period of the water harvest in 2006/2007, in a causalized block design, containing three repetitions; formation the factorial scheme 5 x 8, being constituted of five densities sowing activities (75, 145, 215, 285 and 355 thousand plants per hectare), eight evaluation periods (13, 23, 33, 43, 53, 63, 73, 83 days after emergence) and dry matter accumulation was analyzed not total. As a result of the research of nonlinear models Logística and Gompertz proved adequate to describe the dry mass accumulation. The Logistic model with errors asymmetric normals and asymmetric t errors obtained better results. The Logistic Model with normal asymmetric errors, the densities 145, 285, 355 thousand plants per hectare stand out and for asymmetric t errors the densities 75 and 215 thousand plants per hectare.

Keywords: Skew. Growth model. Bean.

LISTA DE FIGURAS

Figura 2.1 – Função de densidade da normal assimétrica variando o parâmetro de assimetria (λ) (a) assimetria negativa e (b) assimetria positiva.	25
Figura 2.2 – Função de densidade da normal assimétrica com $\sigma^2 = 1$, $\lambda = -8$ e μ variando ((a) valor negativo de μ e (b) valor positivo de μ .)	27
Figura 2.3 – Função de densidade da normal assimétrica com $\mu = 1$, $\lambda = -8$ e σ^2 variando.	27
Figura 2.4 – Função de densidade da normal assimétrica com $\mu = 1$, $\sigma^2 = 2$ e λ variando em que (a) valor negativo de λ e (b) valores positivos de λ	28
Figura 2.5 – Função de densidade da t assimétrica variando λ para $\nu = 1$ (a) e $\nu = 4$ (b).	30
Figura 2.6 – Função de densidade da t assimétrica variando λ para $\nu = 10$ (c) e $\nu = 30$ (d).	31
Figura 2.7 – Função de densidade da t assimétrica variando μ para $\sigma^2 = 1$, $\nu = 1$ e $\lambda = 1$, em que (a) valores negativos de μ e (b) valores positivos para μ	32
Figura 2.8 – Função de densidade da t assimétrica variando σ^2 para $\mu = 1$, $\nu = 3$ e $\lambda = 2$	33
Figura 2.9 – Função de densidade da t assimétrica variando λ ((a-esquerda)- valores negativos e (a-direita) valores positivos) para $\sigma^2 = 4$, $\nu = 4$ e $\mu = 1$	33
Figura 2.10 – Função de densidade da t assimétrica variando ν para $\sigma^2 = 5$, $\lambda = 2$ e $\mu = 1$	34
Figura 2.11 – Relações das distribuições	34
Figura 3.1 – Fluxograma da metodologia adotada nesse estudo	51
Figura 4.1 – Histograma para o acúmulo da matéria seca no total do feijoeiro Bolinha	54
Figura 4.2 – Boxplot para o acúmulo de matéria seca no Total do feijoeiro cv Bolinha	55
Figura 4.3 – Representação gráfica dos valores médios do acúmulo de matéria seca no Total do feijoeiro cv Bolinha da interação densidade de semeadura x dias após emergência	55
Figura 4.4 – Boxplot para o acúmulo de matéria seca para cada densidade de semeadura do feijoeiro cv Bolinha	57
Figura 4.5 – Representação gráfica do acúmulo de matéria seca e ajuste do modelo Logístico (a) e Gompertz (b) com erros normais, normais assimétricos (SN) e t assimétricos em função dos dias após a emergência (DAE) para o acúmulo de matéria seca no total cv. Bolinha na densidade de semeadura de 75 mil.	65

Figura 4.6 – Representação gráfica do acúmulo de matéria seca e ajuste do modelo Logístico (a) e Gompertz (b) com erros normais, normais assimétricos (SN) e t assimétricos em função dos dias após a emergência (DAE) para o acúmulo de matéria seca no total cv. Bolinha na densidade de semeadura de 145 mil.	66
Figura 4.7 – Representação gráfica do acúmulo de matéria seca e ajuste do modelo Logístico (a) e Gompertz (b) com erros normais, normais assimétricos (SN) e t assimétricos em função dos dias após a emergência (DAE) para o acúmulo de matéria seca no total cv. Bolinha na densidade de semeadura de 215 mil.	66
Figura 4.8 – Representação gráfica do acúmulo de matéria seca e ajuste do modelo Logístico (a) e Gompertz (b) com erros normais, normais assimétricos (SN) e t assimétricos em função dos dias após a emergência (DAE) para o acúmulo de matéria seca no total cv. Bolinha na densidade de semeadura de 285 mil.	67
Figura 4.9 – Representação gráfica do acúmulo de matéria seca e ajuste do modelo Logístico (a) e Gompertz (b) com erros normais, normais assimétricos (SN) e t assimétricos em função dos dias após a emergência (DAE) para o acúmulo de matéria seca no total cv. Bolinha na densidade de semeadura de 355 mil.	67
Figura 1 – Resultado para análise com distribuição normal	85
Figura 2 – Resultado para análise com distribuição normal considerando os critérios de comparação	86

LISTA DE TABELAS

Tabela 2.1 – Termo de penalidade c_n de diferentes critérios de avaliação de modelos baseados em informações: AIC, BIC e EDC	46
Tabela 3.1 – Análise química de amostras de solo coletadas na profundidade de 0 a 20 cm do latossolo vermelho distráfico, antes do plantio.*	47
Tabela 4.1 – Estatística descritiva para o acúmulo da matéria seca no total do feijoeiro Bolinha ($kgha^{-1}$)	53
Tabela 4.2 – Análise de variância contendo as fontes de variação (FV), números de graus de liberdade (GL), valores de quadrados médios e respectivas significância do teste F, para matéria seca no Total, no feijoeiro cv. bolinha.	56
Tabela 4.3 – Estimativas e respectivos valor-p dos testes de normalidade (Shapiro -Wilk), homogeneidade de variância (Bartlett) e independência (Durbin-Watson) para matéria seca no Total do feijoeiro cv. Bolinha.	56
Tabela 4.4 – Análise de variância com transformação de postos contendo as fontes de variação (FV), números de graus de liberdade (GL), valores de F e respectivas significância do teste F, para matéria seca no Total, no feijoeiro cv. Bolinha.	57
Tabela 4.5 – Estatística do teste (G) e valor-p do teste de Grubb para o acúmulo de matéria seca no Total em cada densidade de semeadura	58
Tabela 4.6 – Estimativas e respectivos valor-p dos testes de normalidade (Shapiro -Wilk - W), de homoscedasticidade (Breusch-Pagan - BP) e independência (Durbin-Watson- DW) para os modelos Logístico e Gompertz	60
Tabela 4.7 – Estimativas e critérios de seleção para o modelo Logístico (L-75) e Gompertz (G-75) considerando erros normais, normal assimétrico (SN) e t assimétrico (ST) ajustados para o acúmulo de matéria seca no Total do feijoeiro cv. Bolinha na densidade de semeadura de 75 mil plantas.	61
Tabela 4.8 – Estimativas e critérios de seleção para o modelo Logístico (L-145) e Gompertz (G-145) considerando erros normais, normal assimétrico (SN) e t assimétrico (ST) ajustados para o acúmulo de matéria seca no Total do feijoeiro cv. Bolinha na densidade de semeadura de 145 mil plantas.	61

Tabela 4.9 – Estimativas e critérios de seleção para o modelo Logístico (L-215) e Gompertz (G-215) considerando erros normais, normal assimétrico (SN) e t assimétrico (ST) ajustados para o acúmulo de matéria seca no Total do feijoeiro cv. Bolinha na densidade de semeadura de 215 mil plantas.	62
Tabela 4.10 – Estimativas e critérios de seleção para o modelo Logístico (L-285) e Gompertz (G-285) considerando erros normais, normal assimétrico (SN) e t assimétrico (ST) ajustados para o acúmulo de matéria seca no Total do feijoeiro cv. Bolinha na densidade de semeadura de 285 mil plantas.	62
Tabela 4.11 – Estimativas e critérios de seleção para o modelo Logístico (L-355) e Gompertz (G-355) considerando erros normais, normal assimétrico (SN) e t assimétrico (ST) ajustados para o acúmulo de matéria seca no Total do feijoeiro cv. Bolinha na densidade de semeadura de 355 mil plantas.	63

SUMÁRIO

1	INTRODUÇÃO	13
2	REFERENCIAL TEÓRICO	16
2.1	Cultura do feijão	16
2.2	<i>Outliers</i>	18
2.3	Análise de variância com transformação de posto alinhado (ART)	20
2.4	Distribuições assimétricas	23
2.4.1	Distribuição normal assimétrica padrão	24
2.4.2	Distribuição normal assimétrica de posição e escala	26
2.4.3	Distribuição <i>t</i> assimétrica padrão	28
2.4.4	Distribuição <i>t</i> assimétrica posição e escala	32
2.5	Modelos não lineares (MNL)	34
2.6	Modelo não linear normal assimétrico	36
2.7	Modelo não linear <i>t</i> assimétrico	37
2.8	Considerações sobre modelos não lineares aplicados a cultura do feijão	37
2.9	Modelo Logístico	39
2.10	Modelo Gompertz	39
2.11	Métodos de estimação	40
2.11.1	Método iterativo Gauss-Newton	42
2.12	Estimação por máxima verosimilhança	43
2.13	Seleção de modelos	45
2.13.1	Critério de informação de Akaike	45
2.13.2	Critério de informação Bayesiano	45
2.13.3	Critério de determinação eficiente	46
3	MATERIAL E MÉTODOS	47
3.1	Análise de regressão	49
3.1.1	Interface gráfica	51
4	RESULTADOS E DISCUSSÕES	53
4.1	Análise descritiva	53
4.2	Análise de variância com transformação de postos alinhados (ART)	56
4.3	Teste para outliers para cada densidade de semeadura	57
4.4	Ajuste dos modelos não lineares	59

5	CONCLUSÃO	69
	REFERÊNCIAS	70
	APENDICE A – Interface Gráfica	77

1 INTRODUÇÃO

Os modelos não lineares são usados em diversas pesquisas pois apresentam uma estrutura que permitem aplicações nas mais diversas áreas do conhecimento científico como, por exemplo, agricultura, biologia, ciência médica, econômica e entre outras. Na análise de curva de crescimento, os modelos não lineares são utilizados para descrever o tipo de crescimento, pois, geralmente apresentam interpretações práticas para os parâmetros de interesse do pesquisador.

Na literatura, existe a pressuposição que as observações sejam normalmente distribuídas, ou seja, a inferência nos modelos não lineares considera a normalidade dos resíduos. Entretanto, em diversas situações práticas a distribuição Normal não é a mais adequada para modelar os resíduos ou variável resposta, a inferência sob normalidade não é possível, isso pode ocorrer, por exemplo, quando os dados são provenientes de uma distribuição com cauda mais ou menos pesada que a distribuição normal, como é o caso da assimétrica (LABRA et al., 2012). Sendo assim, alguns modelos alternativos vêm surgindo para preservar a estrutura simétrica e permitir a redução da influência de valores atípicos.

Atualmente, tem-se observado que o método mais aplicado para alcançar a normalidade é a transformação de dados. Embora este procedimento tenha sido aplicado com sucesso em alguns casos, nem sempre é conveniente a aplicação deste procedimento. Segundo Azzalini e Capitanio (1999), a transformação de dados pode causar dificuldade na interpretação dos parâmetros do modelo, não ser aplicável em um determinado conjunto de dados e quando existe problema de homogeneidade pode ser que a transformação seja diferente para a alcançar a normalidade.

Caso a distribuição das observações seja caracterizada por uma distribuição simétrica, seja ela univariada ou multivariada, tem-se que a classe das distribuições elípticas pode ser um modelo alternativo para a transformação de variáveis. Essa classe de distribuição, surge como uma extensão das distribuições normais, tais como t-Student, exponencial potência, normal contaminada, logística I, logística II, slash, dentre outras (FANG; KOTZ; NG, 1990).

Apesar dos modelos estatísticos sob classe das distribuições elípticas representarem uma boa alternativa ao modelo com distribuição normal, a metodologia pode não ser adequada dado que em algumas situações a distribuição das observações é assimétrica. Além disso, na prática o interesse é estudar distribuições que sejam menos sensíveis que a distribuição normal a certos

desvios das suposições consideradas, ou seja, distribuições que possam acomodar assimetria e curtose (AZZALINI, 1985).

Quando a distribuição dos dados apresenta um comportamento assimétrico, esta por sua vez, pode ser modelada por meio da classe das distribuições assimétricas. A distribuição normal assimétrica pertence a essa classe, e por isso, talvez tenha sido uma das pioneiras para acomodação de assimetria (AZZALINI, 1985). Partindo do princípio da necessidade de acomodar assimetria presente nos dados, outras distribuições surgiram como Cauchy assimétrica (ARNOLD; BEAVER, 2000), t assimétrica (BRANCO; DEY, 2001), Exponencial potência assimétrica (DICICCIO; MONTI, 2004), Laplace assimétrica (ARYAL; NADARAJAH, 2005), Slash assimétrica (WANG; GENTON, 2006) e etc.

A classe das distribuições assimétricas está sendo utilizada nas mais diversas áreas do conhecimento, tanto na teoria como em aplicações. Vale salientar que, em relação a aplicação é importante a obtenção de resultados que não sejam influenciados por observações atípicas. Por isso, a distribuição t assimétrica mostra-se mais robusta que a distribuição normal assimétrica, ou seja, é uma alternativa para inferência robusta em vários tipos de modelos (MAGHAMI; BAHRAMI; SAJADI, 2020). Entre eles, pode-se citar os modelos não lineares.

É possível substituir a suposição de normalidade para o erro aleatório pela suposição de que a distribuição esteja em uma classe mais flexível de distribuições, sendo que esta classe pode ser formada pelas distribuições assimétricas, permitindo modelar possíveis presença de *outliers* e assimetria, simultaneamente (AZZALINI, 1985).

Como dito anteriormente, a análise de curva de crescimento usa os modelos lineares e não lineares com erro normal para estimar as causas do crescimento. Conforme Vieira et al. (2008), a análise de crescimento com a complementação da marcha de absorção de nutrientes é fundamental na avaliação dos efeitos de fatores de manejo sobre as plantas. Além disso, a análise permite conhecer as quantidades necessárias de nutrientes absorvidas e a absorção em cada fase da cultura, bem como, fornecer informações sobre as épocas mais adequadas para aplicação dos fertilizantes.

Na cultura do feijoeiro pouco destaque tem sido dado à modelagem do acúmulo de matéria com diferente tipo de distribuições para o erro, principalmente para essa cultura que apresenta valores atípicos ao longo do tempo, que pode estar relacionado aos diversos processos fisiológicos da planta. É importante salientar que essas observações atípicas (*outliers*) podem afetar diretamente na distribuição do erro, pois podem causar violação do pressuposto

de normalidade residual. Além disso, pode-se existir alguns *outliers* que afetam as estimativas do modelos (HUBERT; ROUSSEEUW; AELST, 2008).

Diante do exposto, este trabalho tem como objetivo propor diferentes estruturas não lineares com incorporação do erro normal, normal assimétrica e t assimétrica para a descrição do acúmulo de matéria seca do feijoeiro cv “Bolinha”. Adicionalmente, será desenvolvida uma interface gráfica para tal análise.

2 REFERENCIAL TEÓRICO

Este capítulo trata da revisão bibliográfica, contemplando as principais abordagens referentes ao feijão, *outliers*, distribuições assimétricas, aos modelos não lineares, aos modelos não lineares com erros assimétricos e a seleção de modelos, aos quais servirão de subsídios para a discussão dos capítulos subsequentes, bem como da metodologia empregada no experimento.

2.1 Cultura do feijão

O feijão comum (*Phaseolus vulgaris*) é uma leguminosa típica da alimentação dos brasileiros, sendo reconhecida por ser excelente fonte de proteínas, ferro, sais minerais e carboidratos. Ademais, os compostos fenólicos com ação oxidante podem reduzir a incidência de doenças cardiovasculares, alguns tipos de câncer, Mal de Alzheimer e Parkinson (BUTTERFIELD et al., 2002). Além da função relevante na alimentação como citado, o feijão é considerado um dos produtos agrícolas de maior importância no contexto econômico-social, principalmente, devido à mão de obra empregada durante o ciclo da cultura. Por isso, ocupa um lugar de destaque na agricultura brasileira, sendo caracterizado como um forte produto no mercado interno.

No Brasil, destaca-se como uma das principais culturas produzidas por apresentar características técnicas, agronômicas e culturais que o credenciam como uma excelente alternativa de exploração agrícola para pequenos produtores. É crescente, ainda, o cultivo por médios e grandes produtores. O Brasil é o maior produtor e consumidor mundial de feijão com produção média anual de 3,3 milhões de toneladas, com consumo per capita de aproximadamente 17,4 $kg.ano^{-1}$ (MAPA, 2020).

De acordo com CONAB (2020), o consumo nacional tem variado nos anos de 2010 a 2015, entre 3,3 e 3,6, porém na safra 2015/16 ocorreu uma redução para 2,8, sendo a menor historicamente. Esse fato se deve ao aumento do preço provocado para diminuição da área plantada e principalmente por condições climáticas que foram adversas nesse período. Em 2017/18, houve uma leve recuperação do consumo que foi para 3,3 milhões. No entanto, em 2018/19, a expressiva queda dos preços no varejo fez que o consumo recuasse para 3,1 milhões.

A produção do feijoeiro acontece em três safras, conhecidas como: safras das águas, da seca e de inverno. Em que, a primeira corresponde a quadra de agosto à dezembro e concentra altos índices pluviométricos, o plantio ocorre nesse período na região Centro-Sul e no Nordeste no período de outubro à fevereiro. A segunda, ocorre no período de menor índice de chuva no país, em que o plantio coincide nos meses de dezembro à março, tanto na região Centro-Sul

quanto no Nordeste. A terceira é a colheita do feijão irrigado, ocorre no período de abril a junho na Região Centro-Sul do país (MAPA, 2020).

No Brasil, os tipos de feijão mais produzidos são: Carioquinha, Preto, Jalo, Rajado, Roxo e Amarelo. O cultivo do feijão Amarelo, também conhecido como “Bolinha”, é uma cultivar alternativa ademais citada anteriormente, pois se destaca no mercado, já que essa tem preço expressivamente maior em relação as demais cultivares consumidas no Brasil.

Embora o cultivo do feijão e das grandes variedades existentes seja realizado em todo território nacional, a produção de grão no país é ainda considerada baixa. Várias causas podem ser apontadas como responsáveis por essa situação de baixa produtividade, dentre elas tem-se problema de ordem técnica, aplicação racional de fertilizantes, a utilização inadequada das densidades de semeadura e épocas de colheita (TEIXEIRA et al., 2000).

O feijoeiro é considerado uma planta exigente em nutrientes, por causa do seu sistema radicular que é pequeno e pouco profundo, além disso, possui um ciclo curto, ou seja, é necessário um manejo adequado do solo para que o feijão consiga crescer e se desenvolver. A densidade populacional é um fator que está relacionado com a produtividade do feijoeiro e está diretamente relacionada com a fertilidade do solo, pois em certas condições de baixa fertilidade as plantas crescem menos, assim exige a necessidade do aumento da quantidade de plantas na semeadura (VIEIRA et al., 2008).

No caso de solos férteis deve-se optar por menor quantidade de plantas na semeadura, devido à grande magnitude de área foliar. Contudo, com o aumento da capacidade de compensação do feijoeiro, a densidade de semeadura não deve ser exorbitantemente alta, pois a competitividade intraespecífico por água e nutrientes pode levar à morte da planta (KERES et al., 2020).

O desenvolvimento da cultura também é afetado pela densidade populacional da planta que depende, entre outros fatores, da capacidade do solo em fornecer nutrientes. Em geral, são utilizadas populações de 165 a 300 mil plantas por hectare. No caso de feijoeiro de diferentes hábitos de crescimento, a influência da população de plantas tem mostrado que existe aumento de produtividade com o incremento na população (SILVA; LIMA; MENEZES, 2007; SOUZA; ANDRADE; ALVES, 2004) no caso de feijoeiro de diferentes hábitos de crescimento. Contudo, maiores populações nem sempre resultam em grandes produções, o que muitas vezes têm relação com a capacidade de compensação entre os componentes de produção do feijoeiro (VIEIRA et al., 2008; SOUZA et al., 2008).

De acordo com Alves et al. (2009), as densidades populacionais acima de 200 mil plantas por hectare podem ser consideradas ideais. Para a cultivar (cv) Bolinha, essa densidade está em torno de 240 mil plantas por hectares, que proporciona uma produtividade máxima de grãos.

A qualidade dos grãos é outro fator que afeta a produtividade do feijoeiro, e essa é determinada pela época adequada da colheita que envolve maior quantidade de mão de obra e recurso do agricultor. Então, deve-se efetuar a colheita quando as plantas atingirem a maturidade fisiológica (a maturação completa ocorre quando existe a perda da umidade e o máximo acúmulo de matéria seca da semente) (BEVILAQUA et al., 2013). Nesse ponto, ocorre o amarelecimento e queda das folhas, além da mudança na coloração das vagens e, no final desse processo de maturação, as plantas apresentam, em sua maioria, as raízes deterioradas, as hastes e as vagens secas.

Atrasos ou antecipações na colheita podem acarretar diversos prejuízos à qualidade das sementes. Quando ocorre de forma antecipada, existe um risco de redução na produtividade e uma maior dificuldade na triagem do produto, devido à umidade presente na vagem. Por outro lado, se for realizada tardiamente, pode ocorrer a redução da qualidade das sementes, já que o vigor da semente diminui à medida que se prolonga seu tempo de retenção no campo. Além disso, pode ocorrer a perda de sementes, provocada pela deiscência natural das vagens. Grande parte das cultivares de feijoeiro possuem um ciclo de 90 a 100 dias (ANDRADE et al., 2001).

A matéria seca do feijoeiro é um importante parâmetro de crescimento e está associado significativamente com a produtividade de grãos (FAGERIA; BALIGAR, 2005). Além disso, a produtividade da matéria seca está associada com a acumulação de nutrientes pela cultura, o que torna de suma importância, o conhecimento sobre o acúmulo de matéria seca durante o ciclo do cultivo.

Em estudos realizados com matéria seca, é possível observar a presença de *outliers*. Diante disso, torna-se necessário que estudos sejam desenvolvidos para detectar a presença de possíveis valores atípicos relacionados com a matéria seca do feijão.

2.2 Outliers

Os *outliers*, também conhecidos como valores discrepantes, aberrantes ou anormais, são observações que diferenciam marcadamente de todas as outras observações da amostra. Um *outlier* pode ser definido, também, como um valor que desvia-se da normalidade e que pode causar anomalias nos resultados obtidos. Em outras palavras, pode-se dizer que os *outliers*

são pontos fora da curva, pois são observações diferentes das usuais. De acordo com Barnett e Lewis (1974), as observações inconsistentes encontradas nos dados pode ter uma causa da variabilidade natural.

Conforme Barnett e Lewis (1974), pode-se classificar as causas de ocorrência de outliers durante o processo de amostragem das seguintes maneiras: (i) os *outliers* são elementos que pertencem à população, ou seja, variedade inerente à população, (ii) erros de medição, que podem acontecer durante o período de coleta dos dados e pode ser ocasionada, também, por erros humanos, tais como digitação de dados incorretos, e por erros de máquinas.

Já Stevens (1984), distingue dois tipos de outliers: pontos influentes, que ocorrem na variável explicativa e afetam o modelo de regressão, mas esses pontos não necessariamente implicam na produção de *outliers*; e pontos de alavanca, que são valores extremos na variável resposta e que podem dar o indicativo de falhas no modelo. Um ponto de alavanca influencia na estrutura linear descrita pelos dados e influencia na estimação dos coeficiente de regressão. Portanto, uma observação pode ser classificada como valor atípico, ponto de alavanca, ou os dois.

Assim, os *outliers* necessitam de estudo especial, pois geralmente resultam em alguma violação dos pressupostos necessários para adequabilidade do modelo e, conseqüentemente, produzindo efeitos não confiáveis na eficiência dos estimadores (TUKEY, 1960).

A literatura apresenta várias métricas que podem ser usadas para caracterizar pontos como *outliers*. Existem alguns trabalhos escritos para a identificação de outliers em regressão linear, como por exemplo em (ANSCOMBE; TUKEY, 1963; COOK; WEISBERG, 1982; FOX, 1991). Em relação a identificação de *outliers* em modelo não linear, por exemplo, tem-se os trabalhos de Fox, Hinkley e Larntz (1980) que é baseado no método de mínimos quadrados ordinários. Porém, Riazoshams et al. (2009) relata que este tema foi pouco explorado na literatura.

Além disso, Riazoshams et al. (2009) relatam que medidas baseadas nas estimativas de mínimos quadrados ordinários não são eficientes, pois podem causar dois tipos de erros de classificação: o *masking* (falso negativo é quando o resultado indica ausência da condição que está sendo testada quando na realidade ela está presente), no qual um *outlier* não é detectado e o *swamping* (falso positivo é quando o resultado indica a presença da condição que está sendo testada quando na realidade ela não existe), em que um *outlier* não é classificado como *outlier*.

Os *outliers* podem gerar violações dos pressupostos quando consideramos uma análise de variância. Desta forma, para contornar esse problema, utiliza-se um método não paramétrico, tal como, a análise de variância com transformação de posto alinhado (Aligned Ranks Transformation - ART).

2.3 Análise de variância com transformação de posto alinhado (ART)

Um delineamento em esquema fatorial é usado quando um experimento envolve dois ou mais fatores simultaneamente, sendo que cada fator possui dois ou mais níveis. Considerando os planejamentos fatoriais, a sua importância está em permitir uma análise simultânea dos efeitos de fatores principais e suas interações. Assim, todas as combinações possíveis dos níveis dos fatores são investigados em cada repetição de um experimento.

Um efeito principal desse tipo de esquema é definido como uma mudança no nível de um fator, mantendo os fatores restantes em um nível fixo. Já a interação existe entre dois fatores, se a diferença na resposta entre os níveis de um fator não for a mesma em todos os níveis dos outros fatores.

Portanto, o foco da análise de experimento fatorial é testar a hipótese sobre os efeitos principais e interação. Logo, se a interação existir (significativa), os efeitos principais existem conjuntamente e uma influência no outro. No caso da não existência do efeito de interação, os efeitos principais atuam de forma independente, ou seja, um teste para os efeitos principais deve ser realizado. Para realizar tal análise de comparação de tratamentos usa-se a análise de variância, na qual, exigem os pressupostos como: normalidade, homogeneidade e independência dos resíduos (GOMES-PIMENTEL, 2009). Porém, quando os pressupostos não são atendidos, deve-se pesquisar formas alternativas para realização da análise.

Uma alternativa para que o pressuposto de normalidade seja atendido, é utilizar a transformação nos dados, ou então, considerar um outro procedimento no qual a suposição de normalidade não seja necessária. A primeira abordagem, consiste em aplicar uma transformação na variável resposta do tipo logarítmica, raiz quadrada, arco seno e entre outras. O segundo, inclui uma grande quantidade de métodos baseados em atribuições de postos de dados. O procedimento com atribuições de postos fornece um método alternativo robusto e poderoso de análise de dados para os métodos clássicos de mínimos quadrados (CONOVER; IMAN, 1976).

Quando a validade dos testes para a ANOVA são comprometidas, uma opção para obtenção dos resultados é utilizar um teste não paramétrico, pois a violação de pressuposições

geram algumas consequências na confiabilidade dos testes paramétricos, tais como a análise de variância, análise de regressão e comparação de tratamentos ficam afetados, pois pode ocorrer uma certa alteração na probabilidade de ocorrência dos erros tipo I e II e, conseqüentemente, causar falsas interpretações a respeito dos efeitos de tratamentos (MARTIN; STORCK, 2008; STORCK et al., 2000; STEEL; TORRIE; DICKEY, 1997).

Quando várias violações de suposições ocorrem, o poder e o erro do tipo I de um teste paramétrico é frequentemente reduzido e, portanto, requer uma mudança para opções não paramétricas (SAWILOWSKY, 1990).

Leys e Schumann (2010) apresentaram um método não paramétrico para analisar as interações usando o teste transformação de posto alinhado (ART). Além disso, os autores citam a facilidade de condução e robustez desse teste, pois não depende da distribuição das variáveis e baseia-se em uma distribuição F. O ART apresenta-se então, como uma técnica robusta e poderosa, pois não é sensível a *outliers* e violações de suposições sobre a distribuição dos erros.

O ART é baseado no teste de transformação de posto (RT) que foi introduzido por Conover e Iman (1976). O princípio de RT atribui um posto à variável dependente antes de aplicar o teste F paramétrico. Embora essa abordagem produza bons resultados para casos simples, nas situações que envolvem possíveis interações, o método apresenta falhas potenciais. Estudos de simulação apontaram que para grandes tamanhos de efeito, o método possivelmente causaria altas taxas de erro para testes de interação (BLAIR; SAWILOWSKY; HIGGINS, 1987).

O método ART consiste em ajustar as observações, antes de classificá-las. Mansouri (1998) analisou as propriedades assintóticas do ART e apresentou procedimentos de comparações múltiplas. O teste tem poder semelhante ao do teste F para distribuição dos erros, que são aproximadamente normal, e pode ter um poder muito maior para certas distribuições dos erros não normal. No entanto, algumas alterações do valor nominal das taxas de erro tipo I foram observadas para distribuição dos erros assimétricos.

O procedimento ART alinha primeiro todas as N observações, subtraindo todos os efeitos que não são de interesse primário. Em seguida, classifica as N observações alinhadas de 1 a N e calcula o teste F a partir dos postos alinhados. Considerando que as observações são provenientes de um modelo de efeitos fixos balanceados com dois fatores, o modelo é dado por:

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + e_{ijk} \quad (i = 1, \dots, a; j = 1, \dots, b; k = 1, \dots, n)$$

em que y_{ijk} é a variável resposta, μ é a constante do modelo, α_i é o efeito do i -ésimo fator A, β_j é o efeito do j -ésimo fator B, $(\alpha\beta)_{ij}$ é o efeito de interação entre o fator A e fator B e e_{ijk} é erro experimental, em que $N = abn$.

O ajuste para o teste ART é realizado subtraindo estimativas dos mínimos quadrados dos efeitos principais das observações, isto é, a k -ésima resposta alinhada ao i -ésimo nível do fator A e o j -ésimo nível de B é:

$$y'_{ijk} = y_{ijk} - (\bar{y}_{i..} - \bar{y}_{...}) - (\bar{y}_{.j.} - \bar{y}_{...}) - \bar{y}_{...}$$

O modelo correspondente da respostas ajustadas é:

$$y'_{ijk} = (\alpha\beta)_{ijk} + e_{ijk}$$

em que $(\alpha\beta)_{ijk}$ é a média das respostas ajustadas ao efeito de interação.

O teste F de interação pode ser calculado a partir das respostas alinhadas. No caso, é baseado em:

$$QM(\text{interação})/QMerro \sim F_{[(a-1)(b-1), ab(n-1)]}$$

No teste ART, os postos naturais de 1 a n são atribuídos às respostas ajustadas e, posteriormente, a taxa de variação é calculada nos postos ajustados. Para testar um efeito, o teste F paramétrico é aplicado à variável ajustada correspondente, em que apenas esse efeito é examinado, ignorando os outros dois. Como o teste F é usado para testar os efeitos da ART, então surge a questão se a distribuição assintótica é a F. De acordo com Salter e Fawcett (1993), pelo menos para o ART, esses testes são válidos.

Verificando o efeito de interação significativo para realizar a análise quando os dados são quantitativos, usa-se a regressão. Sabendo-se que, possíveis presenças de *outliers* podem influenciar na variável resposta e no erro, tem-se as opções de usar métodos paramétrico ou não paramétrico.

Nesse estudo serão abordadas distribuições que acomodam problemas com valores atípicos, por isso, nas próximas seções serão apresentadas as distribuições assimétricas e os modelos não lineares com erro assimétrico. Um princípio que se adotou nesse trabalho foi de denotar algumas distribuições por sua abreviatura em inglês. Por exemplo, a distribuição normal assimétrica será referenciada como SN (skew-normal) e t assimétrica por ST (skew-t).

2.4 Distribuições assimétricas

A teoria clássica dos modelos lineares e não lineares assume que o erro aleatório possui distribuição normal, o que muitas vezes pode não ocorrer pela natureza dos dados. Desta forma, surge a necessidade de utilizar modelos alternativos que superem as possíveis “falhas” apresentadas pelo modelo com distribuição normal, como de não acomodar a assimetria, valores atípicos (outliers) e heterogeneidade de variância. Assim, existe a possibilidade de substituir a suposição de normalidade para o erro aleatório por uma distribuição que esteja em uma classe mais flexível de distribuições.

Na falta de normalidade dos erros, uma alternativa é considerar outras distribuições para os mesmos, como as distribuições assimétricas, que dão suporte tanto para assimetria quanto para as caudas pesadas. Sendo assim, escolhe-se as distribuições que estão na classe da família assimétrica, que por sua vez, consiste em uma extensão da distribuição normal, incorporando a possibilidade de assimetria (esta característica em conjunto permite modelar outliers e assimetria ao mesmo tempo).

De acordo com Cruz e Branco (2009), alguns trabalhos vêm sendo desenvolvidos para propor distribuições mais flexíveis para o erro, que podem retratar melhor os dados analisados, pois algumas distribuições assimétricas preservam as propriedades das distribuições que foram geradas. Como por exemplo, a distribuição t assimétrica e a normal assimétrica, ambas possuem as propriedades da distribuição t-Student e normal, respectivamente. Os autores analisaram modelos não lineares com enfoque bayesiano e erros com distribuições assimétricas em dados clínicos de gestantes, e concluíram que a distribuição t assimétrica para os erros possuiu o melhor ajuste para os dados. Desta maneira, alguns autores usaram as distribuições assimétricas para modelagem de dados com a presença de caudas mais pesadas ou leves que a distribuição normal.

Guedes et al. (2014) ajustaram modelos de regressão com erros normais e normais assimétricos a dados de alturas de plantas observados no tempo e cultivados em vasos de diferentes tamanhos e concluíram que o modelo normal assimétrico não é o mais adequado e sugerindo o modelo de regressão com erro t assimétrico.

Louzada, Ferreira e Diniz (2014) propuseram quatro modelos de crescimento (Morgan-Mercer-Flodin, von Bertalanffy, Gompertz e Richards) com erros normais e normais assimétricos, no qual os erros foram testados com três estruturas de covariâncias diferentes e mostraram

que o melhor modelo foi o modelo von Bertalanffy heterocedástico multiplicativo com erros normais assimétricos.

Já Chen, Zeng e Song (2014), analisaram um modelo não linear com erro t normal assimétrico, em que essa distribuição é proposta dentro de uma classe de distribuição flexível assimétrica-simétrica, e obtiveram que essa distribuição acomoda os valores atípicos e a multimodal.

Um estudo do modelo não linear com abordagem bayesiana foi aplicado em dados de peso de codornas machos e fêmeas por Rossi e Santos (2014), consideraram as distribuições assimétricas e simétricas para os erros, e verificaram que os modelos de Gompertz com erros normais assimétricos e t assimétricos são os que melhores se ajustam aos dados.

Ferreira e Lachos (2016) propuseram uma extensão de modelos não lineares usando distribuições normais de mistura com assimetria e escala. Concluíram que os modelos não lineares utilizando essa distribuição de caudas pesadas são mais robustos a captação de observações atípicas em comparação com a distribuição normal assimétrica. Como pôde-se observar, os estudos considerando o erro com distribuições assimétricas são frequentemente usados nos mais diversos enfoques.

2.4.1 Distribuição normal assimétrica padrão

A ideia de uma distribuição normal assimétrica foi introduzida por Azzalini (1985) para o caso univariado. Ao longo do tempo diversos trabalhos foram desenvolvidos como em Nelson (1964), Roberts (1966), Aigner, Lovell e Schmidt (1977) e O'hagan e Leonard (1976) no contexto bayesiano para modelagem estatística.

Uma extensão da normal assimétrica ao caso multivariado é apresentada por Azzalini e Valle (1996), enquanto Arellano-Valle, Genton e Loschi (2009) desenvolveram a normal assimétrica multivariada usando a abordagem bayesiana, apresentando proposições que podem ser utilizadas nas distribuições de predição e posteriori do vetor de observações e parâmetro de assimetria quando os parâmetros de localização e escala são conhecidos. Já Wang, Li e Gupta (2009), utilizaram a normal assimétrica multivariada para definir a distribuição qui-quadrado assimétrica não-central generalizada. Arellano-Valle, Branco e Genton (2006) desenvolveram uma estrutura unificada para distribuições multivariadas assimétricas e estabeleceram ligações com as definições já existentes dessas distribuições, e Genton (2004) apresenta diversas aplicações da distribuição normal assimétrica.

Segundo Azzalini (1985), Z tem distribuição normal assimétrica padrão com parâmetro de assimetria $\lambda \in \mathbb{R}$, e denota-se por $Z \sim SN(\lambda)$, se sua função densidade de probabilidade (f.d.p.) é dada por:

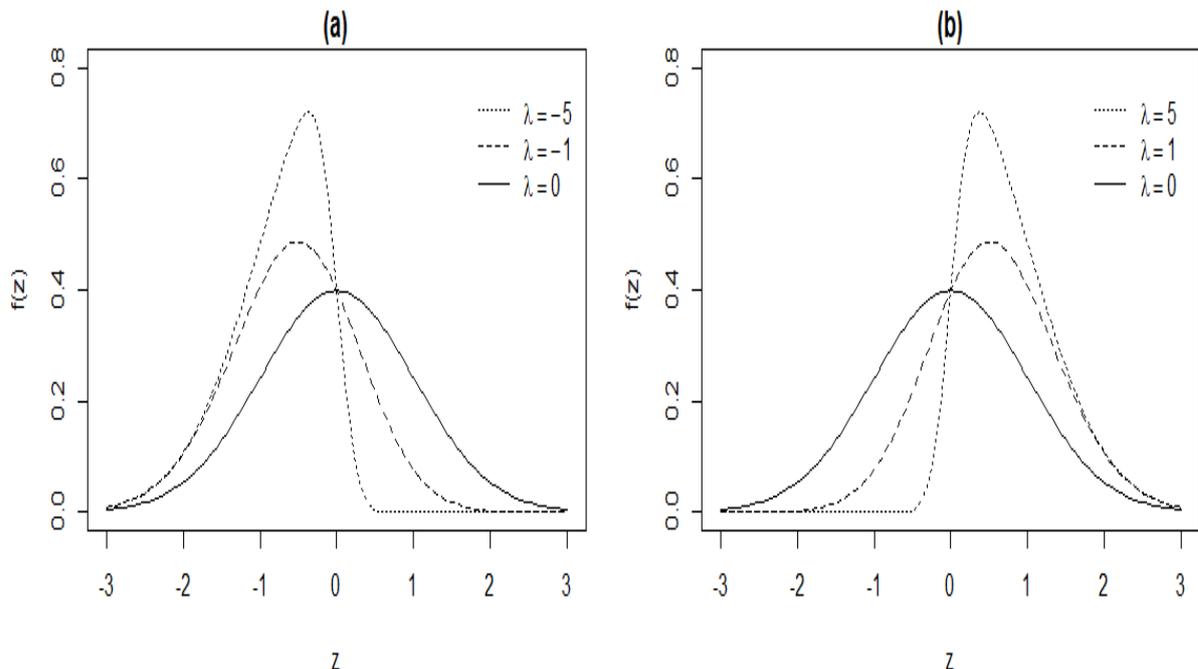
$$f_Z(z; \lambda) = 2\phi(z)\Phi(\lambda z)\mathbb{1}_{\mathbb{R}}(z) \quad (2.1)$$

em que as funções $\phi(\cdot)$ e $\Phi(\cdot)$ representam, respectivamente, a função de densidade e função acumulada da distribuição normal padrão. Já $\mathbb{1}_{\mathbb{R}}(\cdot)$ é função indicadora da distribuição em que

$$\mathbb{1}_{\mathbb{R}}(z) = \begin{cases} 1, & \text{se } z \in \mathbb{R} \\ 0, & \text{caso contrário.} \end{cases}$$

Na Figura 2.1, observa-se o comportamento da função de densidade para diferentes valores do parâmetro λ que define assimetria. Sabe-se, que a distribuição normal padrão é um caso particular da distribuição normal assimétrica padrão quando $\lambda = 0$. Além disso, observa-se no gráfico que para valores positivos de λ tem-se assimetria à direita (Figura 2.1 (a)), para valores negativos de λ tem-se assimetria à esquerda (ver Figura 2.1 (b)) e para λ igual a zero tem-se uma distribuição simétrica.

Figura 2.1 – Função de densidade da normal assimétrica variando o parâmetro de assimetria (λ) (a) assimetria negativa e (b) assimetria positiva.



Azzalini e Capitanio (2014) mostraram algumas propriedades da distribuição normal assimétrica, como :

1. Se $\lambda = 0$ em (2.1) tem-se que $Z \sim N(0,1)$.
2. Se $Z \sim SN(\lambda)$, então $-Z \sim SN(-\lambda)$.
3. Se $Z \sim SN(\lambda)$, então $|Z|$ tem distribuição half-normal, isto é $|Z| \sim HN(0, 1)$
4. Se $Z \sim SN(\lambda)$, então $Z^2 \sim \chi_1^2$
5. Se $Z \sim SN(1)$ então $F_Z(z) = [\Phi(z)]^2$

2.4.2 Distribuição normal assimétrica de posição e escala

Conforme Azzalini (1985), Y tem distribuição normal assimétrica com parâmetros de posição $\mu \in \mathbb{R}$ e de escala $\sigma^2 > 0$, denotada por $Y \sim SN(\mu, \sigma^2, \lambda)$, se sua função densidade de probabilidade (f.d.p.) é dada por:

$$f_Y(y) = \frac{2}{\sigma} \phi\left(\frac{y-\mu}{\sigma}\right) \Phi\left(\lambda \frac{y-\mu}{\sigma}\right) \mathbb{1}_{\mathbb{R}}(y). \quad (2.2)$$

em que $\mathbb{1}(\cdot)$ é função indicadora da distribuição com $y \in \mathbb{R}$.

A distribuição normal assimétrica de posição e escala tem a seguinte propriedade: se $Z \sim SN(\lambda)$ e $Y = \mu + \sigma Z$, então $Y \sim SN(\mu, \sigma^2, \lambda)$.

Nas Figuras 2.2, 2.3 e 2.4, apresentam-se os gráficos da densidade da distribuição normal assimétrica fixando dois parâmetros e variando o outro. Na Figura 2.2, conclui-se que quando fixa-se os parâmetros de escala e assimetria e varia-se o de posição, os gráficos da densidade se deslocam no eixo horizontal tanto para valores positivo quanto negativos de μ .

Já na Figura 2.3, em que considera-se posição e assimetria fixos, a medida que σ^2 aumenta o gráfico da densidade vai ficando mais achatado. Assim, pode-se observar pelas Figuras 2.2 e 2.3, que existe uma ideia da variação da posição e da escala quando altera-se os valores para tais parâmetros da distribuição. Desta forma, consegue-se analisar o comportamento da distribuição da normal assimétrica para posição e escala.

Figura 2.2 – Função de densidade da normal assimétrica com $\sigma^2 = 1$, $\lambda = -8$ e μ variando ((a) valor negativo de μ e (b) valor positivo de μ .)

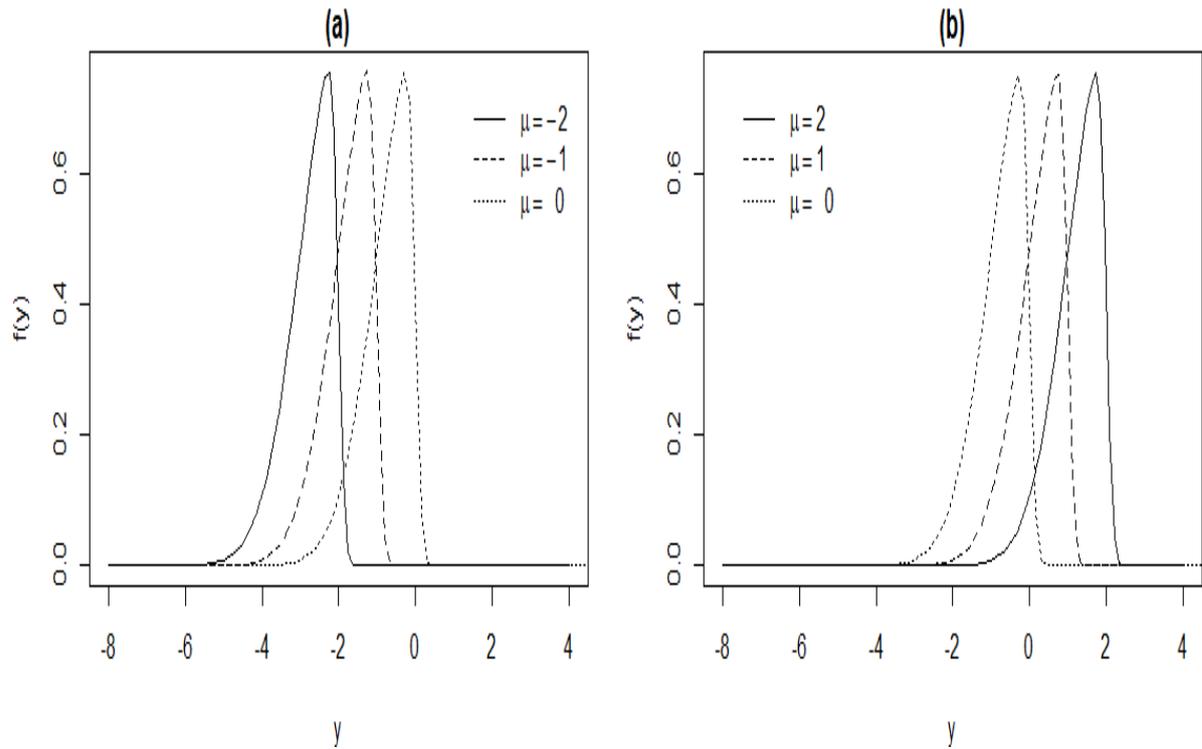
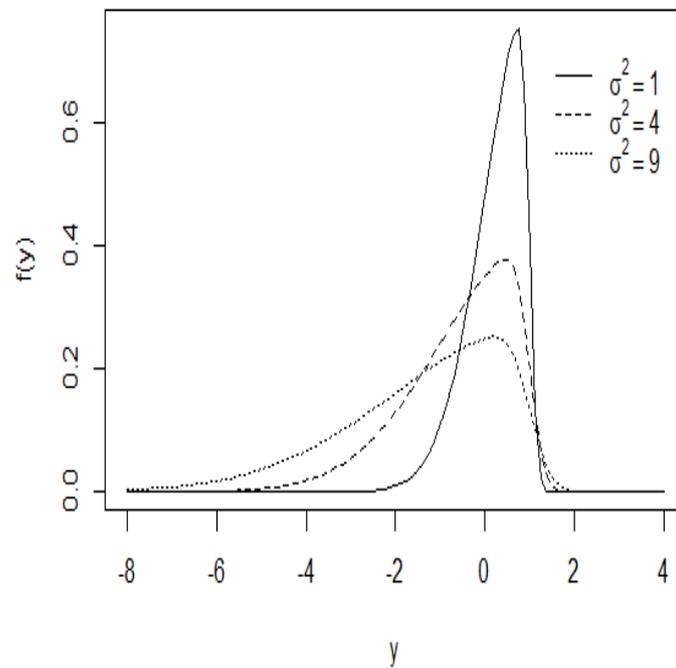
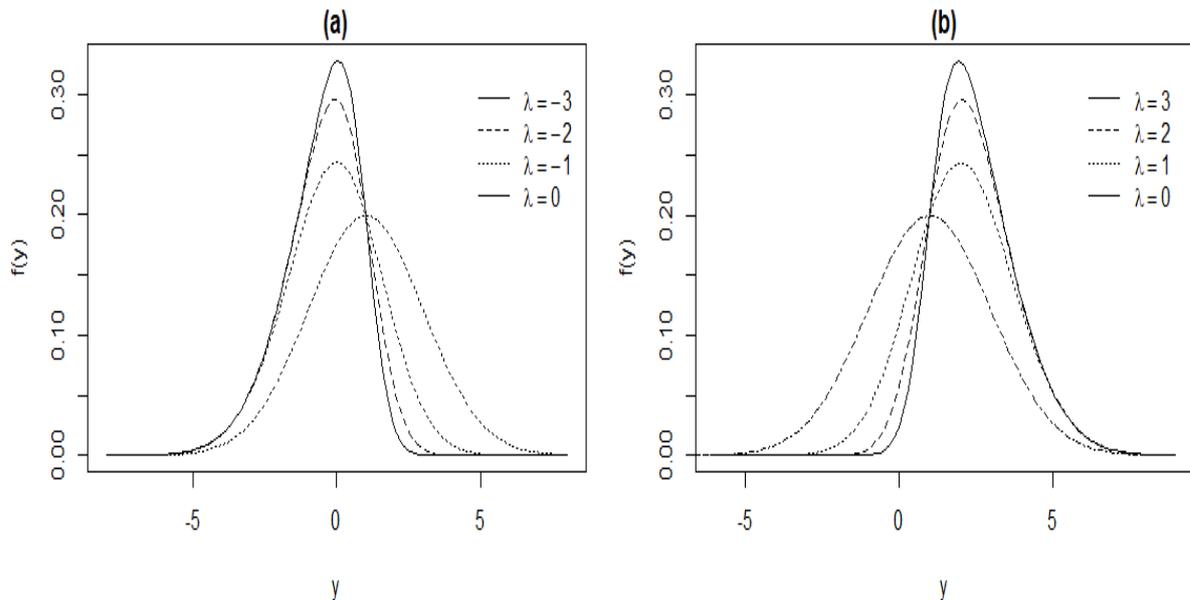


Figura 2.3 – Função de densidade da normal assimétrica com $\mu = 1$, $\lambda = -8$ e σ^2 variando.



Considerando que a posição e a escala são fixos, pode-se observar por meio da Figura 2.4, que na medida que λ vai aumentando, a densidade vai ficando assimétrica à direita, ou seja, para valores positivos. Por outro lado, quando λ assume valores negativos, tem-se assimetria à esquerda, como mostra a Figura 2.4(a). E para $\lambda = 0$, o comportamento gráfico é de uma distribuição simétrica que coincide com Normal com média μ e variância σ^2 .

Figura 2.4 – Função de densidade da normal assimétrica com $\mu = 1$, $\sigma^2 = 2$ e λ variando em que (a) valor negativo de λ e (b) valores positivos de λ .



2.4.3 Distribuição t assimétrica padrão

Conforme Branco e Dey (2001), a distribuição t assimétrica representa uma extensão da distribuição t de Student, e assim como, a distribuição normal assimétrica, existe o acréscimo de um parâmetro λ que regula o seu grau de assimetria, para maiores detalhes ver em (GODOI, 2007). Diz-se que uma variável aleatória Y segue uma distribuição t assimétrica padrão com ν graus de liberdade, denota-se de $Y \sim ST(\nu, \lambda)$, se sua função densidade de probabilidade é dada por:

$$f_Y(y) = 2t_{(\nu)}(y)T_{(\nu+1)} \left(\lambda y \sqrt{\frac{1+\nu}{\nu+y^2}} \right) \mathbb{I}_{\mathbb{R}}(y) \quad (2.3)$$

em que as funções $t_{(\nu)}(\cdot)$ e $T_{(\nu)}(\cdot)$ representam, respectivamente, a função de densidade e função acumulada da distribuição t de Student e $\mathbb{I}(\cdot)$ é função indicadora da distribuição com $y \in \mathbb{R}$.

Para valores positivos de λ , é possível observar que as distribuições possuem assimetria à direita. E para valores negativos de λ as distribuições possuem assimetria à esquerda.

Quando $\lambda = 0$, obtém-se uma distribuição simétrica que coincide com a distribuição t padrão. Sendo assim, pode-se verificar esses comportamentos da distribuição t assimétrica padrão através dos gráficos da sua função de densidade. A distribuição t assimétrica possui as seguintes propriedades:

1. $ST(v, 0) \stackrel{D}{=} t_{(v)}$, em que $\stackrel{D}{=}$ representa igual a distribuição.
2. Se $Z \sim SN(\lambda)$ e $W \sim Gama(\frac{v}{2}, \frac{v}{2})$, em que Z tem distribuição normal assimétrica e W possui distribuição gama são independentes, então:

$$Y = \frac{Z}{\sqrt{W}} \sim ST(v, \lambda).$$

Para maiores detalhes ver em [Azzalini e Capitanio \(2014\)](#).

As Figuras 2.5 e 2.6 apresentam os gráficos da função de densidade da distribuição t assimétrica padrão nos quais os valores de λ e v sofreram algumas variações. Quando $v = 1$ e $\lambda = 0$ tem-se um caso particular, a distribuição de Cauchy. Além disso, é possível verificar que com um aumento do valor de v a distribuição “perde” a característica de cauda mais pesada.

Figura 2.5 – Função de densidade da t assimétrica variando λ para $\nu = 1$ (a) e $\nu = 4$ (b).

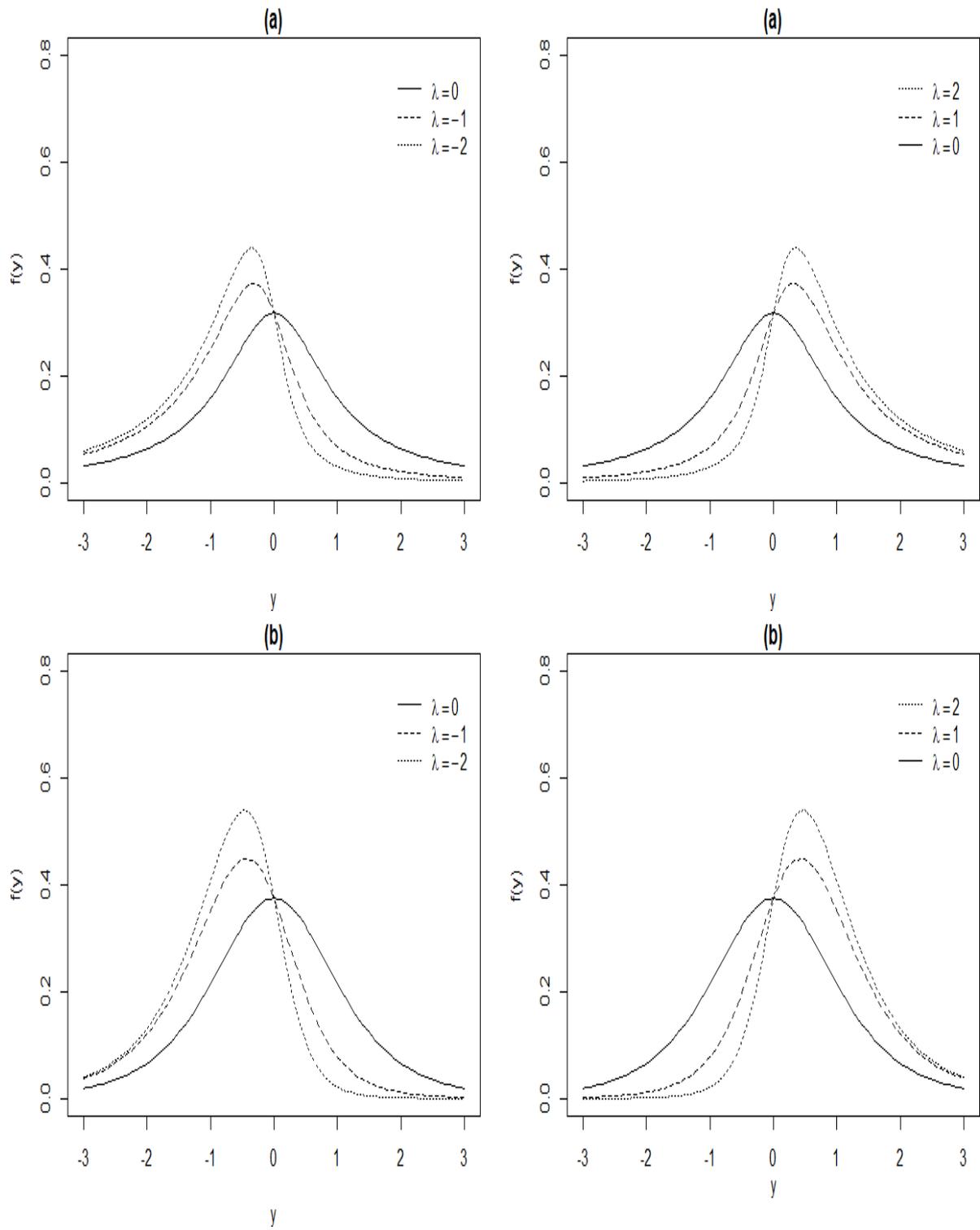
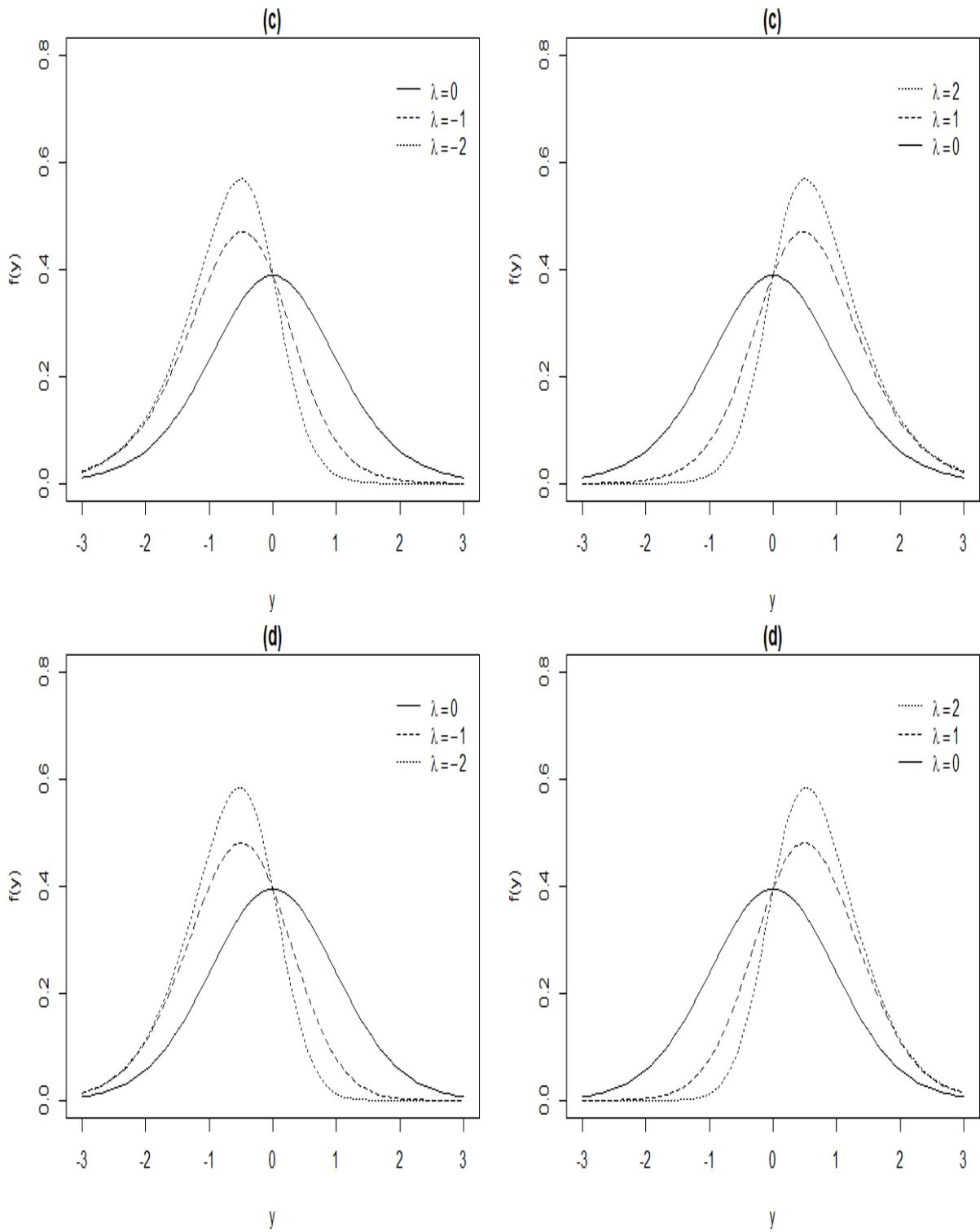


Figura 2.6 – Função de densidade da t assimétrica variando λ para $\nu = 10$ (c) e $\nu = 30$ (d).

2.4.4 Distribuição t assimétrica posição e escala

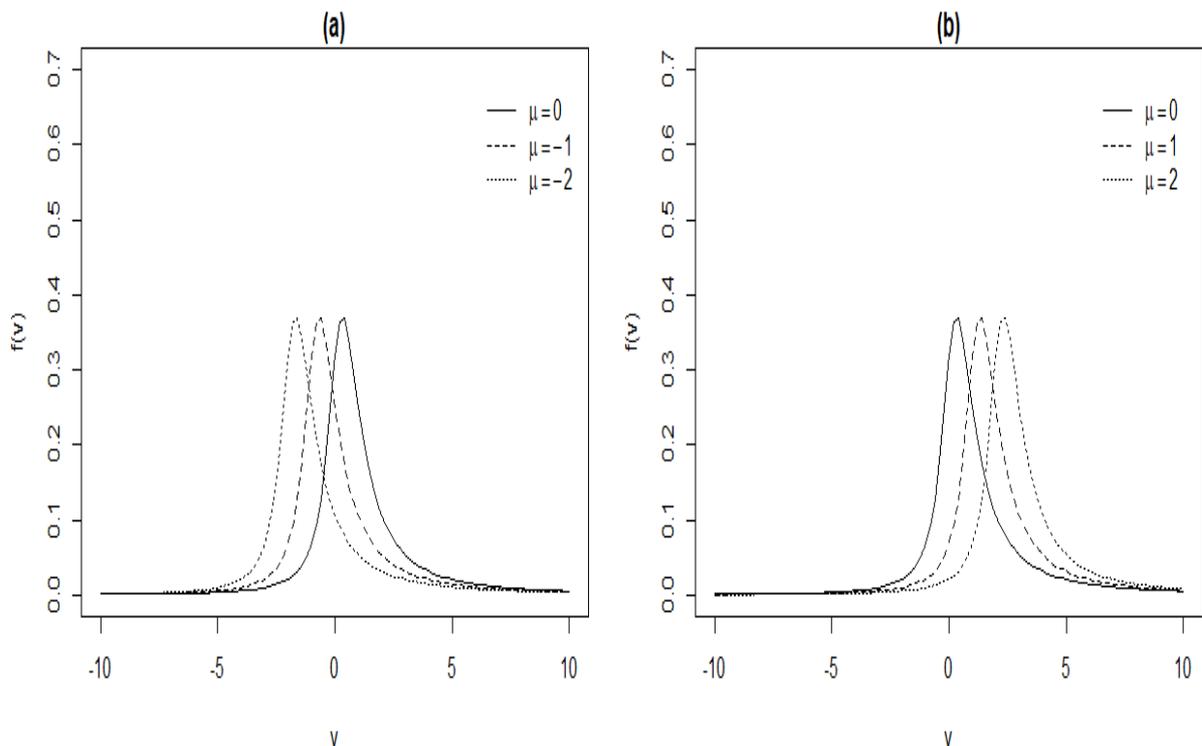
De acordo com Branco e Dey (2001), a distribuição t assimétrica posição e escala é uma extensão da distribuição t assimétrica com a adição de dois parâmetros, o parâmetro de posição ($\mu \in \mathbb{R}$) e o parâmetro de escala ($\sigma^2 > 0$). Diz-se que uma variável aleatória V segue uma distribuição t assimétrica com parâmetro de posição $\mu \in \mathbb{R}$ e parâmetro de escala $\sigma^2 > 0$, denotado por $Y \sim ST(\mu, \sigma^2, \nu, \lambda)$, se sua função densidade de probabilidade é dada pela seguinte expressão:

$$f_V(v) = \frac{2}{\sigma} t_{(\nu)} \left(\frac{v - \mu}{\sigma} \right) T_{(\nu+1)} \left[\lambda \left(\frac{v - \mu}{\sigma} \right) \sqrt{\frac{\nu + 1}{\nu + \sigma^{-2}(v - \mu)^2}} \right] \mathbb{1}_{\mathbb{R}}(v) \quad (2.4)$$

A distribuição também possui a seguinte propriedade se $Y \sim ST(\nu, \lambda)$ e $V = \mu + \sigma Y$, então $V \sim ST(\mu, \sigma^2, \nu, \lambda)$.

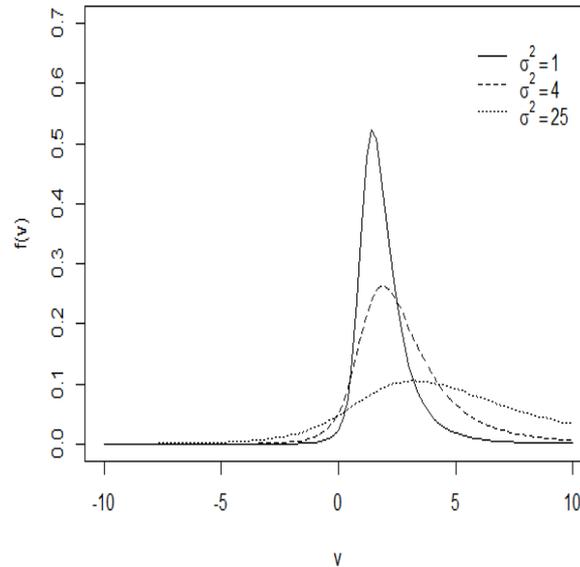
As Figuras 2.7, 2.8, 2.9 e 2.10 apresentam os gráficos da função de densidade da distribuição t assimétrica para alguns valores de μ , σ^2 , λ , ν .

Figura 2.7 – Função de densidade da t assimétrica variando μ para $\sigma^2 = 1$, $\nu = 1$ e $\lambda = 1$, em que (a) valores negativos de μ e (b) valores positivos para μ .



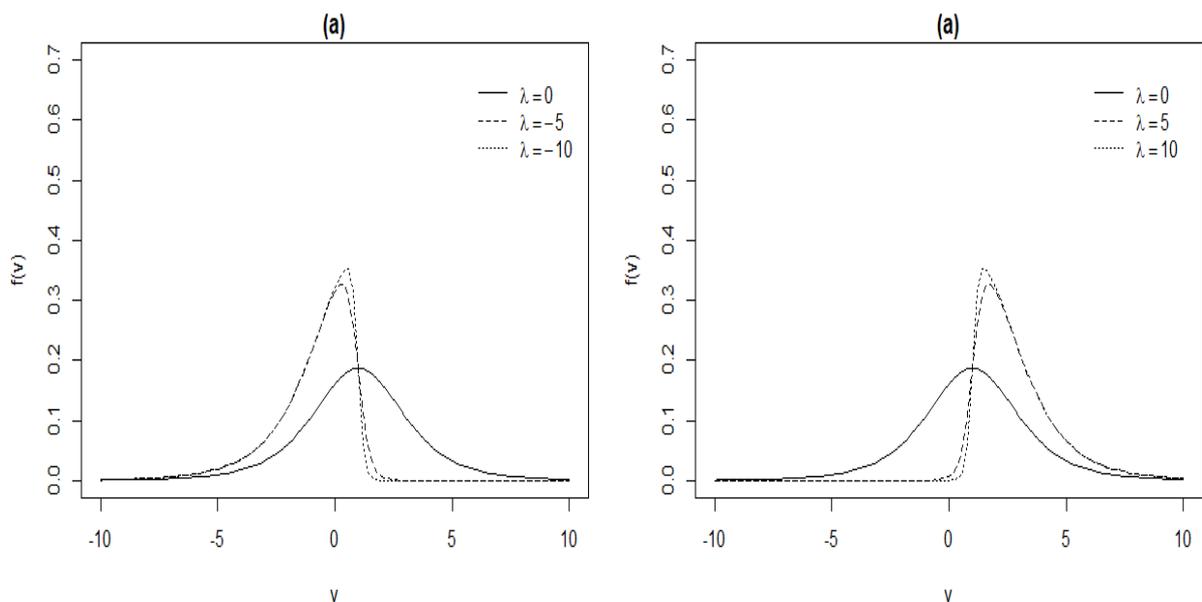
Pela Figura 2.7 (a) e (b), pode-se observar que alterações nos valores de μ implica em um deslocamento ao longo do eixo horizontal, sem alterações na forma inicial.

Figura 2.8 – Função de densidade da t assimétrica variando σ^2 para $\mu = 1$, $\nu = 3$ e $\lambda = 2$.



Através da Figura 2.8 pode-se observar uma achamento da distribuição para valores altos de σ^2 .

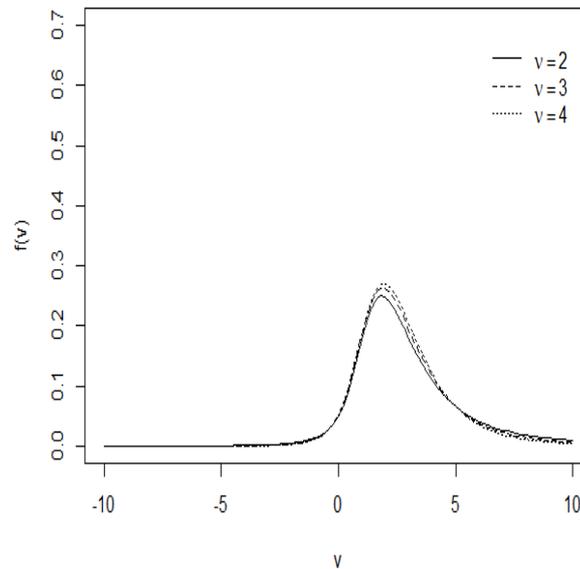
Figura 2.9 – Função de densidade da t assimétrica variando λ ((a-esquerda)- valores negativos e (a-direita) valores positivos) para $\sigma^2 = 4$, $\nu = 4$ e $\mu = 1$.



Quando ocorre uma alteração no parâmetro de assimetria λ , tem-se que para valores negativos existe uma assimetria à esquerda e caso contrário, assimetria à direita. Vale ressaltar,

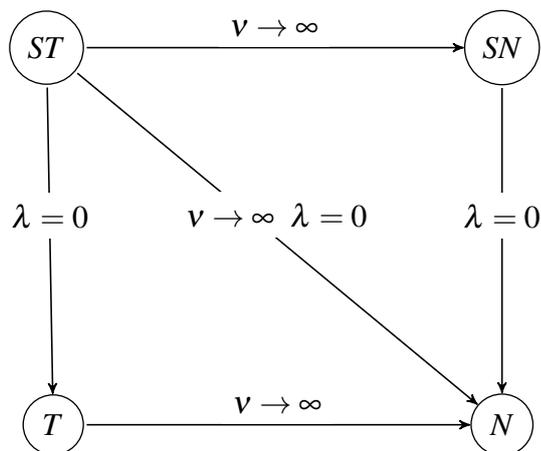
que quando o valor de λ é igual a zero tem-se a distribuição t assimétrica de posição e escala (ver Figura 2.9).

Figura 2.10 – Função de densidade da t assimétrica variando ν para $\sigma^2 = 5$, $\lambda = 2$ e $\mu = 1$.



Já a Figura 2.10, mostra a distribuição t assimétrica posição e escala com diferentes graus de liberdade. Na Figura 2.11 pode-se observar as seguintes relações das distribuições. Além disso, verificou-se que a distribuição normal é caso particular para outras distribuições com a normal assimétrica (SN), t assimétrica (ST) e t -Student (T).

Figura 2.11 – Relações das distribuições



2.5 Modelos não lineares (MNL)

Conforme Draper e Smith (1998), os modelos de regressão podem ser classificados de três maneiras, as quais estão relacionadas com a derivação dos parâmetros.

- Lineares - quando as derivadas parciais em relação aos parâmetros do modelo não dependem deles.
- Não lineares - quando as derivadas parciais em relação aos parâmetros do modelo dependem dos parâmetros.
- Linearizáveis - quando podem ser linearizados através de alguma transformação, como por exemplo a transformação logarítmica \ln .

Os modelos não lineares podem ser utilizados para descrever variáveis físicas e sistemas biológicos. Uma das vantagens desses modelos é que geralmente seus parâmetros possuem interpretações práticas, o que contribui para uso frequente nas diversas áreas do conhecimento. Draper e Smith (1998) apresentam MNL normal da seguinte forma:

$$Y_i = \eta(\beta, X_i) + \varepsilon_i \quad i = 1, 2, \dots, n \quad (2.5)$$

em que Y_i é o vetor da variável resposta (variável dependente), β é o vetor de parâmetros, $\eta(\cdot)$ é a forma funcional conhecida, X_i é o vetor de uma ou mais variáveis explicativas (variáveis independentes) e ε_i é o vetor de erros aleatórios associados ao modelo, ou seja, por suposição tem-se que $\varepsilon_i \sim \mathbb{N}(0, \sigma^2)$.

Diversos modelos não lineares podem descrever o crescimento de culturas, dentre eles, tem-se o Logístico, von Bertalanfy, Richards, Gompertz e Brody, pois esses modelos possuem um comportamento de uma curva sigmoideal ou crescente. Esse tipo de crescimento é característico em estudo com feijão (LIMA et al., 2017), milho (MANGUEIRA et al., 2016), entre outras cultivares.

A análise da curva de crescimento é de interesse nas mais diversas áreas de pesquisa, como na agricultura, biologia, química, ciências sociais, dentre outras áreas. Além disso, essa técnica auxilia os pesquisadores a inferir sobre as contribuições dos diversos processos presentes no desenvolvimento vegetal e animal (SEBER; WILD, 1989). Na biologia, o uso da análise da curva de crescimento está relacionado com a descrição do crescimento dos animais e das plantas. Na química, é utilizada no estudo das isotermas de adsorção. Já na agricultura, como são evidentes as vantagens econômicas e administrativas em conhecer como a produção cresce e o quão rápido ela cresce, a análise da curva de crescimento é empregada para verificar como esses fatores reagem às condições ou tratamentos ambientais.

A curva de crescimento geralmente é caracterizada por um sigmóide que possui como característica um crescimento monotonicamente no início do desenvolvimento até passar por um ponto de inflexão e atingir um máximo em uma assíntota (SILVA et al., 2011). Quando se modela a curva de crescimento, o objetivo é obter informações sobre a interpretação biológica dos parâmetros, com a finalidade de construir um modelo parcimonioso para as observações em estudo.

A análise de curvas de crescimento usando modelos não lineares tem o intuito de estimar as causas de crescimento e inferir sobre as possíveis contribuições nos diversos processos presentes no desenvolvimento vegetal. De acordo com Fernandes et al. (2014), o uso de modelos não lineares para as curvas de crescimento possibilita sumarizar informação a partir de um conjunto dados e resumí-las em alguns parâmetros geralmente com interpretações práticas.

Dentre os modelos não lineares, o modelo Logístico e Gompertz, são usados para descrever vários processos biológicos por meio da curva de crescimento.

2.6 Modelo não linear normal assimétrico

Conforme Cancho, Lachos e Ortega (2010), o modelo não linear normal assimétrico pode ser definido como:

$$Y_i = \eta(\beta, X_i) + \varepsilon_i \quad i = 1, 2, \dots, n. \quad (2.6)$$

em que Y_i é o vetor da variável resposta, $\eta(\cdot)$ é forma funcional conhecida, X_i é o vetor da variável explicativa, β é vetor de parâmetros e $\varepsilon_i \sim SN\left(-\sqrt{\frac{2}{\pi}}\lambda, \sigma^2, \lambda\right)$. Então tem-se que

$$Y_i \sim SN\left(\eta(\beta, X_i) - \sqrt{\frac{2}{\pi}}\lambda, \sigma^2, \lambda\right).$$

Considerando que a distribuição normal assimétrica é unimodal, ou seja, possui uma única moda, então, a distribuição possui assimetria negativa quando $\lambda < 0$, caso contrário, compreende-se a assimetria positiva. Vale ressaltar que quando $\lambda = 0$ tem-se a distribuição normal $N(\mu, \sigma^2)$.

Uma forma de estimar os parâmetros do modelo não linear quando os dados são distribuídos por uma normal assimétrica, é utilizar o método de máxima verossimilhança. De acordo com Cancho, Lachos e Ortega (2010), o logaritmo natural da função de verossimilhança, considerando um vetor β dada observações de uma amostra $Y = (y_1, \dots, y_n)^\top$ é dado por $l(\beta) = \sum_{i=1}^n l_i(\beta)$, em que:

$$l_i(\beta) = \log 2 - \frac{1}{2} \log 2\pi - \frac{1}{2} \log \psi - \frac{d_i}{2} + \log \Phi(K_i)$$

em que $\psi = \sigma^2 + \lambda^2$, $d_i = \frac{\left(Y_i - \eta(\beta, X_i) + \sqrt{\frac{2}{\pi}}\lambda\right)^2}{\sigma^2 + \lambda^2} \sim \chi_1^2$, $K_i = \frac{\lambda}{\sigma\psi^{1/2}} \left(Y_i - \eta(\beta, X_i) + \sqrt{\frac{2}{\pi}}\lambda\right)$, χ_1^2 é uma distribuição qui-quadrado com um grau de liberdade.

2.7 Modelo não linear t assimétrico

De acordo com Contreras-Reyes, Arellano-Valle e Canales (2014), assumindo que o erro aleatório ε_i possui distribuição t assimétrica, o modelo não linear t assimétrico é dado por:

$$Y_i = \eta(\beta, X_i) + \varepsilon_i \quad i = 1, 2, \dots, n. \quad (2.7)$$

em que Y_i é o vetor da variável resposta, $\eta(\cdot)$ é forma funcional conhecida, X_i é o vetor da variável explicativa, β é vetor de parâmetros e $\varepsilon_i \sim ST\left(-\sqrt{\frac{2}{\pi}}k_1\sigma\delta, \sigma^2, \lambda, \nu\right)$, em que $k_1 = \frac{\nu \Gamma((\nu-1)/2)}{2 \Gamma(\nu/2)}$, $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$ e $\nu > 1$. Com isso, $Y_i \sim ST\left(\eta(\beta, X_i) - \sqrt{\frac{2}{\pi}}k_1\sigma\delta, \sigma^2, \lambda, \nu\right)$.

A estimação de modelos que consideram erros com distribuição t assimétrica pode ser realizada, assim como na normal assimétrica, pelo método de máxima verossimilhança. A função de verossimilhança é descrita por:

$$L(\beta, \nu, \lambda; y) = \prod_{i=1}^n 2t(z_i; \nu) T\left(\lambda z_i \sqrt{\frac{1+\nu}{\nu+z_i^2}}; \nu+1\right)$$

em que $z_i = (Y_i - \eta(\beta, X_i) - \sqrt{\frac{2}{\pi}}k_1\sigma\delta)/\sigma^2$ e $t(z_i, \nu) = \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)\sqrt{\pi\nu}} \left(1 + \frac{z_i^2}{\nu}\right)^{-(\nu+1)/2}$

é a densidade de uma distribuição t-Student com ν graus de liberdade e $T(\cdot; \cdot)$ é a distribuição acumulada da t-Student.

2.8 Considerações sobre modelos não lineares aplicados a cultura do feijão

Alguns trabalhos desenvolvidos relacionados com a cultura do feijão utilizaram a modelagem não linear, por exemplo, o estudo envolvendo a estabilidade fenotípica da produtividade do feijão, realizado por Rosse e Vencovsky (2000). Nesse estudo, foi possível concluir que

o modelo foi útil para classificar os materiais genéticos segundo suas produtividades e seus padrões de resposta.

Com o objetivo de avaliar o crescimento de duas cultivares de feijão, MartinsFilho et al. (2008) utilizaram a abordagem bayesiana para ajustar o modelo não linear logístico. Nos resultados obtidos, concluíram que a metodologia permitiu comparar as curvas de crescimento de forma eficiente, sem o uso de teoria assintótica.

Moura et al. (2008) analisaram o acúmulo de fitomassa seca das culturas do feijão-caupi e do milho, submetidos a dois tipos de sistemas de plantio, o exclusivo e consorciado, nas condições climáticas do Semi-árido brasileiro, utilizando os modelos de crescimento Expolinear, Logístico e de Gompertz. Nesta pesquisa, observou-se que os modelos apresentaram bom desempenho e que os parâmetros biológicos gerados pelo modelo expolinear sugeriram que o sistema de plantio exclusivo é o mais adequado para o cultivo de feijão-caupi e milho.

Lucio, Nunes e Rego (2016) ajustaram modelos de regressão não linear para a produção de feijão-de-vagem e também identificaram o tamanho da parcela com melhor poder de explicação. Os modelos ajustados foram o Logístico e von Bertalanffy para o peso médio de vagens acumulado nas múltiplas colheitas e com diferentes tamanhos de parcelas. Os autores observaram que as estimativas para ambos modelos foram semelhantes entre si e, concluíram também, que o modelo Logístico proporcionou estimativas mais próximas da realidade, apresentando melhor desempenho na descrição do comportamento do peso médio de vagens no decorrer do ciclo produtivo da cultura.

Silva et al. (2018) modelaram o acúmulo do macronutriente de cálcio do feijoeiro cultivar Jalo em cinco densidades de semeadura usando modelos não lineares Gompertz e Logístico (Verhulst). Esses pesquisadores concluíram que o modelo Logístico foi o mais adequado para descrever os dados em estudo.

Por fim, Lima et al. (2019) ajustaram os modelos não lineares Logístico e von Bertalanffy para o acúmulo de biomassa ao longo do tempo para a cultivar de feijoeiro Jalo. Os autores verificaram que os modelos testados são adequados para a descrição do acúmulo da matéria seca, e ressaltaram que o modelo Logístico obteve um melhor ajuste.

Os autores elencado acima utilizaram pressuposição de normal nos erros. Dentre modelos não lineares existentes serão apresentados o Logístico e o Gompertz considerando a incorporação de outras distribuições para o erro com normal assimétrica e t-assimétrica.

2.9 Modelo Logístico

O modelo Logístico foi inicialmente proposto por Verhulst em 1838, para estudar o crescimento de populações humanas. Ao longo dos anos, seu uso se difundiu na descrição do crescimento de animais e plantas (VERHULST, 1838). Esse modelo apresenta algumas reparametrizações que pode ser encontrada em Ratkowsky (1983). Segundo Fernandes et al. (2015), o modelo reparametrizado com interpretação direta de todos os parâmetros é dado por:

$$Y_i = \frac{a}{1 + e^{k(b-x_i)}} + \varepsilon_i$$

em que Y_i é a i -ésima observação da variável resposta, a é assíntota superior ou o valor assintótico, que pode ser interpretado como tamanho máximo, acúmulo máximo, entre outros; b é a abscissa do ponto de inflexão; k está relacionado com crescimento ou indica o índice de maturidade, quanto maior for o valor de k menos tempo será necessário para a planta atingir a assíntota superior (α), x_i é o i -ésimo valor assumido pela variável explicativa; o ε_i é o erro aleatório associado à i -ésima observação.

O modelo apresenta forma sigmoidal e sendo simétrico em relação ao ponto de inflexão, o qual é dado por $\hat{x} = \hat{b}$, e é nesse ponto que ocorre a mudança de concavidade da curva. Desta forma, a taxa de crescimento atinge seu valor máximo. Ao substituir a expressão do ponto de inflexão na função logística obtém-se o valor da função $\hat{Y} = \frac{\hat{a}}{2}$, ou seja, é nesse valor que Y atinge a metade do valor máximo assintótico.

2.10 Modelo Gompertz

De acordo com Winsor (1932), o modelo foi proposto por Benjamin Gompertz, no qual, seu estudo apresentou grande avanço para o cálculo de anuidade e seguros contra morte. Conforme Fernandes et al. (2015), uma das reparametrizações mais adequadas é dada por:

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

em que, Y_i é a i -ésima observação da variável resposta; a é a assíntota superior (valor máximo a ser atingindo pela variável em estudo) ou valor assintótico máximo que pode ser interpretado como o peso máximo que a planta pode atingir; b é a abscissa do ponto de inflexão; k está relaci-

onado ao crescimento ou índice de maturidade; x_i os valores observados da variável explicativa e ε_i é o erro aleatório associado a observação.

Como característica o modelo apresenta formato sigmoidal, entretanto, não apresenta simetria em relação ao ponto de inflexão. Já com relação a estimativa do ponto de inflexão, o qual é dado por $\hat{x} = \hat{b}$, e é nesse ponto que ocorre a mudança de concavidade da curva. Desta forma, a taxa de crescimento atinge seu valor máximo. Ao substituir a expressão do ponto de inflexão na função logística obtém-se o valor da função $\hat{Y} = \frac{\hat{a}}{e}$, ou seja, é nesse valor que Y atinge aproximadamente 37% do valor máximo assintótico. Como isso, o ponto de inflexão é diferente do Logístico, pois não ocorre na metade da assíntota, mas sim em aproximadamente 37%. Assim, tem-se que $\frac{a}{e} < \frac{a}{2}$, em que e é a constante de Euler.

2.11 Métodos de estimação

As estimativas dos parâmetros podem ser realizadas de diversas maneiras, dentre elas, por meio dos métodos dos mínimos quadrados ou máxima verossimilhança. Esses métodos proporcionam o sistema de equações normais, que no caso de modelos não lineares, não possuem uma solução explícita, sendo necessário o auxílio de métodos iterativos para obtenção dessas estimativas.

No método de mínimos quadrados, sob determinadas condições, os seus estimadores são não viesados e apresentam variância mínima em relação a outros estimadores não viesados (DRAPER; SMITH, 1998). Esse método consiste em encontrar o melhor ajuste para os dados tentando minimizar a soma de quadrados dos resíduos entre a curva ajustada e os valores observados.

Para o uso dos métodos de mínimos quadrados não se faz necessariamente pressuposições. No entanto, para que as estimativas dos parâmetros sejam úteis para fazer algum tipo de inferência, é necessário levar em consideração algumas pressuposições em relação ao vetor de erros, tais como, erros independentes e identicamente distribuídos, que seguem uma distribuição normal de média zero e variâncias homogêneas. Desta forma, Draper e Smith (1998) concluíram que para a estimação dos parâmetros do modelo deve-se levar em consideração esses pressupostos.

Os modelos de regressão podem ser classificados de acordo com a estrutura dos seus erros em: modelos ordinários, nos quais, sua estrutura de erros não viola nenhuma das pressuposições, ou seja, $\boldsymbol{\varepsilon} \stackrel{i.i.d}{\sim} N(\mathbf{0}, \mathbf{I}\sigma^2)$, em que $\mathbf{I}_{n \times n}$ é a matriz identidade, e modelos ponderados,

cuja estrutura dos erros viola a pressuposição de homogeneidade de variância (homocedasticidade), $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{D}\sigma^2)$, sendo $\mathbf{D}_{n \times n}$ uma matriz diagonal, positiva definida, em que os elementos da diagonal ponderam a variância.

Conforme Hoffman e Vieira (1988), os modelos generalizados são aqueles em que a estrutura dos erros violam a pressuposição de independência ou/e a de homogeneidade de variância. Desta maneira, tem-se que os erros são correlacionados e possivelmente heteroscedásticos, $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{W}\sigma^2)$, em que $\mathbf{W}_{n \times n}$ é uma matriz simétrica positiva definida que representa as variâncias e covariância dos erros.

Para exemplificar o método de mínimos quadrados ordinários, considere o modelo na notação matricial. Um modelo não linear é definido por :

$$\mathbf{Y}_{n \times 1} = \mathbf{F}(\boldsymbol{\beta})_{n \times 1} + \boldsymbol{\varepsilon}_{n \times 1} \quad (2.8)$$

sendo que

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}; \mathbf{F}(\boldsymbol{\beta}) = \begin{bmatrix} f(X_1, \beta) \\ f(X_2, \beta) \\ \vdots \\ f(X_n, \beta) \end{bmatrix}; \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

em que \mathbf{Y} é um vetor de valores observados, $\mathbf{F}(\boldsymbol{\beta})$ é o vetor que contém a forma funcional conhecida e $\boldsymbol{\varepsilon}$ é o vetor de erro aleatório.

Desse modo, a partir do modelo matricial pode-se escrever o vetor de erros, que é dado por:

$$\boldsymbol{\varepsilon} = \mathbf{Y} - \mathbf{F}(\boldsymbol{\beta}). \quad (2.9)$$

O estimador para o vetor de parâmetros $\boldsymbol{\beta}$ é dado por $\hat{\boldsymbol{\beta}}$ que minimiza a soma dos quadrados dos erros

$$SQRes(\boldsymbol{\beta}) = \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = [\mathbf{Y} - \mathbf{F}(\boldsymbol{\beta})]'[\mathbf{Y} - \mathbf{F}(\boldsymbol{\beta})]$$

Conforme Draper e Smith (1998) e Bates e Watts (1988), essa minimização resulta em sistemas de equações normais, dado por:

$$\mathbf{J}'\mathbf{F}(\boldsymbol{\beta}) = \mathbf{J}'\mathbf{Y}$$

Esses métodos proporcionam o sistema de equações normais não lineares. Como o sistema não possui solução explícita, deve-se recorrer aos métodos iterativos para a obtenção da solução. Como se sabe, $\mathbf{F}(\boldsymbol{\beta})$ e \mathbf{J} dependem de $\boldsymbol{\beta}$, assim, é impossível obter uma forma fechada para a solução $\hat{\boldsymbol{\beta}}$, sendo necessário o uso de métodos computacionais para uma aproximação desta solução. Vale ressaltar que \mathbf{J} é a matriz de derivada parcial de \mathbf{F} em relação a $\boldsymbol{\beta}$ (DRAPER; SMITH, 1998).

Com o avanço da tecnologia, diversos métodos iterativos surgiram, porém os mais usados são o de Gauss-Newton, ou método de linearização, o método Steepest-Descent, ou método gradiente e o método de Marquardt. Segundo Bates e Watts (1988), os métodos diferenciam pela forma como $\boldsymbol{\beta} - \boldsymbol{\beta}^0$ é calculado para assegurar as atualizações no vetor de parâmetros, em que $\boldsymbol{\beta}^0$ é o vetor com valores iniciais. De um modo geral, os critérios básicos são:

- Gradiente: $\boldsymbol{\beta} - \boldsymbol{\beta}^0 = \mathbf{J}'\boldsymbol{\varepsilon}$.
- Gauss-Newton: $\boldsymbol{\beta} - \boldsymbol{\beta}^0 = (\mathbf{J}'\mathbf{J})^{-1}\mathbf{J}'\boldsymbol{\varepsilon}$.
- Newton: $\boldsymbol{\beta} - \boldsymbol{\beta}^0 = \mathbf{G}^{-1}\mathbf{J}'\boldsymbol{\varepsilon}$ em que \mathbf{G}^{-1} é a inversa da matriz gradiente.
- Maquardt: $\boldsymbol{\beta} - \boldsymbol{\beta}^0 = [\mathbf{J}'\mathbf{J} + \delta \text{diag}(\mathbf{J}'\mathbf{J})]^{-1}\mathbf{J}'\boldsymbol{\varepsilon}$, em que $(\mathbf{J}'\mathbf{J})^{-1}$ é matriz inversa generalizada.

Na utilização desses métodos, o algoritmo possivelmente vai convergir quando os valores iniciais forem adequados. Esses valores podem ser obtidos por estudos anteriores do problema, embasamento teórico ou por uma grade de valores que minimize a soma de quadrados residuais.

2.11.1 Método iterativo Gauss-Newton

Suponha que $\boldsymbol{\beta}^0$ é uma aproximação de $\hat{\boldsymbol{\beta}}$ por mínimos quadrados para um modelo não linear. Para $\boldsymbol{\beta}$ próximo de $\boldsymbol{\beta}^0$, considera-se uma expansão da série de Taylor de primeira ordem como

$$f(\mathbf{X}, \boldsymbol{\beta}) \approx f(\mathbf{X}, \boldsymbol{\beta}^0) + \mathbf{F}(\boldsymbol{\beta}^0)(\boldsymbol{\beta} - \boldsymbol{\beta}^0)$$

em que $\mathbf{F}(\boldsymbol{\beta}^0)$ é a matriz de primeiras derivadas parciais em relação a cada parâmetro $\boldsymbol{\beta}$. Agora definindo $\mathbf{e}(\boldsymbol{\beta})$ como sendo um vetor de erros, pode-se escrever

$$\mathbf{e}(\boldsymbol{\beta}) = \mathbf{Y} - f(\mathbf{X}, \boldsymbol{\beta}) \approx \mathbf{e}(\boldsymbol{\beta}^0) - \mathbf{F}(\boldsymbol{\beta}^0)(\boldsymbol{\beta} - \boldsymbol{\beta}^0)$$

Com o intuito realizar a minimização da soma dos quadrados dos resíduos $SQRes(\boldsymbol{\beta}) = \mathbf{e}(\boldsymbol{\beta})' \mathbf{e}(\boldsymbol{\beta})$, então tem-se que

$$SQRes(\boldsymbol{\beta}) \approx [\mathbf{e}(\boldsymbol{\beta}^0) - \mathbf{F}(\boldsymbol{\beta}^0)(\boldsymbol{\beta} - \boldsymbol{\beta}^0)]' [\mathbf{e}(\boldsymbol{\beta}^0) - \mathbf{F}(\boldsymbol{\beta}^0)(\boldsymbol{\beta} - \boldsymbol{\beta}^0)]$$

$$SQRes(\boldsymbol{\beta}) \approx \mathbf{e}(\boldsymbol{\beta}^0)' [\mathbf{e}(\boldsymbol{\beta}^0) - 2\mathbf{e}(\boldsymbol{\beta}^0)\mathbf{F}(\boldsymbol{\beta}^0)(\boldsymbol{\beta} - \boldsymbol{\beta}^0) + (\boldsymbol{\beta} - \boldsymbol{\beta}^0)' \mathbf{F}'(\boldsymbol{\beta}^0)\mathbf{F}(\boldsymbol{\beta}^0)(\boldsymbol{\beta} - \boldsymbol{\beta}^0)]$$

Com isso, o $SQRes(\boldsymbol{\beta})$ será minimizado quando

$$(\boldsymbol{\beta} - \boldsymbol{\beta}^0) = [\mathbf{F}'(\boldsymbol{\beta}^0)\mathbf{F}(\boldsymbol{\beta}^0)]^{-1} \mathbf{F}'(\boldsymbol{\beta}^0)\mathbf{e}(\boldsymbol{\beta}^0).$$

Assim, o próximo passo será dado por:

$$\boldsymbol{\beta}^{(0+1)} = \boldsymbol{\beta}^0 + [\mathbf{F}'(\boldsymbol{\beta}^0)\mathbf{F}(\boldsymbol{\beta}^0)]^{-1} \mathbf{F}'(\boldsymbol{\beta}^0)\mathbf{e}(\boldsymbol{\beta}^0).$$

O processo é repetido até se obter a convergência, a partir do critério de parada estabelecido pelo pesquisador. Vale ressaltar que mesmo com o critério definido, a convergência não é garantida. Essa convergência depende diretamente dos valores iniciais estabelecidos, ou seja, valores ruins aumentam o número de iterações para a convergência. Além disso, a escolha dos valores iniciais tem uma grande importância na estimação dos parâmetros no contexto não linear, pois uma má escolha pode resultar em uma convergência para um mínimo local, ou até mesmo não convergir. Draper e Smith (1998) mostraram algumas técnicas para a obtenção de bons valores iniciais que viabilizam a convergência rápida do processo de estimação. Entre essas técnicas, seria analisar o comportamento dos parâmetros do modelo analiticamente ou visualmente, por meio de gráficos.

2.12 Estimação por máxima verossimilhança

Um outro método para a estimação dos parâmetros em modelo não lineares é de máxima verossimilhança, que diferente do método de mínimos quadrados, consiste em maximizar a função de verossimilhança para a obtenção dos estimadores dos parâmetros. Conforme Bolfarine

e Sandoval (2001), considerando que X_1, \dots, X_n sejam amostras aleatórias da variável aleatória \mathbf{X} com função de densidade $f(x|\boldsymbol{\beta})$, então a função de verossimilhança de $\boldsymbol{\beta}$ é dada por:

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

O estimador $\hat{\boldsymbol{\beta}}$ de máxima verossimilhança de $\boldsymbol{\beta}$ é dado pelo valor que maximiza a $L(\boldsymbol{\beta}, x)$.

Usa-se o logaritmo natural de $L(\boldsymbol{\beta}, x)$, pois em geral é mais simples chegar aos mesmos resultados da maximização da função original. O logaritmo natural da função de verossimilhança de $\boldsymbol{\beta}$ é dado por:

$$l(\boldsymbol{\beta}, x) = \log(L(\boldsymbol{\beta}, x))$$

Para encontrar a solução do estimador de máxima verossimilhança basta resolver $U(\boldsymbol{\beta}) = 0$, em que $U(\boldsymbol{\beta})$ é conhecida como a função score que é dada por:

$$U(\boldsymbol{\beta}) = \frac{\partial l(\boldsymbol{\beta}; x)}{\partial \boldsymbol{\beta}}$$

Suponha que os $\varepsilon \stackrel{\text{iid}}{\sim} \mathbf{N}(\mathbf{0}, \mathbf{I}\sigma^2)$ então, sua função de verossimilhança é dada por:

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

Agora ignorando as constantes, a função log-verossimilhança é dada por:

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

Dado que σ^2 é máximo com relação a $\boldsymbol{\beta}$ quando $SQRes(\boldsymbol{\beta})$ é minimizado, em outras palavras, quando $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$ (estimativa de mínimos quadrados) (SEBER; WILD, 1989). Além de que $\frac{\partial l}{\partial \sigma^2} = 0$ tem solução $\sigma^2 = SQRes(\boldsymbol{\beta})/n$, esse valor é máximo (para $\boldsymbol{\beta}$) quando a segunda derivada é negativa. Ou seja, sugerindo que $\hat{\boldsymbol{\beta}}$ e $\hat{\sigma}^2 = SQRes(\hat{\boldsymbol{\beta}})/n$ são estimativas de máxima verossimilhança.

Ademais, pode-se usar o algoritmo EM como uma solução numérica das estimativas de máxima verossimilhança dos parâmetros dos modelos não lineares (MONAHAN, 2011). Conforme Montgomery, Peck e Vining (2012) a estimação por máxima verossimilhança é uma alternativa quando a distribuição dos erros de um modelo é diferente da distribuição normal.

2.13 Seleção de modelos

Nas pesquisas é comum o ajuste de vários modelos para descrever o fenômeno em questão, porém é necessário algum critério de seleção que permita ao pesquisador comparar e selecionar o modelo que melhor descreve os dados. Entre esses critérios de seleção está: critério de informação de Akaike (AIC), critério de informação Baeyiano (BIC) e Critério de Determinação Eficiente (EDC). Esses critérios foram levados em consideração, pois usam o mesmo critério estatístico que é a função de máxima verossimilhança como medida de ajustamento, entretanto definem valores críticos diferentes.

2.13.1 Critério de informação de Akaike

O critério de informação de Akaike (AIC) foi proposto por Akaike (1974). O AIC é baseado na penalização dos modelos com relação ao seu número de parâmetros, ou seja, evita que os modelos possuam o excesso de parâmetros. Esse critério é definido como

$$AIC = -2l(\hat{\theta}) + 2p \quad (2.10)$$

sendo que $l(\hat{\theta})$ é o logaritmo da função de máxima verossimilhança do modelo e p é o número de parâmetro. Desta forma, quanto menor for o valor do AIC, melhor será o modelo e, ainda, tem-se que o modelo descreve o fenômeno em estudo.

2.13.2 Critério de informação Bayesiano

O critério de informação bayesiano (BIC), proposto por Schwarz et al. (1978) é dada por:

$$BIC = -2l(\hat{\theta}) + p \ln(n) \quad (2.11)$$

em que p é o número de parâmetros a serem estimados e n é o número de observações da amostra. Assim como no AIC, quanto menor for o valor do BIC, mais o modelo será considerado como plausível para o estudo.

2.13.3 Critério de determinação eficiente

O Critério de Determinação Eficiente, proposto originalmente por Bai, Krishnaiah e Zhao (1989), é uma generalização dos critérios AIC e BIC, sendo definido por:

$$-2l(\hat{\theta}) + pc_n \quad (2.12)$$

em que p é o número de parâmetros, n é o número de observação da amostra e c_n é o termo de penalidade para critério associado, conforme listado na Tabela 2.1.

Tabela 2.1 – Termo de penalidade c_n de diferentes critérios de avaliação de modelos baseados em informações: AIC, BIC e EDC

Critério	c_n
AIC	2
BIC	$\ln(n)$
EDC	$0,2\sqrt{n}$

Para selecionar o modelo, usa-se o menor valor desse critério, dando um indicativo de que o modelo é razoável para representar o fenômeno em estudo.

3 MATERIAL E MÉTODOS

Os dados utilizados nesse trabalho são oriundos de um experimento realizado na área experimental do Departamento de Agricultura da Universidade Federal de Lavras (UFLA), em Lavras-MG, em um solo classificado como Latossolo vermelho distrófico, originalmente sob cerrado, em sistema de plantio convencional com uma aração e duas gradagens.

A cidade de Lavras está situada na região Sul de Minas Gerais, nas coordenadas 21°14' de latitude S; 45°00' de longitude W e altitude média de 918 m. O experimento foi realizado na safra das águas (primavera-verão) nos anos 2006/2007 e faz parte do projeto coordenado e conduzido pelo professor Dr. Messias José Bastos de Andrade juntamente com seus orientandos, parte do estudo pode ser visto em Alves et al. (2009) e Vieira et al. (2009).

Na Tabela 3.1, é possível visualizar a análise de fertilidade relativa a amostras de solo coletadas à profundidade de 0 a 20cm, realizadas no laboratório de Fertilidade do Solo do Departamento de Ciência do Solo da UFLA, antes da semeadura do feijão. A interpretação das estimativas é: AcF = acidez fraca, AcM = acidez média, BA= baixo, M = médio, B = bom, MB = muito bom, PC = plantio convencional.

Tabela 3.1 – Análise química de amostras de solo coletadas na profundidade de 0 a 20 cm do latossolo vermelho distrófico, antes do plantio.*

Características	(PC)
pH em H ₂ O	5,3 AcM
P (Mehlich 1) mg.dm ⁻³	8,9 Ba
K (Mehlich 1) mg.dm ⁻³	62,0 M
Ca cmolc.dm ⁻³	1,6 M
Mg cmolc.dm ⁻³	0,4 Ba
Al cmolc.dm ⁻³	0,3 Ba
S mg.dm ⁻³	31,7 MB
Matéria Orgânica dag.kg ⁻¹	3,4 M
V (%)	30,2 Ba

Fonte:Vieira (2006)

O experimento foi instalado no delineamento em blocos causalizados, com três repetições, sendo os tratamentos dispostos no esquema fatorial 5 x 8, com os fatores constituídos por cinco densidades de semeadura: 75, 145, 215, 285 e 355 mil plantas ha⁻¹ e por oito períodos de avaliação: 13, 23, 33, 43, 53, 63, 73 e 83 dias após a emergência (DAE) das plantas. Foi usado a cultivar de feijão Bolinha, que apresenta grãos amarelos e arredondados, com hábito de crescimento do tipo II de porte ereto, com cerca de 32 a 33 g por 100 grãos e ciclo médio (ALVES et al., 2009).

A parcela experimental foi constituída por quatro linhas de cinco metros de comprimento, espaçadas entre si por 0,5 metro (m), com área total de 10 m². Como área útil da parcela foram utilizadas as duas linhas centrais, perfazendo 5m².

Para obter o acúmulo de massa seca da parte aérea ao longo do ciclo da cultura, foram realizadas coletas a cada 10 dias a partir do 13° dia após a emergência (DAE). Em cada coleta foram amostradas cinco plantas por meio de corte a 1 cm do solo, separando-se as partes das plantas em haste (H), folhas (F), flor+vagens (V) e grãos (G); sendo posteriormente o material seco em estufa de circulação forçada de ar a 65-70°C, até peso constante. Esses materiais foram pesados em balança de precisão, determinando o peso da matéria seca das partes da planta, e em seguida, transformado em acúmulo de matéria seca por hectare, em $kg\ ha^{-1}$.

O progresso do acúmulo de nutrientes na parte aérea ao longo do ciclo da cultura foi obtido utilizando amostras da matéria seca, trituradas e encaminhadas ao laboratório de Análise Foliar no Departamento de Ciência do Solo da UFLA, no qual foi medida a porcentagem de cada nutriente nas parcelas, as quais, foram multiplicadas pelas suas respectivas matérias secas, obtendo, assim, a quantidade de nutriente em cada parcela, sendo os valores expressos em $kg\ ha^{-1}$. Em seguida, foram obtidos os valores observados de acúmulo de matéria seca nas diversas partes da planta, pela soma de cada parte, obtendo-se o peso de matéria seca e dos nutrientes na haste (H), na folha (F), na haste + folha (HF), na haste + folha + vagens (HFV) e no total (HFVG).

Como o maior interesse era estudar o efeito das épocas de avaliação sobre o acúmulo de matéria seca no total, e descrever como esse acúmulo ocorre, em cada densidade de semeadura, realizou-se o desdobramento da interação analisando as épocas dentro de cada densidade de semeadura, sendo as análises realizadas conforme sugestão de Gomes-Pimentel (2009) e Steel, Torrie e Dickey (1997). Assim, os dados de cada variável, inicialmente, foram submetidos à análise de variância (ANOVA tradicional) com o intuito de verificar o efeito das épocas de avaliação em cada densidade de semeadura utilizado-se as sugestões de Steel, Torrie e Dickey (1997) e Gomes-Pimentel (2009). Com relação análise residual da ANOVA tradicional em que os pressupostos não foram atendidos optou-se por usa um método não paramétrico como a ART ANOVA.

O dado de acúmulo da matéria seca no total foi submetido à análise de variância para um esquema fatorial, utilizando-se o seguinte modelo linear:

$$y_{ijkl} = \mu + b_j + d_i + a_k + da_{ik} + \varepsilon_{ijkl} \quad (3.1)$$

no qual,

y_{ijkl} é o valor observado de acúmulo de matéria seca no i -ésimo sistema de plantio, j -ésimo bloco, k -ésima densidade de semeadura e l -ésima época de avaliação;

μ é uma constante inerente;

b_j é o efeito do j -ésimo bloco ($j = 1, 2, 3$);

d_i é o efeito da i -ésima densidade de semeadura ($k = 1, 2, 3, 4, 5$);

a_k é o efeito da i -ésima época de avaliação $k = 1, 2, \dots, 8$;

da_{ik} é o efeito de interação entre k -ésima densidade de semeadura e i -ésima época de avaliação;

ε_{ijkl} é o efeito do erro experimental associado às observações y_{ijkl} .

Após a análise de variância foi realizada uma análise de resíduos, com a finalidade de verificar as pressuposições de normalidade e homocedasticidade. Para testar a normalidade residual foi utilizado o teste de Shapiro-Wilk (SHAPIRO; WILK, 1965) e a verificação de homogeneidade dos resíduos usou-se o teste de Bartlett (BARTLETT, 1937). A presença de autocorrelação residual foi avaliada pelo teste de Durbin-Watson (DURBIN; WATSON, 1950) e executou-se uma verificação da ausência de valores discrepantes através do teste de Grubbs (URVOY; AUTRUSSEAU, 2014). Além disso, para análise de resíduos em que os pressupostos não foram atendidos, foi utilizado o método não paramétrico de análise de variância com transformação de posto alinhando (SAWILOWSKY, 1990).

3.1 Análise de regressão

Os modelos não lineares Logístico (3.2) e Gompertz (3.3) utilizados para descrever o acúmulo de matéria seca do feijoeiro foram:

$$Y_i = \frac{a}{1 + e^{k(b-x_i)}} + \varepsilon_i \quad (3.2)$$

$$Y_i = ae^{-e^{k(b-x_i)}} + \varepsilon_i \quad (3.3)$$

em que Y_i é o acúmulo de matéria seca no total, a é acúmulo máximo; b é abscissa do ponto de inflexão; k está relacionado com crescimento ou indica o índice de maturidade, x_i é o dias de

após a emergência, em que i representa as épocas de avaliação (13, 23, 33, 43, 53, 63, 73 e 83 DAE) e o 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). Os modelos não lineares com erro assimétrico e normal foram ajustados aos dados de acúmulos de matéria seca no total em relação à idade da planta, em dias após a emergência, para cada densidade de semeadura.

A análise residual é importante na modelagem dos dados caso alguma pressuposição não seja atendida. Então, a verificação dos pressuposições de normalidade, independência e homoscedasticidade dos resíduos foi realizada através dos testes de Shapiro-Wilk, Breusch-Pagam, Durbin-Watson, respectivamente.

A seleção e comparação dos modelos foram realizadas utilizando os critérios AIC, BIC e EDC. Sendo assim, o modelo que apresentou a menor estimativa para os critérios, foi considerado o mais adequado para este estudo.

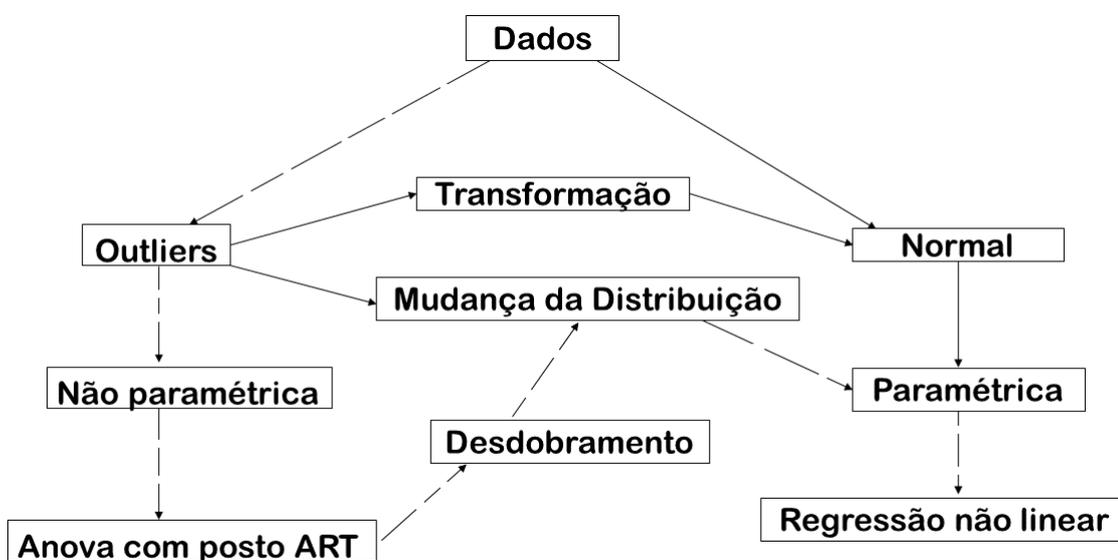
Para a estimação de máxima verossimilhança em modelos não lineares pode ter alguns problemas computacionais, no sentido de que o método pode não fornecer soluções globais máxima, se os valores iniciais para o algoritmo EM no contexto de modelos não linear forem distantes, vale ressaltar que esses valores desempenham um grande importância na estimativa dos parâmetros. No caso, considera-se os seguintes procedimento para calcular as estimativas de EM sob os modelos não lineares normal, normal assimétrica e t assimétrica.

- Calculou-se os valores iniciais de β^0 e $\sigma^{2(0)}$ ajustando os modelos usando o método de mínimos quadrados, no R foi utilizada a função `nls()`, considerando a normalidade residual.
- Calculou-se o valor inicial de λ^0 considerou-se o valor do coeficiente de assimetria dos resíduos obtido na etapa anterior.

Assim usou-se as estimativas de máxima verossimilhança usando o algoritmo EM para os modelos não lineares normais, normal assimétrico e t assimétrico.

Na Figura 3.1 observa-se um fluxograma para resumir a metodologia que foi aplicada nesse estudo. No caso, as linhas tracejada foram as etapas adotadas para a análise de dados.

Figura 3.1 – Fluxograma da metodologia adotada nesse estudo



3.1.1 Interface gráfica

Nesta pesquisa, foi ainda desenvolvida uma interface gráfica amigável, para que usuários pudessem carregar uma base com os dados de seus experimentos, escolher entre as três distribuições (normal, normal assimétrica e t assimétrica) e os dois modelos não lineares (Logístico e Gompertz). Por fim, tem-se como resultado a análise dos dados e a indicação da distribuição e modelo mais adequado para seu tipo de experimento. A seguir, seguem as descrições dos elementos da interface do sistema e suas respectivas funcionalidades:

- Botão para seleção de arquivo de entrada e área de visualização do nome e caminho do arquivo. Os arquivos a serem selecionados pelo usuário deverão estar no formato ".txt".
- Seleção do tipo de modelos não lineares: Logístico (1) e Gompertz (2).
- Seletor para variável resposta e explicativa. Os usuários podem selecionar .txt que contém os dados a serem utilizados, a coluna com os dados que devem ser usados para a variável resposta e a coluna da variável explicativa.
- Seletor para as distribuições dos erros. O usuário poderá escolher as distribuições Normal, skew normal (normal assimétrica) e Skew t (t assimétrica).
- Seletor para o tipo de regressão. O usuário pode selecionar o tipo de regressão como Homoscedastic (Homocedástico) sem problema na variância dos erros e Heteroscedastic

(Heterocedástico) com problema na variância dos erros. Além disso, pode-se selecionar a função de variância já predefinida e qual é a variável Z para modelar a variância.

- Lista com os valores iniciais para shape (assimetria), rho, nu (curtose) e betas. O usuário pode definir os valores iniciais baseados em um estudo prévio ou de conhecimento anterior de outros estudos.
- Lista para selecionar os modelos não lineares tem-se 1 para escolha do modelo Logístico e 2 para o modelo Gompertz.
- Seletor para os critérios de comparação. O usuário poderá escolher entre os critérios AIC, BIC e EDC
- Botão para realização da análise.

Assim como foi apresentada, a ferramenta oferece diversas funcionalidades aos usuários, tais como: facilidade no manuseio, robustez e eficácia. Na medida em que, poderá ser utilizada por pesquisadores para auxiliar na escolha de modelos assimétricos não lineares que mais se adequem ao tipo de experimento estudado, sem a necessidade desses pesquisadores desenvolverem qualquer tipo de programação de dados.

A interface foi desenvolvida no software R (ver Apêndice A) e adaptada de Duarte e Aoki (2019) e toda a programação da ferramenta que envolvem os parâmetros, os modelos não lineares e seleção foram implementados na linguagem de programação em R (R Core Team, 2019). Os pacotes *outliers* (teste de Grubbs), *ARTool* (ANOVA postos alinhados), *nlsmsn* (modelos não lineares assimétrico), *lmtest* (teste de Breusch-Pagan) e *tcltk2* (desenvolvimento da interface gráfica).

4 RESULTADOS E DISCUSSÕES

Procedeu-se inicialmente com uma análise descritiva e gráfica para verificar o comportamento do acúmulo de matéria seca do feijoeiro Bolinha. Posteriormente, foi realizada uma análise de variância com transformação de posto para verificar o efeito de interação no estudo. Além disso, foram ajustados os modelos não lineares Logístico e Gompertz com erro normal, erro normal assimétrico e erros t assimétrico.

4.1 Análise descritiva

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).

Tabela 4.1 – Estatística descritiva para o acúmulo da matéria seca no total do feijoeiro Bolinha ($kgha^{-1}$)

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

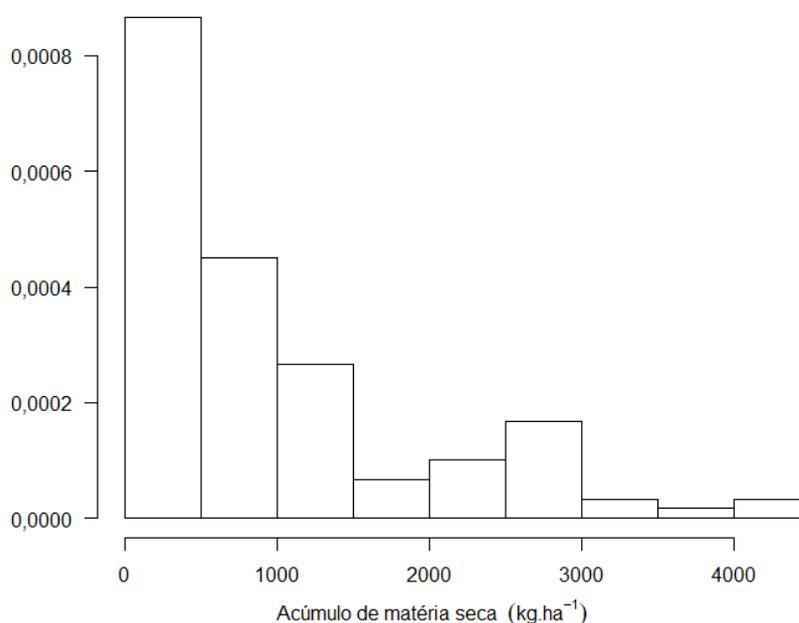
A assimetria é um valor adimensional que está associado com o grau de desvio ou afastamento da simetria de uma distribuição em função da média. Por isso, para revelar informações sobre a distribuição dos dados, é importante determinar o grau e a intensidade de assimetria.

Nesse sentido, a distribuição dos dados pode ser considerada como simétrica ou assimétrica, de tal modo que, quando o coeficiente de assimetria é igual a 0, a distribuição é classificada como simétrica. Caso o coeficiente de simetria esteja entre os valores 0 e 0,15, a distribuição é considerada assimétrica fraca. Por outro lado, se o coeficiente de assimetria pertence ao intervalo $[0,15;1)$, tem-se que a assimetria é moderada. Por fim, se o coeficiente é maior ou igual a 1, classifica-se a distribuição como assimétrica forte (PETRUCCI; OLIVEIRA, 2017). Um ponto importante a ser destacado é que se o coeficiente for maior que zero, a distribuição é assimétrica positiva ou à direita e, se o coeficiente for menor que zero, a distribuição é assimétrica negativa ou à esquerda.

Sendo assim, através dos dados apresentados na Tabela 4.1, nota-se que a distribuição para a variável em estudo é assimétrica forte e classificada como positiva (ou à direita), tendo em vista que o valor de assimetria encontrado (1,40) é maior que zero. Para uma melhor visualização dos dados, segue na Figura 4.1 o histograma dos dados para o acúmulo de matéria seca no total do feijoeiro.

O histograma apresentado na Figura 4.1 mostra que a distribuição dos dados não caracteriza uma distribuição normal. Note que existe um elevado agrupamento de dados nos valores mais baixos deixando a distribuição dos dados assimétrica. Assim, a cauda longa da distribuição fica à direita, sugerindo a ocorrência de valores altos com baixa frequência. Por meio da Figura 4.1, também pode-se verificar que os dados para o acúmulo de matéria seca no total do feijoeiro possui uma assimetria à direita (distribuição com assimetria positiva).

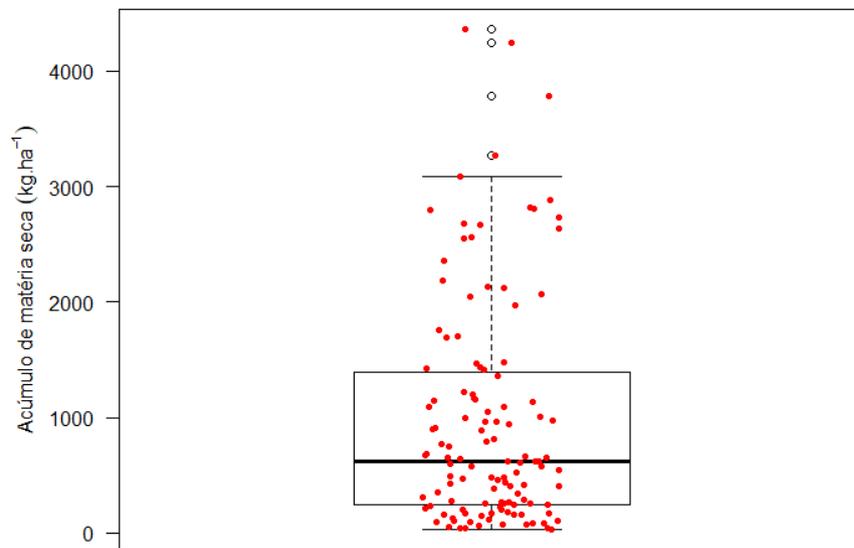
Figura 4.1 – Histograma para o acúmulo da matéria seca no total do feijoeiro Bolinha



De acordo com Schwertman, Owens e Adnan (2004), o boxplot é um método gráfico de identificação de *outliers* mais simples, pois não requer tanto esforço computacional. Os autores ainda ressaltam a importância de identificar valores atípicos que possam ter influência substancial na análise estatística, levando a possíveis distorções e conclusões imprecisas.

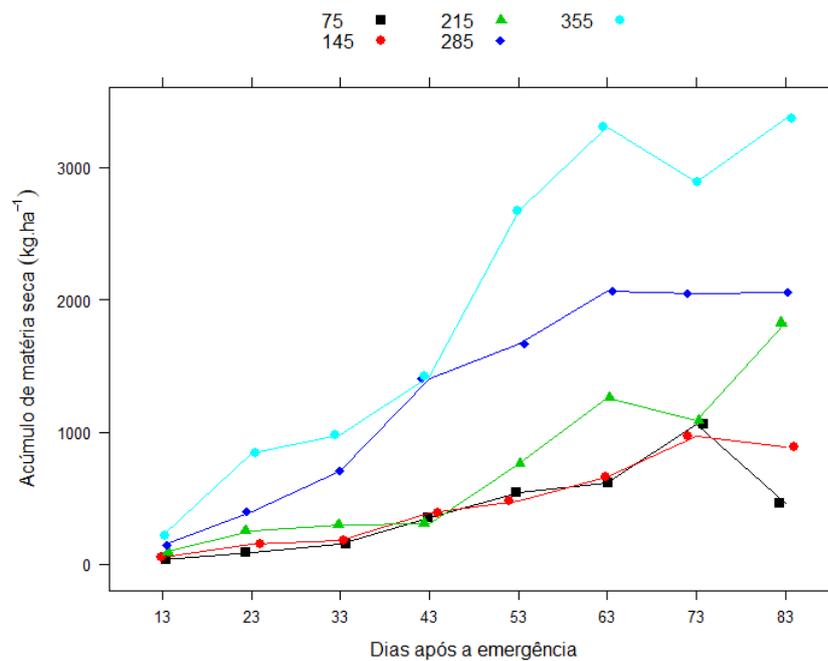
Desta forma, optou-se por usar o boxplot para análise visual de detecção de *outliers* e o teste de Grubbs para confirmação da visualização. Na Figura 4.2, pode-se observar um indicativo de que a distribuição dos dados é assimétrica (as bolinhas vermelhas representam as observações) e que os dados contém possíveis valores atípicos (as bolinhas brancas). A presença de *outliers* foi confirmada pelo teste de Grubbs (valor-p =0,025).

Figura 4.2 – Boxplot para o acúmulo de matéria seca no Total do feijoeiro cv Bolinha



Os valores médios do acúmulo de matéria seca, mostrados na Figura 4.3, sugerem um comportamento sigmoidal do acúmulo de matéria seca em relação aos dias após à emergência em cada densidade de semeadura, dando indicativo para o uso de modelos não lineares, como o Logístico, Gompertz, von Bertalanffy, Brody e etc, com a finalidade de estimar, por exemplo, o acúmulo máximo.

Figura 4.3 – Representação gráfica dos valores médios do acúmulo de matéria seca no Total do feijoeiro cv Bolinha da interação densidade de semeadura x dias após emergência



4.2 Análise de variância com transformação de postos alinhados (ART)

A análise de variância (ANOVA) de matéria seca obtida no Total está apresentada na Tabela 4.2. Através da ANOVA, verifica-se os efeitos significativos de densidade, DAE e interação sobre a variável analisada no estudo.

Tabela 4.2 – Análise de variância contendo as fontes de variação (FV), números de graus de liberdade (GL), valores de quadrados médios e respectivas significância do teste F, para matéria seca no Total, no feijoeiro cv. bolinha.

FV	GL	QMTotal
Bloco	2	853045*
Dens	4	10295275**
DAE	7	6082287**
Dens x DAE	28	549709**
Resíduo	78	177401

(ns) não significativo, (*) e (**) significativo pelo teste F ao nível de 5% e 1%, respectivamente

A verificação dos pressupostos da análise de variância relativos à normalidade, homogeneidade de variância e independência dos erros para as variáveis podem ser visto na Tabela 4.3. Pelos resultados obtidos a partir dos testes estatísticos, verificou-se a violação dos pressupostos da normalidade e homogeneidade de variância dos erros, já o teste de Durbin-Watson para independência indicou que os erros são independentes. A não rejeição da hipótese de independência dos erros, provavelmente, está relacionada ao método de amostragem realizado, pois de uma avaliação para outra, a amostragem é feita em parcelas distintas, e em experimentos desse tipo as amostras são destrutivas.

Tabela 4.3 – Estimativas e respectivos valor-p dos testes de normalidade (Shapiro -Wilk), homogeneidade de variância (Bartlett) e independência (Durbin-Watson) para matéria seca no Total do feijoeiro cv. Bolinha.

	Shapiro-Wilk		Bartlett		Durbin-Watson	
	W	valor-p	B	valor-p	DW	valor-p
Total	0,93	< 0,001	69,13	< 0,001	2,82	0,2452

Como pode ser visto, os dois pressupostos da ANOVA não foram atendidos. Desta maneira, uma alternativa é a utilização da transformação nos dados para estabilizar a variância e a normalidade dos resíduos, ou emprego de alguma técnica de análise não paramétrica, em que os pressupostos são mais brandos com relação a ANOVA.

Lima et al. (2017) obtiveram resultado semelhante para o feijoeiro do cultivar Jalo e propuseram uma transformação na variável resposta. Já Luepsen (2020) relataram que diversos

métodos de ANOVA não paramétrica, como a de transformação de posto alinhado, podem ser aplicados usando os procedimentos paramétrico padrão da ANOVA.

Wobbrock et al. (2011), avaliaram que a técnica ART foi considerada bastante eficiente com relação a sua aplicabilidade para experimento com esquema fatorial.

Nesse estudo, foi utilizada uma alternativa não paramétrica com transformação de posto, como pode ser visto na Tabela 4.4, na qual, verifica-se que a interação entre densidade e dias após emergência é significativa, bem como outras fontes de variação. Isto significa que, existe um efeito de dependência entre densidade de semeadura e dias após a emergência.

Tabela 4.4 – Análise de variância com transformação de postos contendo as fontes de variação (FV), números de graus de liberdade (GL), valores de F e respectivas significância do teste F, para matéria seca no Total, no feijoeiro cv. Bolinha.

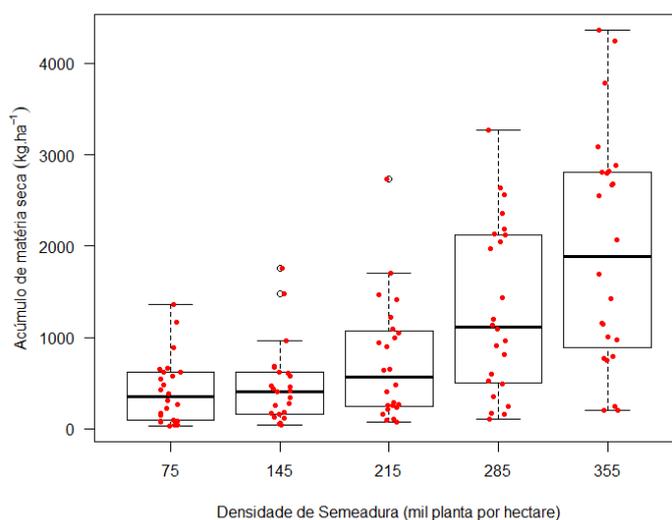
FV	GL	F
Bloco	2	5,89*
Dens	4	56,69**
DAE	7	56,17**
Dens x DAE	28	4,16**

(ns) não significativo, (*) e (**) significativo pelo teste F

4.3 Teste para outliers para cada densidade de semeadura

Foi realizado uma análise gráfica por meio do boxplot e do teste de Grubbs para verificar a presença de valores atípicos dentro de cada densidade de semeadura, como pode ser visto na Tabela 4.5 e Figura 4.4, em que no boxplot, os pontos em vermelho são observações para cada densidade de semeadura.

Figura 4.4 – Boxplot para o acúmulo de matéria seca para cada densidade de semeadura do feijoeiro cv Bolinha



Por meio da análise visual, pode-se observar a presença de possíveis pontos atípicos nas densidades 145 e 215, já na 75 verifica-se um comportamento assimétrico. Com relação as densidades 285 e 355 aparentemente a distribuição é simétrica, porém com maior variabilidade que as demais. (Figura 4.4).

Freitas et al. (2008) utilizaram técnicas de análise exploratória em dados da produção de matéria seca da alfafa. Entre essas técnicas, tem-se o boxplot que foi usado para verificar a ocorrência de *outliers*. Com isso, constataram prováveis ocorrências deles.

Dantas et al. (2015) avaliaram a produção de matéria seca e o controle de plantas daninhas por espécies de leguminosas. Nas três avaliações da massa de matéria seca da plantas no dias 29, 85 e 142 após a semeadura, usaram o boxplot e identificaram a presença de valores aberrantes ou atípicos.

Santos et al. (2020) usaram o gráfico de boxplot em diferentes períodos do desenvolvimento da cultura do milho. Assim, identificaram apresentação de alguns valores atípicos durante a fase desenvolvimento da planta em estudo.

Tabela 4.5 – Estatística do teste (G) e valor-p do teste de Grubb para o acúmulo de matéria seca no Total em cada densidade de semeadura

Densidade	G	valor-p
75	0,468	0,013
145	0,311	< 0,001
215	0,451	< 0,001
285	0,694	0,437
355	0,676	0,360

Todos os dados considerando a densidade de semeadura foram submetidos ao teste de Grubbs para verificação da existência de valores atípicos. Pelos resultados obtidos, pode-se verificar que existem *outliers* na densidade 75,145 e 215, pois obtiveram valores-p < 0,05.

Matoso e Kusdra (2014) realizaram uma análise estatística para a verificação de dados discrepantes na nodulação e crescimento do feijoeiro em resposta à aplicação de molibdênio e inoculante rizobiano usando teste de Grubbs para detecção de pontos atípicos.

Barbosa, Fernandes e Ferreira (2019) utilizaram o teste de Grubbs para verificação de *outlier* e obtiveram a presença do valor atípico no seu conjunto de dados. Além disso, concluíram que os substratos compostos orgânicos apresentaram um ph de 5,4 influenciam diretamente na absorção dos nutrientes nas plantas e, desta forma, proporcionam valores altos de massa seca da parte da raiz e do total.

Outros trabalhos que usaram o teste de Grubbs para a verificação da presença de dados discrepantes foram: Silva et al. (2016), na produção de mudas de couve manteiga do cultivar Geórgia, Uchôa et al. (2018), na avaliação de substratos alternativos na qualidade de mudas e produtividade do almeirão em cultivo orgânico, na qual verificaram que o composto orgânico propiciou valores superiores de massas secas da parte aérea e da raiz.

Galea, Paula e Cysneiros (2005) apresentaram resultados abordando inferência e diagnóstico na classe de modelos não lineares com erros simétricos independentes, com a finalidade de tentar reduzir a influência de observações atípicas nas estimativas dos coeficientes, e concluiu que uma das alternativas é assumir para os erros distribuições com caudas mais pesadas do que a normal.

Já Ferreira e Lachos (2016), apresentaram uma classe generalizada para modelos não lineares simétricos, pois abrangem as distribuições assimétricas. Essas distribuições com caudas pesadas são mais robustas a observações atípicas.

4.4 Ajuste dos modelos não lineares

Com a finalidade de avaliar o crescimento do feijoeiro em acúmulo de matéria seca, procedeu-se o desdobramento da interação entre DAE x Dens, analisando-se o efeito de DAE em cada densidade de semeadura e foi ajustado o modelo Logístico e Gompertz considerando os erros normais. Pela Figura 4.3, pode-se verificar um comportamento sigmoidal, então ajustou-se os modelos não lineares Logístico e Gompertz. Vários autores também usaram esses modelos como Prado, Savian e Muniz (2013) no coqueiro anão verde, Fernandes et al. (2014) no cafeeiro, Muianga et al. (2016) no cajueiro, Ribeiro et al. (2018) no pequi e Lima et al. (2019) no feijoeiro Jalo.

As pressuposições dos modelos foram verificadas para cada densidade de semeadura, ou seja, foram avaliadas a normalidade residual, homogeneidade de variância e independência. Na Tabela 4.6, tem-se as estimativas e os respectivos valores-p dos testes de normalidade (Shapiro-Wilk), de homoscedasticidade (Breusch-Pagan) e independência (Durbin-Watson).

Os resultados dos testes de Shapiro indicaram que não existe violação da normalidade para as densidades 285 e 355 em ambos modelos, pois apresentaram valores-p > 0,05 para o teste de Shapiro-Wilk. Não foi detectado problema com homogeneidade de variância e nem com a independência, isso foi observado através dos teste de Breusch-Pagan (valor-p > 0,05) e Durbin-Watson(valor-p > 0,05), respectivamente.

Tabela 4.6 – Estimativas e respectivos valor-p dos testes de normalidade (Shapiro -Wilk - W), de homoscedasticidade (Breusch-Pagan - BP) e independência (Durbin-Watson- DW) para os modelos Logístico e Gompertz

Densidade	Modelo	W	valor-p	BP	valor-p	DW	valor-p
75	L	0,906	0,029	5,910	0,052	2,100	0,430
75	G	0,905	0,028	5,360	0,069	2,070	0,410
145	L	0,853	0,002	4,680	0,096	2,640	0,880
145	G	0,846	0,002	4,170	0,120	2,650	0,890
215	L	0,895	0,017	6,940	0,074	2,130	0,460
215	G	0,902	0,024	4,720	0,094	2,140	0,460
285	L	0,953	0,311	5,320	0,070	2,620	0,870
285	G	0,947	0,230	4,760	0,093	2,570	0,850
355	L	0,965	0,547	5,830	0,054	2,570	0,850
355	G	0,971	0,697	5,250	0,072	2,50	0,790

Pode-se analisar pela Tabela 4.6, que as densidades 75, 145, 215 apresentaram problema com a falta de normalidade residual. Além disso, anteriormente verificou-se também a presença de possíveis *outliers* nessas densidades. Assim pode-se observar que os pontos atípicos podem ter afetado diretamente a normalidade residual. Mangueira et al. (2016) analisaram a altura da planta de milho ao longo do tempo e verificaram que a falta de normalidade residual. Assim, ressaltaram a importância de incorporação de outras distribuições para o erro, porém não realizaram teste para captação de possíveis valores atípicos.

Com isso, ajustou-se os modelos Logístico e Gompertz com erros normais, normais assimétrico (SN) e t assimétrico (ST) com 4 graus de liberdade, no qual esse valor foi baseado na estimativa da curtose do acúmulo da matéria seca total, para cada densidade de semeadura, os resultados obtidos são apresentados nas Tabelas 4.7 e 4.8, 4.9, 4.10 e 4.11.

Pelas estimativas obtidas do parâmetro α apresentadas nas Tabelas 4.7, 4.8, 4.9, 4.10 e 4.11, é possível analisar a ocorrência de um aumento da densidade e um aumento no acúmulo assintótico em ambos os modelos. Além disso, verificou-se um acúmulo gradativo da matéria seca ao longo do ciclo de vida do feijoeiro, e corroborando com o resultado desenvolvido por Vieira et al. (2008)

Tabela 4.7 – Estimativas e critérios de seleção para o modelo Logístico (L-75) e Gompertz (G-75) considerando erros normais, normal assimétrico (SN) e t assimétrico (ST) ajustados para o acúmulo de matéria seca no Total do feijoeiro cv. Bolinha na densidade de semeadura de 75 mil plantas.

	Parâmetros	Normal	SN	ST
L-75	a	737,031	666,154	639,288
	b	43,493	37,145	38,995
	k	0,122	0,093	0,123
	λ	0,000	2,764	0,493
	AIC	337,658	335,217	333,605
	BIC	343,548	341,107	340,673
	EDC	332,557	330,116	327,484
	G-75	a	748,782	694,277
b		38,387	29,856	34,346
k		0,086	0,055	0,084
λ		0,000	2,965	0,335
AIC		338,228	335,985	334,898
BIC		344,119	341,876	341,966
EDC		333,127	330,884	328,776

Tabela 4.8 – Estimativas e critérios de seleção para o modelo Logístico (L-145) e Gompertz (G-145) considerando erros normais, normal assimétrico (SN) e t assimétrico (ST) ajustados para o acúmulo de matéria seca no Total do feijoeiro cv. Bolinha na densidade de semeadura de 145 mil plantas.

	Parâmetros	Normal	SN	ST
L-145	a	1075,407	1075,354	8406,701
	b	53,649	53,647	207,229
	k	0,069	0,068	0,018
	λ	0,000	0,097	19,532
	AIC	347,577	347,574	324,792
	BIC	353,467	353,464	331,860
	EDC	342,476	342,473	318,670
	G-145	a	1382,104	1382,086
b		51,692	51,691	417,454
k		0,032	0,032	0,004
λ		0,000	0,009	12,163
AIC		347,762	347,762	325,555
BIC		353,653	353,653	332,623
EDC		342,661	342,661	319,434

Tabela 4.9 – Estimativas e critérios de seleção para o modelo Logístico (L-215) e Gompertz (G-215) considerando erros normais, normal assimétrico (SN) e t assimétrico (ST) ajustados para o acúmulo de matéria seca no Total do feijoeiro cv. Bolinha na densidade de semeadura de 215 mil plantas.

	Parâmetros	Normal	SN	ST
L-215	<i>a</i>	3057,336	2070,905	1690,998
	<i>b</i>	77,388	63,265	57,102
	<i>k</i>	0,050	0,045	0,050
	λ	0,000	4,452	2,347
	AIC	354,116	348,451	345,040
	BIC	360,006	354,342	352,110
	EDC	349,015	343,350	338,920
G-215	<i>a</i>	9834,875	6333,197	4376,834
	<i>b</i>	122,059	113,886	93,307
	<i>k</i>	0,014	0,012	0,013
	λ	0,000	4,702	2,986
	AIC	353,993	348,859	346,160
	BIC	359,884	354,749	353,230
	EDC	348,892	343,758	340,040

Tabela 4.10 – Estimativas e critérios de seleção para o modelo Logístico (L-285) e Gompertz (G-285) considerando erros normais, normal assimétrico (SN) e t assimétrico (ST) ajustados para o acúmulo de matéria seca no Total do feijoeiro cv. Bolinha na densidade de semeadura de 285 mil plantas.

	Parâmetros	Normal	SN	ST
L-285	<i>a</i>	2107,358	2107,343	2485,803
	<i>b</i>	37,980	37,979	45,531
	<i>k</i>	0,107	0,107	0,039
	λ	0,000	0,050	7,882
	AIC	379,592	379,590	373,371
	BIC	385,482	385,480	380,439
	EDC	374,491	374,490	367,249
G-285	<i>a</i>	2210,495	2210,487	2994,576
	<i>b</i>	32,515	32,515	39,075
	<i>k</i>	0,066	0,065	0,021
	λ	0,000	0,040	7,453
	AIC	379,877	379,880	373,632
	BIC	385,767	385,770	380,700
	EDC	374,776	374,780	367,510

Tabela 4.11 – Estimativas e critérios de seleção para o modelo Logístico (L-355) e Gompertz (G-355) considerando erros normais, normal assimétrico (SN) e t assimétrico (ST) ajustados para o acúmulo de matéria seca no Total do feijoeiro cv. Bolinha na densidade de semeadura de 355 mil plantas.

	Parâmetros	Normal	SN	ST
L-355	a	3432,483	3330,525	3290,185
	b	42,265	40,411	40,443
	k	0,089	0,082	0,084
	λ	0,000	1,981	0,720
	AIC	376,905	376,005	376,322
	BIC	382,795	381,896	383,391
	EDC	371,804	370,904	370,201
G-355	a	3814,359	3738,305	3702,188
	b	36,648	34,479	34,760
	k	0,048	0,042	0,044
	λ	0,000	2,229	0,800
	AIC	378,441	377,421	378,657
	BIC	384,331	383,311	385,725
	EDC	373,340	372,320	372,535

Na Tabela 4.7 para o modelo Logístico em L-75, o parâmetro a (assintótico superior) para as distribuições Normal e ST foi de 737,031 e 639,288, respectivamente. Assim, representando uma redução de 13,26% para esta estimativa. Já comparando a distribuição Normal e SN tem-se o valor máximo de 737,031 e 666,154, respectivamente. Com isso, verifica-se uma redução de 9,62% na estimativa de a . Além disso, pode-se observar que para as densidades 75 (Tabela 4.7), 145 (Tabela 4.8) e 215 (Tabela 4.9) com detecção de pontos atípicos uma possível redução na estimativa quando comparada com a distribuição normal para ambos os modelos.

Já nas Tabelas 4.10 e 4.11 verifica-se que os valores de estimativas para os parâmetros estão bem próximo da distribuição Normal. Vale ressaltar que as densidades 285 e 355 não tinha a presença de *outliers*

Pela Tabela 4.8, verificou-se que não ocorreu uma superestimação na estimativa do acúmulo máximo de matéria seca (a) e na abscissa associada ao ponto de inflexão (b). Assim, tornando-as não tão confiáveis. Na Tabela 4.9, o mesmo problema foi identificado para a densidade 215 para o modelo Gompertz com todas as distribuições estudadas. Esses resultados corroboram com Lima et al. (2017), que encontraram altos valores nas estimativas dos parâmetros para os modelos Logístico e von Bertalanffy.

Nas densidades mais baixas (75 e 145 mil planta por hectare) analisa-se que o acúmulo da matéria seca é menor considerando as demais densidades. Na densidade de 215, observa-se

ainda um incremento no acúmulo de massa seca, não atingindo uma estabilização de desenvolvimento (Figura 4.7).

A densidade de planta é influenciada por diversos fatores, dentre os quais se destacam o porte, a arquitetura e o hábito de crescimento da planta. Ademais, esses fatores apontam o nível de competição e o período do desenvolvimento que isso se intensifica entre as plantas (AMARO et al., 2014).

Com relação a , pode-se observar que o acúmulo máximo da matéria seca variou entre 630 kg ha^{-1} a 3850 kg ha^{-1} . Vale ressaltar que as densidades 215 e 285 mil planta por hectare apresentaram os maiores valores de acúmulo de matéria seca. Por outro lado, para as densidades de semeadura 75, 145, 215 e 355 mil plantas por hectare verificou-se uma redução na estimativa de a quando comparado as distribuições do erro.

De acordo com Albuquerque (2011), usando modelo cúbico encontrou os valores de máximo acúmulo de matéria seca para cada densidade na cultivar bolinha, em que obteve os seguintes valores: 977 kg ha^{-1} , 1651 kg ha^{-1} , 1667 kg ha^{-1} , 2454 kg ha^{-1} e 3331 kg ha^{-1} para 75, 145, 215, 285 e 355 mil plantas por hectare, respectivamente. Além disso, observou também a respectiva idade da planta que se atinge o máximo de acúmulo de matéria seca 65 DAE e 66 DAE para as densidade 75 e 215, e para as demais 73 DAE.

Com relação à abscissa associado ao ponto de inflexão (b), pode-se avaliar uma variação entre 29 a 77,5 DAE. Além disso, observa-se no modelo Gompertz esse ponto ocorre antes do que no Logístico, pelo fato de não ser simétrico. Com exceção da densidade 215 no qual ocorreu um problema de estimação.

Já a taxa de maturidade k , teve pouca variação considerando as diferentes distribuições nos dois modelos e variando de 0,01 a 0,10.

Com relação ao parâmetro λ que está associado a assimetria da distribuição, pode-se observar que os modelos para cada densidade de semeadura exibiram um valor de assimetria positiva, ou seja, os resíduos dos modelos apresentaram uma leve assimetria à direita. Os valores variaram de 0 à 19,6, ressaltando que zero indica uma distribuição normal.

Para a escolha entre os modelos Logístico e Gompertz, observou-se os menores valores de AIC, BIC e EDC presente nas tabelas 4.7 a 4.11, que indica um bom ajuste para o modelo. Ademais, retirou-se da seleção os modelos que apresentaram uma superestimação. Desse modo, o modelo Logístico com distribuição SN foi selecionada para as densidades 145, 285 e 355. Já para as densidades 75 e 215, o mesmo modelo com distribuição ST.

Os valores do acúmulo de matéria seca e as curvas ajustadas desse acúmulo em função dos dias após a emergência para cada densidade de semeadura podem ser visualizados nas Figuras 4.5, 4.6, 4.7, 4.8 e 4.9.

Figura 4.5 – Representação gráfica do acúmulo de matéria seca e ajuste do modelo Logístico (a) e Gompertz (b) com erros normais, normais assimétricos (SN) e t assimétricos em função dos dias após a emergência (DAE) para o acúmulo de matéria seca no total cv. Bolinha na densidade de semeadura de 75 mil.

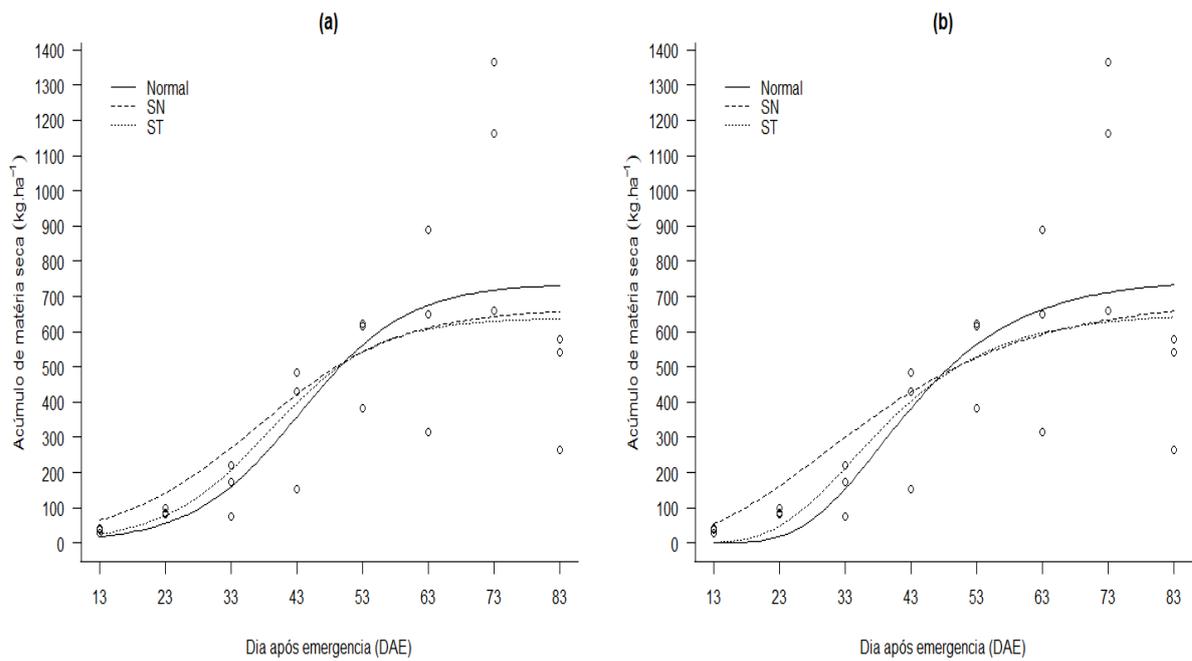


Figura 4.6 – Representação gráfica do acúmulo de matéria seca e ajuste do modelo Logístico (a) e Gompertz (b) com erros normais, normais assimétricos (SN) e t assimétricos em função dos dias após a emergência (DAE) para o acúmulo de matéria seca no total cv. Bolinha na densidade de semeadura de 145 mil.

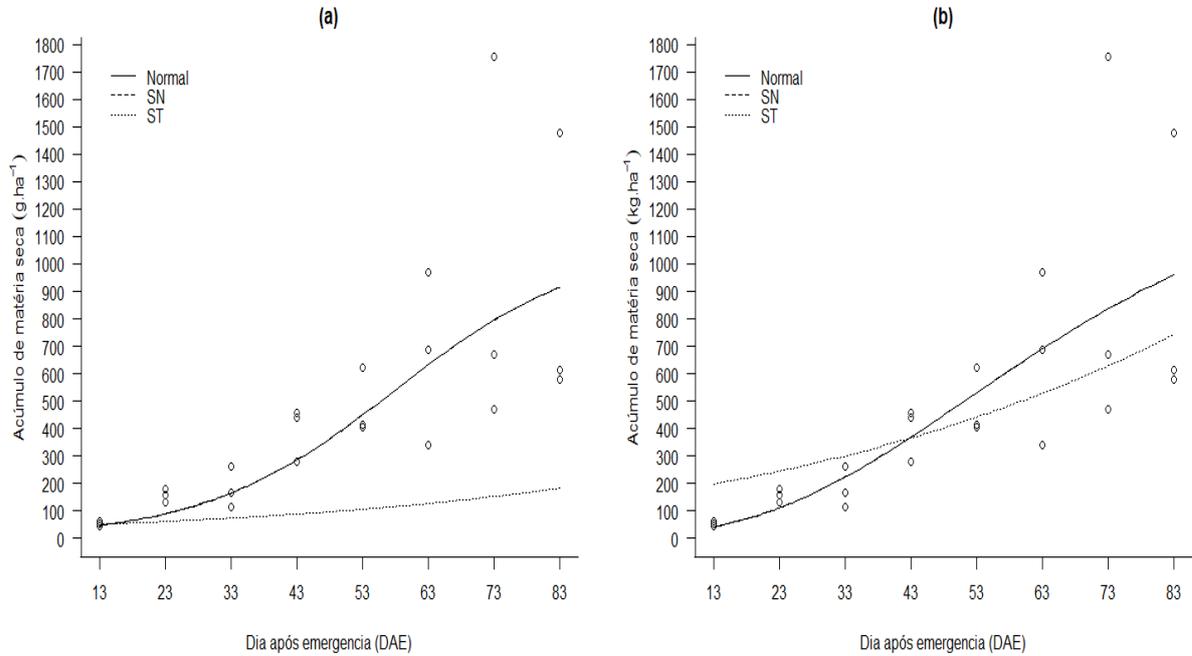


Figura 4.7 – Representação gráfica do acúmulo de matéria seca e ajuste do modelo Logístico (a) e Gompertz (b) com erros normais, normais assimétricos (SN) e t assimétricos em função dos dias após a emergência (DAE) para o acúmulo de matéria seca no total cv. Bolinha na densidade de semeadura de 215 mil.

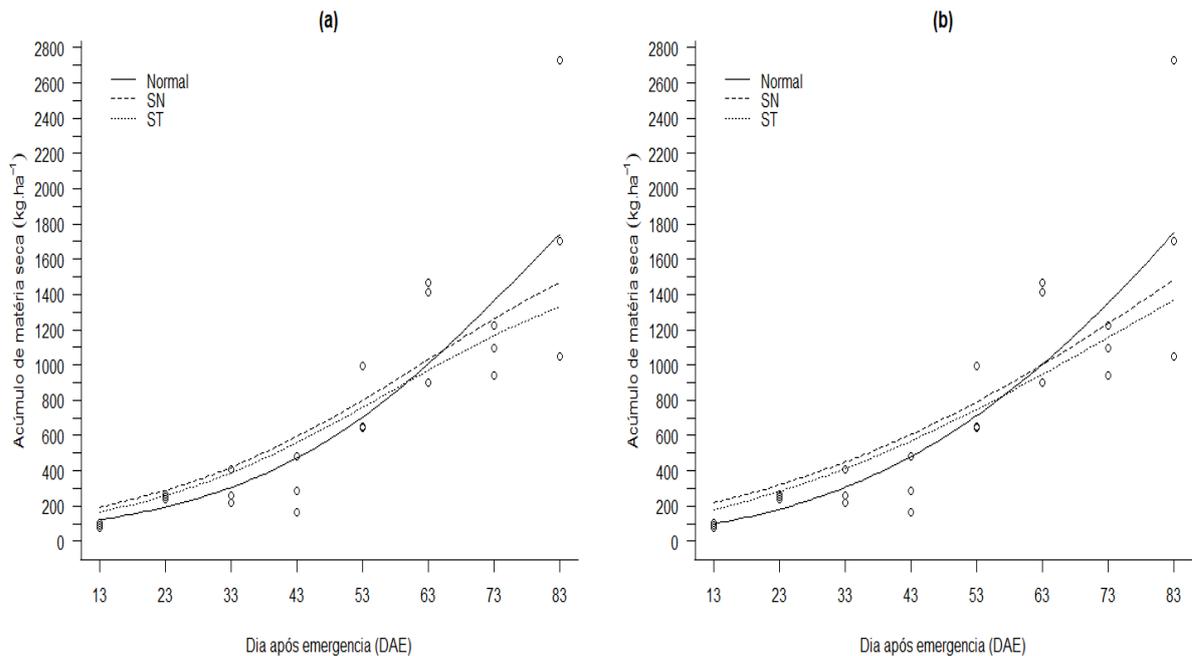


Figura 4.8 – Representação gráfica do acúmulo de matéria seca e ajuste do modelo Logístico (a) e Gompertz (b) com erros normais, normais assimétricos (SN) e t assimétricos em função dos dias após a emergência (DAE) para o acúmulo de matéria seca no total cv. Bolinha na densidade de semeadura de 285 mil.

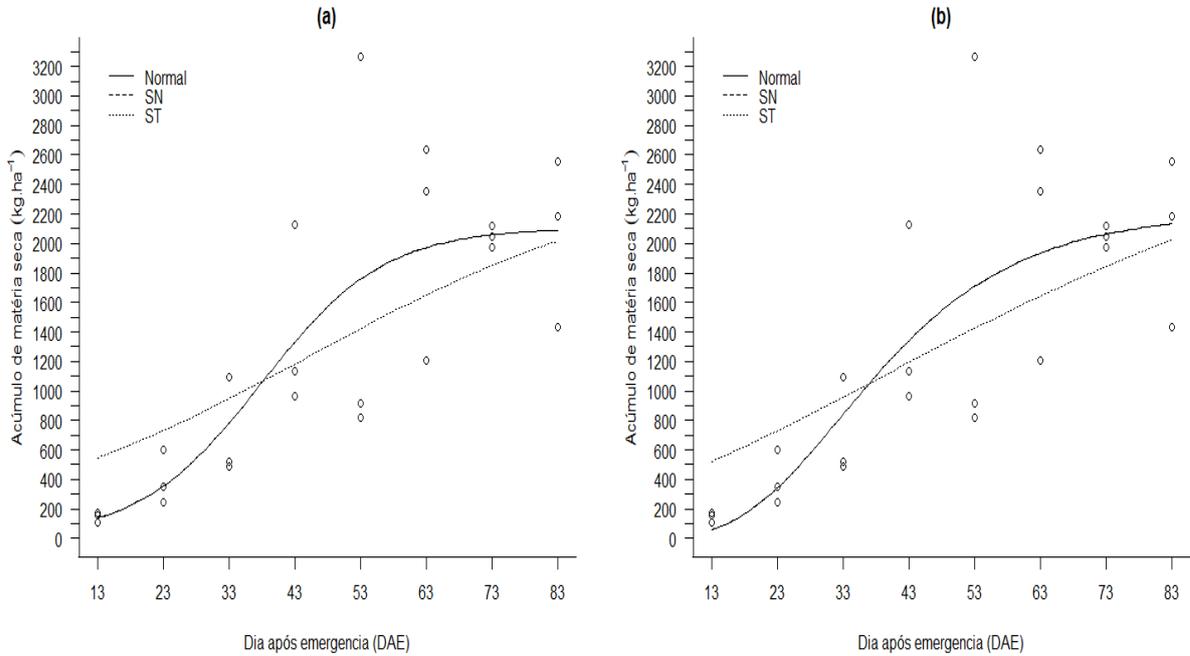
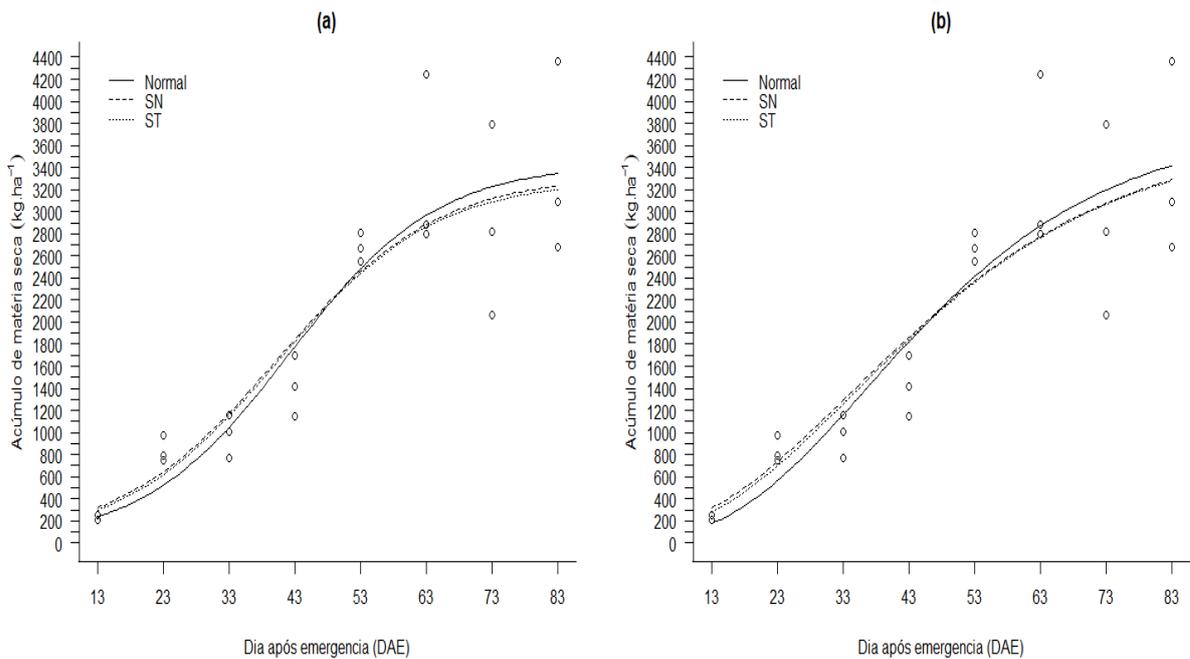


Figura 4.9 – Representação gráfica do acúmulo de matéria seca e ajuste do modelo Logístico (a) e Gompertz (b) com erros normais, normais assimétricos (SN) e t assimétricos em função dos dias após a emergência (DAE) para o acúmulo de matéria seca no total cv. Bolinha na densidade de semeadura de 355 mil.



Por meio das Figuras 4.5, 4.7 e 4.9, pode-se observar que os modelos com erros normais tendem a trazer as curvas ajustadas para perto dos *outliers*. Já quando considera-se as distribuições assimétricas, as curvas não sofrem tanta influência dos pontos atípicos. Além disso, pode-se observar que a distribuição ST não levou em conta no ajuste as observações (atípicas e não atípicas). No caso, das Figuras 4.6 e 4.8 observou-se a modelagem com distribuição ST como livre das observações discrepantes, assim ajustando quase uma reta sendo mais evidente na Figura 4.8.

Pode-se observar ainda que a SN e ST fornecem graficamente resultados bem semelhantes retirando os casos que ocorreu superestimação das estimativas do parâmetros. Assim, tem-se que os erros com as distribuições SN e ST depositaram menos peso em suas caudas, com isso são distribuições mais propensa e resistentes a outliers.

5 CONCLUSÃO

Os modelos Logísticos e Gompertz foram adequados para descrever o acúmulo de matéria seca nas cinco densidades de semeadura (75, 145, 215, 255, 285 e 355 mil plantas por hectares). Além disso, a distribuição normal assimétrica e t assimétrica conseguiram captar os valores atípicos encontrados e detectar assimetria positiva nos dados. Houve superestimação com relação a estimativa dos parâmetros para alguns modelos.

O acúmulo máximo total da matéria seca, estimado pelo modelo Logístico com erro normal, foi de $1075,354 \text{ kg ha}^{-1}$, $2107,343 \text{ kg ha}^{-1}$ e $3330,525 \text{ kg ha}^{-1}$, para as densidades 145, 285 e 355, respectivamente. Considerando a distribuição t assimétrica, esse valor foi $639,28 \text{ kg ha}^{-1}$ e $1769,456 \text{ kg ha}^{-1}$, para densidade 75 e 215, respectivamente.

Através do ajuste dos modelos, ficou evidente a importância de se conhecer o comportamento dos dados, tendo em vista que as distribuições normal assimétrica e t assimétrica apresentaram um comportamento igual ou superior a normal, mesmo em cenários onde não foram detectados outliers.

Além disso, a interface gráfica pode facilitar o usuário leigo em programação a ajustar os modelos não lineares com distribuição normal, normal assimétrica e t assimétrica para tomada de decisão na sua pesquisa.

REFERÊNCIAS

- AIGNER, D.; LOVELL, C. K.; SCHMIDT, P. Formulation and estimation of stochastic frontier production function models. **Journal of econometrics**, v. 6, n. 1, p. 21–37, 1977.
- AKAIKE, H. A new look at the statistical model identification. **IEEE transactions on automatic control**, v. 19, n. 6, p. 716–723, 1974.
- ALBURQUERQUE, A. D. **Acúmulo de matéria seca e macronutrientes em diferentes órgãos de cultivares de feijoeiro em função de populações de plantas e ambientes de cultivo**. 2011. Tese (Doutorado), UFLA (Universidade Federal de Lavras), Lavras, Brasil.
- ALVES, A. F. et al. Densidades populacionais para cultivares alternativas de feijoeiro no norte de minas gerais. **Ciência e Agrotecnologia**, v. 33, n. 06, p. 1495–1502, 2009.
- AMARO, H. T. R. et al. Qualidade fisiológica de sementes de cultivares de feijão em função de densidades populacionais. **Semina: Ciências Agrárias**, v. 35, n. 3, p. 1241–1248, 2014.
- ANDRADE, M. J. B. et al. Comunicação épocas de colheita em cinco cultivares de feijoeiro. ii. efeitos sobre a qualidade da semente. **Ciência e Agrotecnologia**, v. 33, n. 3, p. 683–689, 2001.
- ANSCOMBE, F. J.; TUKEY, J. W. The examination and analysis of residuals. **Technometrics**, v. 5, n. 2, p. 141–160, 1963.
- ARELLANO-VALLE, R. B.; BRANCO, M. D.; GENTON, M. G. A unified view on skewed distributions arising from selections. **Canadian Journal of Statistics**, v. 34, n. 4, p. 581–601, 2006.
- ARELLANO-VALLE, R. B.; GENTON, M. G.; LOSCHI, R. H. Shape mixtures of multivariate skew-normal distributions. **Journal of Multivariate Analysis**, v. 100, n. 1, p. 91–101, 2009.
- ARNOLD, B. C.; BEAVER, R. J. The skew-cauchy distribution. **Statistics & probability letters**, v. 49, n. 3, p. 285–290, 2000.
- ARYAL, G.; NADARAJAH, S. On the skew laplace distribution. **Journal of information and optimization sciences**, v. 26, n. 1, p. 205–217, 2005.
- AZZALINI, A. A class of distributions which includes the normal ones. **Scandinavian journal of statistics**, p. 171–178, 1985.
- 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.
- AZZALINI, A.; CAPITANIO, A. **The Skew-Normal and Related Families**. New York: Cambridge University Press, 2014.
- AZZALINI, A.; VALLE, A. D. The multivariate skew-normal distribution. **Biometrika**, v. 83, n. 4, p. 715–726, 1996.
- BAI, Z. .; KRISHNAIAH, P. R.; ZHAO, L. . On rates of convergence of efficient detection criteria in signal processing with white noise. **IEEE Transactions on Information Theory**, v. 35, n. 2, p. 380–388, 1989.

- BARBOSA, C. S.; FERNANDES, N. C. L.; FERREIRA, R. L. F. Crescimento e qualidade de mudas de cubiuzeiro (*solanum sessiliflorum* dunal) com uso de substratos em Rio Branco, Acre. **Scientia Naturalis**, v. 1, n. 5, 2019.
- BARNETT, V.; LEWIS, T. **Outliers in statistical data**. New York: John Wiley & Sons, 1974.
- BARTLETT, M. S. Properties of sufficiency and statistical tests. **Proceedings of the Royal Society of London. Series A**, v. 160, n. 901, p. 268–282, 1937.
- BATES, D. M.; WATTS, D. G. **Nonlinear regression analysis and its applications**. New York: John Wiley, 1988. 365 p.
- BEVILAQUA, G. et al. Indicações técnicas para produção de sementes de feijão para a agricultura familiar. **Embrapa Clima Temperado-Circular Técnica (INFOTECA-E)**, Pelotas: Embrapa Clima Temperado, 2013., 2013.
- BLAIR, R. C.; SAWILOWSKY, S. S.; HIGGINS, J. J. Limitations of the rank transform statistic in tests for interactions: Limitations of the rank transform statistic. **Communications in Statistics-Simulation and Computation**, v. 16, n. 4, p. 1133–1145, 1987.
- BOLFARINE, H.; SANDOVAL, M. C. **Introdução à inferência estatística**. São Paulo: SBM, 2001.
- BRANCO, M. D.; DEY, D. K. A general class of multivariate skew-elliptical distributions. **Journal of Multivariate Analysis**, v. 79, n. 1, p. 99–113, 2001.
- BUTTERFIELD, D. A. et al. Nutritional approaches to combat oxidative stress in alzheimer's disease. **The Journal of nutritional biochemistry**, v. 13, n. 8, p. 444–461, 2002.
- CANCHO, V. G.; LACHOS, V. H.; ORTEGA, E. M. A nonlinear regression model with skew-normal errors. **Statistical papers**, v. 51, n. 3, p. 547–558, 2010.
- CHEN, X.; ZENG, Q.; SONG, Q. Likelihood inference of nonlinear models based on a class of flexible skewed distributions. **Abstract and Applied Analysis**, v. 2014, p. 1–8, 2014.
- CONAB. **Companhia Nacional de Abastecimento**. 2020. Acesso em 07.05.2020. Disponível em: <<https://github.com/jdanielnd/gnlsmn/>>.
- CONOVER, W.; IMAN, R. L. On some alternative procedures using ranks for the analysis of experimental designs. **Communications in Statistics-Theory and Methods**, v. 5, n. 14, p. 1349–1368, 1976.
- CONTRERAS-REYES, J. E.; ARELLANO-VALLE, R. B.; CANALES, T. M. Comparing growth curves with asymmetric heavy-tailed errors: Application to the southern blue whiting (*micromesistius australis*). **Fisheries research**, v. 159, p. 88–94, 2014.
- COOK, R. D.; WEISBERG, S. **Residuals and influence in regression**. New York: Chapman and Hall, 1982.
- 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.

- DANTAS, R. d. A. et al. Produção de matéria seca e controle de plantas daninhas por leguminosas consorciadas com cana-de-açúcar em cultivo orgânico. **Pesquisa Agropecuária Brasileira**, v. 50, n. 8, p. 681–689, 2015.
- 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, 2004.
- DRAPER, N. R.; SMITH, H. **Applied regression analysis**. 3. ed. New York: John Wiley, 1998.
- DUARTE, J.; AOKI, R. **Graphic package for fitting non-linear models with scale mixture of skew-normal (NL-SMSN) distributions**. 2019. Acesso em 30.01.2020. Disponível em: <<https://github.com/jdanielnd/gnlsmsn/>>.
- DURBIN, J.; WATSON, G. S. Testing for serial correlation in least squares regression. **Biometrika**, v. 37, n. 3/4, p. 409–428, 1950.
- FAGERIA, N.; BALIGAR, V. Enhancing nitrogen use efficiency in crop plants. **Advances in agronomy**, v. 88, p. 97–185, 2005.
- FANG, K.; KOTZ, S.; NG, K. **Symmetric multivariate and related distributions**. London: Chapman & Hall, 1990. (Monographs on statistics and applied probability, 36). ISBN 0412314304. Disponível em: <http://gso.gbv.de/DB=2.1/CMD?ACT=SRCHA&SRT=YOP&IKT=1016&TRM=pfn+027504395&sourceid=fbw_bibsonomy>.
- FERNANDES, T. J. et al. Parameterization effects in nonlinear models to describe growth curves. **Acta Scientiarum. Technology**, v. 37, n. 4, p. 397–402, 2015.
- FERNANDES, T. J. et al. Seleção de modelos não lineares para a descrição das curvas de crescimento do fruto do cafeeiro. **Coffee Science**, v. 9, n. 1, 2014.
- FERREIRA, C. S.; LACHOS, V. H. Nonlinear regression models under skew scale mixtures of normal distributions. **Statistical Methodology**, v. 33, p. 131–146, 2016.
- FOX, J. **Regression diagnostics: An introduction**. California: SAGE Publishing, 1991. v. 79.
- FOX, T.; HINKLEY, D.; LARNTZ, K. Jackknifing in nonlinear regression. **Technometrics**, v. 22, n. 1, p. 29–33, 1980.
- FREITAS, A. R. d. et al. Técnicas de análises exploratórias em dados de cultivares de alfafa. **Revista Brasileira de Zootecnia**, v. 37, n. 9, p. 1531–1536, 2008.
- GALEA, M.; PAULA, G. A.; CYSNEIROS, F. J. A. On diagnostics in symmetrical nonlinear models. **Statistics & Probability Letters**, Elsevier, v. 73, n. 4, p. 459–467, 2005.
- GENTON, M. G. **Skew-elliptical distributions and their applications: a journey beyond normality**. New York: Chapman and Hall, 2004.
- GODOI, L. G. d. **A Distribuição t-assimétrica univariada : propriedade e inferência**. Dissertação (Mestrado) — São Paulo: Instituto de Matemática e Estatística, Universidade de São Paulo, 2007. Dissertação de Mestrado.
- GOMES-PIMENTEL, F. **Curso de estatística experimental**. São Paulo: FEALQ, 2009.

- GUEDES, T. A. et al. Applying regression models with skew-normal errors to the height of bedding plants of *stevia rebaudiana* (bert) bertonii. **Acta Scientiarum. Technology**, v. 36, n. 3, 2014.
- HOFFMAN, R.; VIEIRA, S. **Uma introdução á econometria**. São Paulo: HUCITEC, 1988. 379 p.
- 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.
- KERES, G. et al. Desempenho de feijoeiro mungo em densidades populacionais e espaçamento entre linhas no mato grosso. **Scientia Agraria Paranaensis**, v. 18, n. 3, p. 251, 2020.
- LABRA, F. V. et al. Estimation and diagnostics for heteroscedastic nonlinear regression models based on scale mixtures of skew-normal distributions. **Journal of Statistical Planning and Inference**, v. 142, n. 7, p. 2149–2165, 2012.
- LEYS, C.; SCHUMANN, S. A nonparametric method to analyze interactions: The adjusted rank transform test. **Journal of Experimental Social Psychology**, v. 46, n. 4, p. 684–688, 2010.
- LIMA, K. et al. Uso de modelos não lineares na descrição do acúmulo de boro em diferentes partes do feijoeiro cultivar jalo. **Revista Brasileira de Biometria**, v. 35, n. 4, p. 834–861, 2017.
- LIMA, K. P. et al. Modelagem não linear da biomassa seca do feijoeiro cv. jalo. **Sigmae**, v. 8, n. 2, p. 359–369, 2019.
- LOUZADA, F.; FERREIRA, P. H.; DINIZ, C. A. Skew-normal distribution for growth curve models in presence of a heteroscedasticity structure. **Journal of Applied Statistics**, v. 41, n. 8, p. 1785–1798, 2014.
- LUCIO, A. D.; NUNES, L. F.; REGO, F. Nonlinear regression and plot size to estimate green beans production. **Horticultura Brasileira**, v. 34, n. 4, p. 507–513, 2016.
- LUEPSEN, H. **Varianzanalysen-Prüfung der Voraussetzungen und Übersicht der nichtparametrischen Methoden sowie praktische Anwendungen mit R und SPSS**. 2020. Acesso em 05.05.2018. Disponível em: <<http://www.uni-koeln.de/~luepsen/statistik/texte/nonpar-anova.pdf>>.
- MAGHAMI, M. M.; BAHRAMI, M.; SAJADI, F. A. On bias reduction estimators of skew-normal and skew-t distributions. **Journal of Applied Statistics**, p. 1–23, 2020.
- MANGUEIRA, R. A. F. et al. O modelo logístico considerando diferentes distribuições para os erros aplicado a dados de altura do milho. **Revista Brasileira de Biometria**, v. 34, n. 2, p. 317–333, 2016.
- MANSOURI, H. Multifactor analysis of variance based on the aligned rank transform technique. **Computational statistics & data analysis**, v. 29, n. 2, p. 177–189, 1998.
- MAPA. **Ministério da Agricultura, Pecuária e Abastecimento - Feijão**. 2020. Acessado: 07-05-2020. Disponível em: <<http://www.agricultura.gov.br/vegetal/culturas/feijao/saiba-mais>>.

MARTIN, T.; STORCK, L. Análise das pressuposições do modelo matemático em experimentos agrícolas no delineamento blocos ao acaso. In: II SEMINÁRIO: SISTEMAS DE PRODUÇÃO AGROPECUÁRIA. **Anais...** Paraná: UTFPR, 2008. p. 177–196.

MARTINSFILHO, S. et al. Abordagem bayesiana das curvas de crescimento de duas cultivares de feijoeiro. **Ciência Rural**, v. 38, n. 6, p. 1516–1521, set 2008.

MATOSO, S. C.; KUSDRA, J. F. Nodulação e crescimento do feijoeiro em resposta à aplicação de molibdênio e inoculante rizobiano. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 18, n. 6, p. 567–573, 2014.

MONAHAN, J. F. **Numerical methods of statistics**. Reino Unido: Cambridge University Press, 2011.

MONTGOMERY, D. C.; PECK, E. A.; VINING, G. G. **Introduction to linear regression analysis**. New York: John Wiley & Sons, 2012.

MOURA, M. S. B. de et al. Modelos de crescimento para o feijão-caupi e o milho, sob sistemas de plantio exclusivo e consorciado, no semiárido brasileiro. **Revista Brasileira de Agrometeorologia**, v. 16, n. 3, p. 275–284, 2008.

MUIANGA, C. A. et al. Descrição da curva de crescimento de frutos do cajueiro por modelos não lineares. **Revista Brasileira de Fruticultura**, v. 38, n. 1, p. 22–32, 2016.

NELSON, L. S. The sum of values from a normal and a truncated normal distribution. **Technometrics**, v. 6, p. 469–471, 1964.

O'HAGAN, A.; LEONARD, T. Bayes estimation subject to uncertainty about parameter constraints. **Biometrika**, v. 63, n. 1, p. 201–203, 1976.

PETRUCCI, E.; OLIVEIRA, L. A. de. Coeficientes de assimetria e curtose nos dados de vazão média mensal da bacia do Rio Preto-BA. **Os Desafios da Geografia Física na Fronteira do Conhecimento**, v. 1, p. 158–170, 2017.

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**, v. 43, n. 5, p. 803–809, 2013.

R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2019. Disponível em: <<https://www.R-project.org/>>.

RATKOWSKY, D. A. **Nonlinear regression modeling: a unified practical approach. Statistics: textbooks and monographs**. New York: John Wiley, 1983.

RIAZOSHAMS, A. H. et al. On the outlier detection in nonlinear regression. **Engineering and Technology**, v. 60, p. 264–270, 2009.

RIBEIRO, T. D. et al. Description of the growth of pequi fruits by nonlinear models. **Revista Brasileira de Fruticultura**, v. 40, n. 4, 2018.

ROBERTS, C. A correlation model useful in the study of twins. **Journal of the American Statistical Association**, v. 61, n. 316, p. 1184–1190, 1966.

ROSSE, L. N.; VENCOVSKY, R. Modelo de regressão não-linear aplicado ao estudo da estabilidade fenotípica de genótipos de feijão no estado do paraná. **Bragantia**, v. 59, n. 1, 2000.

ROSSI, R. M.; SANTOS, L. A. D. Bayesian modeling growth curves for quail assuming skewness in errors. **Semina: Ciências Agrárias**, v. 35, n. 3, p. 1637–1648, 2014.

SALTER, K.; FAWCETT, R. The art test of interaction: a robust and powerful rank test of interaction in factorial models. **Communications in Statistics-Simulation and Computation**, v. 22, n. 1, p. 137–153, 1993.

SANTOS, R. A. dos et al. Remote sensing as a tool to determine biophysical parameters of irrigated seed corn crop. **Semina: Ciências Agrárias**, v. 41, n. 2, p. 435–446, 2020.

SAWILOWSKY, S. S. Nonparametric tests of interaction in experimental design. **Review of Educational Research**, v. 60, n. 1, p. 91–126, 1990.

SCHWARZ, G. et al. Estimating the dimension of a model. **The annals of statistics**, v. 6, n. 2, p. 461–464, 1978.

SCHWERTMAN, N. C.; OWENS, M. A.; ADNAN, R. A simple more general boxplot method for identifying outliers. **Computational statistics & data analysis**, v. 47, n. 1, p. 165–174, 2004.

SEBER, G.; WILD, C. **Nonlinear regression**. [S.l.]: New York: John Wiley, 1989.

SHAPHIRO, S.; WILK, M. An analysis of variance test for normality. **Biometrika**, v. 52, n. 3, p. 591–611, 1965.

SILVA, A.; LIMA, E.; MENEZES, H. Rendimento de grãos de feijão (*phaseolus vulgaris* L.), cultivado em diferentes densidades de plantio. **Revista das Faculdades Integradas de Bebedouro**, v. 3, p. 1–5, 2007.

SILVA, F. d. L. et al. Curvas de crescimento em vacas de corte de diferentes tipos biológicos. **Pesquisa Agropecuária Brasileira**, v. 46, n. 3, p. 262–271, 2011.

SILVA, L. M. et al. Comparison of the gompertz and verhulst models in the accumulation of calcium in common bean cultivar jalo. **International Journal of Engineering Inventions**, v. 7, p. 11–14, 2018.

SILVA, N. M. et al. Condicionadores alternativos de substrato na qualidade da muda e produtividade de couve manteiga. **Revista Verde de Agroecologia e Desenvolvimento Sustentável**, v. 11, n. 5, 2016.

SOUZA, A. B. et al. Densidades de semeadura e níveis de npk e calagem na produção do feijoeiro sob plantio convencional, em ponta grossa, paraná. **Pesquisa Agropecuária Tropical**, v. 38, n. 1, p. 39–43, 2008.

SOUZA, A. B. de; ANDRADE, M. J. B. de; ALVES, V. G. Populações de plantas, adubação e calagem para o feijoeiro (cv. iapar 81) em um gleissolo de ponta grossa, estado do paraná. **Acta Scientiarum. Agronomy**, v. 26, n. 3, p. 347–352, 2004.

STEEL, R. G.; TORRIE, J. H.; DICKEY, D. A. **Principles and procedures of statistics: A biological approach**. New York: McGraw-Hill College, 1997. 451 p.

STEVENS, J. P. Outliers and influential data points in regression analysis. **Psychological Bulletin**, v. 95, n. 2, p. 334, 1984.

STORCK, L. et al. **Experimentação vegetal**. Santa Maria: UFSM, 2000.

TEIXEIRA, I. R. et al. Resposta do feijoeiro (*phaseolus vulgaris* L. cv. pérola) a diferentes densidades de semeadura e doses de nitrogênio. **Ciência e Agrotecnologia**, v. 24, n. 2, p. 399–408, 2000.

TUKEY, J. W. A survey of sampling from contaminated distributions. **Contributions to probability and statistics**, p. 448–485, 1960.

UCHÔA, T. L. et al. Produtividade de almeirão orgânico produzido com substratos alternativos. **Revista Verde de Agroecologia e Desenvolvimento Sustentável**, v. 13, n. 2, p. 132–137, 2018.

URVOY, M.; AUTRUSSEAU, F. Application of grubbs' test for outliers to the detection of watermarks. In: 2ND ACM WORKSHOP ON INFORMATION HIDING AND MULTIMEDIA SECURITY. **Proceedings...** New York: ACM Press, 2014. p. 177–196.

VERHULST, P.-F. Notice sur la loi que la population suit dans son accroissement. **Corresp. Math. Phys.**, v. 10, p. 113–126, 1838.

VIEIRA, N. et al. Accumulation of macronutrients by different common bean cultivars grown in different plant densities in no-tillage crop system. **Annual Report of the Bean Improvement Cooperative**, v. 52, n. 1, p. 132–133, 2009.

VIEIRA, N. M. B. **Crescimento e marcha da absorção de nutrientes no feijoeiro cvs. BRS-MG Talismã e Ouro Negro, em plantio direto e convencional**. Tese (145 f.) — Curso de Doutorado em Agronomia, Departamento de Agricultura, Universidade Federal de Lavras, 2006. Dissertação (Mestrado).

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**, v. 32, n. 6, p. 1687–1693, 2008.

WANG, J.; GENTON, M. G. The multivariate skew-slash distribution. **Journal of Statistical Planning and Inference**, v. 136, n. 1, p. 209–220, 2006.

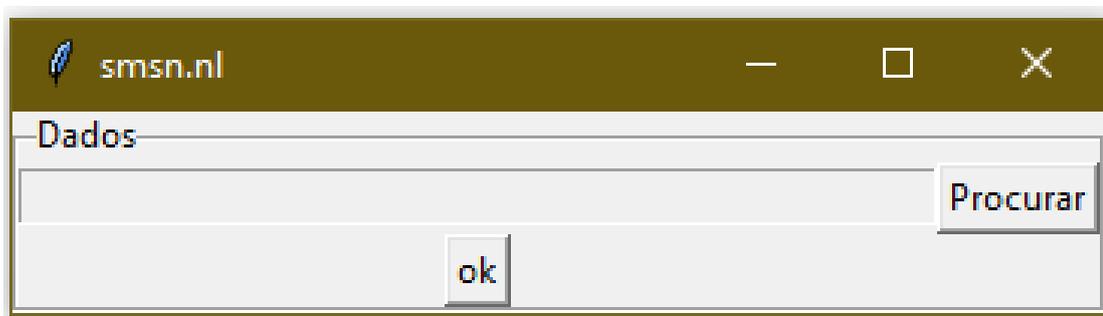
WANG, T.; LI, B.; GUPTA, A. K. Distribution of quadratic forms under skew normal settings. **Journal of Multivariate Analysis**, v. 100, n. 3, p. 533–545, 2009.

WINSOR, C. P. The gompertz curve as a growth curve. **Proceedings of the national academy of sciences**, v. 18, n. 1, p. 1–8, 1932.

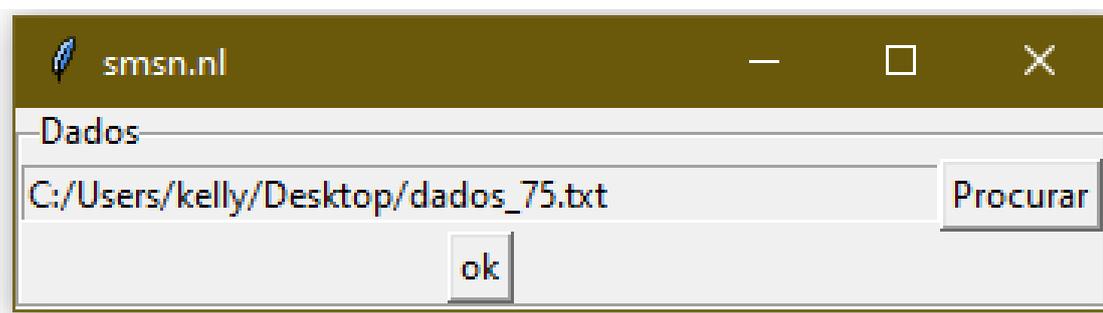
WOBBROCK, J. O. et al. The aligned rank transform for nonparametric factorial analyses using only anova procedures. **Computer Human Interaction- CHI**, p. 143–146, 2011.

APÊNDICE A – Interface Gráfica

Para solucionar o problema da falta de uma interface de interação entre a linguagem R e os usuários que estão fora do grupo de estatístico e programadores, e, portanto, não possuem conhecimento necessário para a utilização através de linhas de comandos foi desenvolvido uma interface gráfica. Para o uso da interface gráfica é necessário a instalação do R e/ou RStudio e futuramente ficará disponível para uso no github. Os pacotes necessário serão automaticamente instalado quando for usado a interface. A partir do comando *gsmsn.nl* () foi gerado uma interface gráfica amigável. É um tipo de interface, que permite a interação com dispositivos digitais por meio de elementos gráficos como ícones.



Para entrar com o arquivo, é necessário que ele esteja em extensão .txt (bloco de notas e a separação do decimal é ponto(.), será necessário salva no mesmo arquivo .txt uma coluna para variável resposta e uma para a explicativa. clica-se no botão Procurar e busca-se o arquivo em seu diretório.



Como pode ser visto, após o uso da função e inserção dos dados surgirá uma interface gráfica. Nessa interface tem-se a caixa e seletor para o usuário preencher. Então, gera-se uma janela com a distribuição dos erros que foram usados nesse trabalho.

Tem-se a caixa para selecionar a distribuição, a variável resposta em estudo e a variável explicativa. Além disso, pode-se modelar a variância quando existe problema com a homogeneidade residual, para isso usa-se os tipo de regressão, função e Z variables para selecionar a variável relacionada a essa variância. Os valores iniciais que são de conhecimento prévio do pesquisador. E modelos que foram programados o Logístico e Gompertz para desenvolvimento dessa tese.

The screenshot shows the 'smsn.nl' software interface with the following configuration options:

- Distribuição:** Radio buttons for Skew.t, Skew.normal, and Normal.
- Variável resposta:** A dropdown menu.
- Variável explicativa:** Checkboxes for 'tot' and 'dae'.
- Tipo de Regressão:** A dropdown menu set to 'Homoscedastic'.
- Função de variância:** A dropdown menu set to 'exp(z*rho)'. Below it, checkboxes for 'Z variables' 'tot' and 'dae'.
- Valores Iniciais:** Input fields for 'shape' (0), 'rho' (0), 'nu' (0), and 'betas' (c(1,1,1)).
- Modelos:** A large empty area for model selection.
- Crítério de Comparação:** A checkbox.
- Ok:** A button at the bottom.

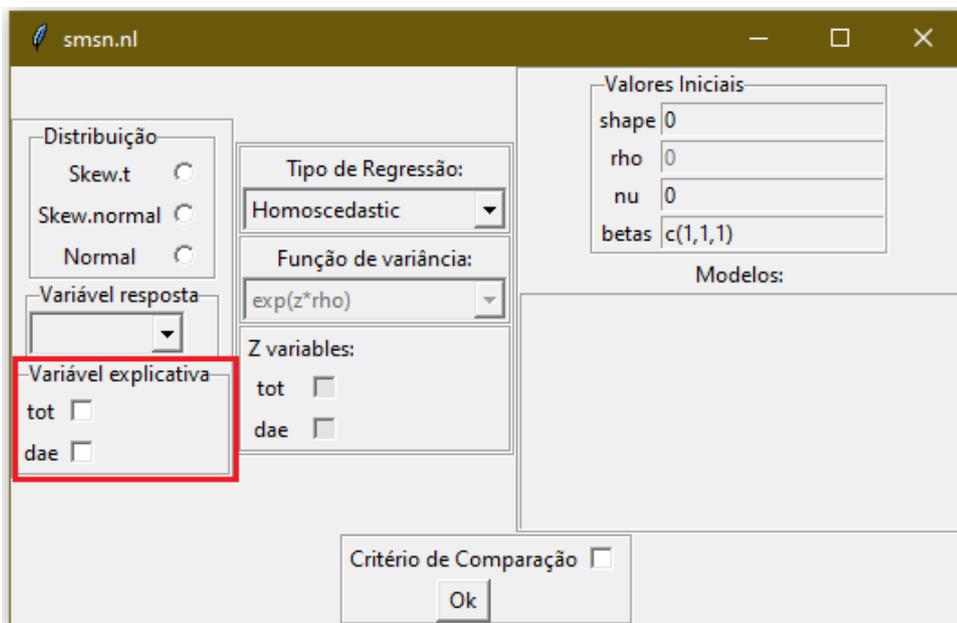
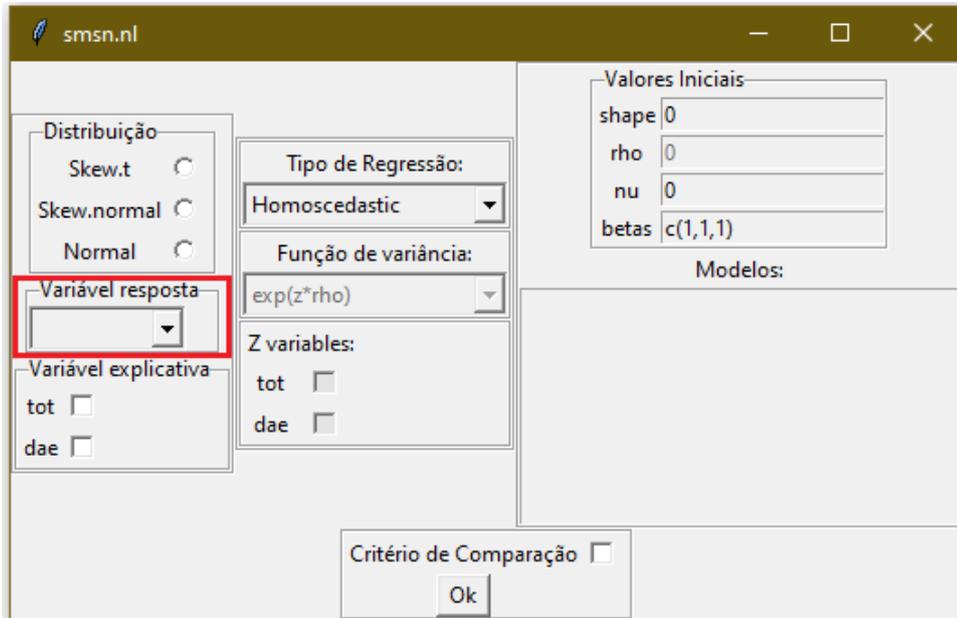
Na quadro tem-se as distribuições existentes, no qual, pode-se escolher as três possíveis distribuições para o erro associado ao modelo, que são t assimétrica (skew.t), normal assimétrica (skew.normal) e a Normal.

The screenshot shows the 'smsn.nl' software window with the following settings:

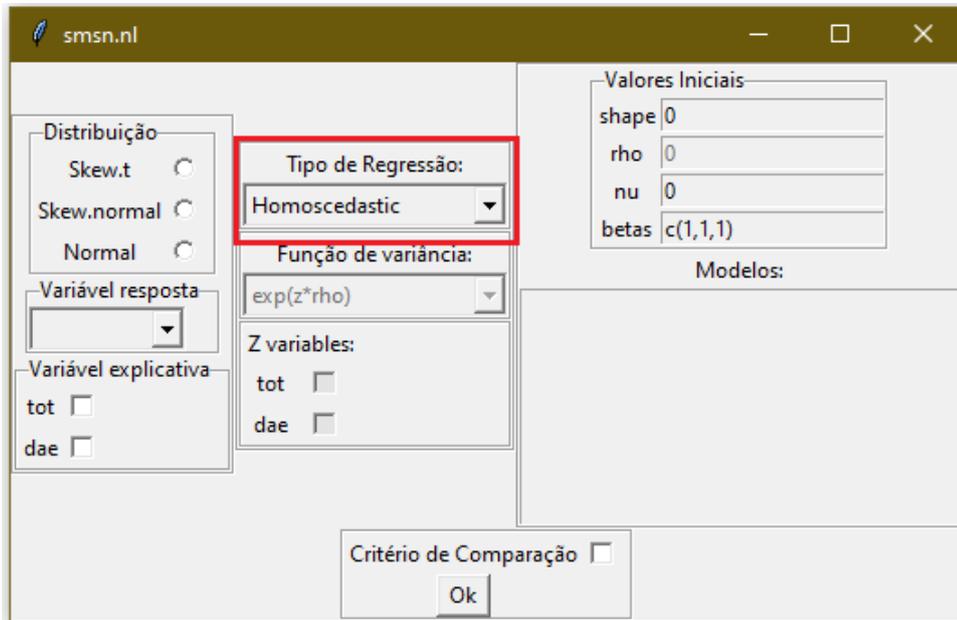
- Distribuição:** A red box highlights three radio button options: Skew.t, Skew.normal, and Normal.
- Tipo de Regressão:** A dropdown menu set to 'Homoscedastic'.
- Função de variância:** A dropdown menu set to 'exp(z*rho)'. Below it, 'Z variables' are listed with checkboxes for 'tot' and 'dae', both of which are currently unchecked.
- Valores Iniciais:** A table with the following values:

shape	0
rho	0
nu	0
betas	c(1,1,1)
- Modelos:** An empty text area.
- Other settings:** 'Variável resposta' is empty, 'Variável explicativa' has checkboxes for 'tot' and 'dae' (both unchecked), and 'Critério de Comparação' is unchecked.
- Buttons:** An 'Ok' button is located at the bottom center.

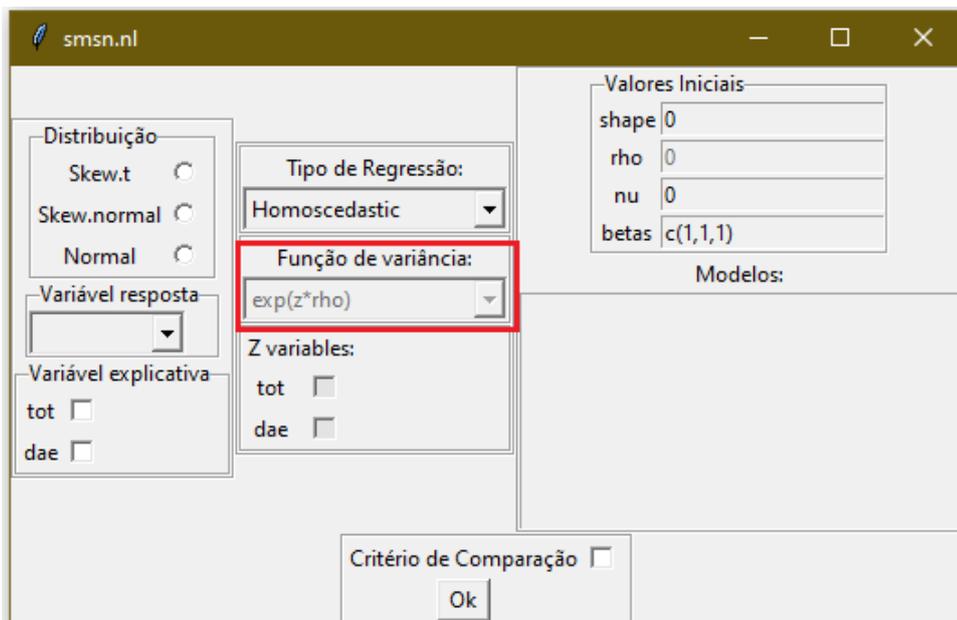
Já na variável resposta seleciona-se na caixa a variável de resposta do seu estudo. Na caixa da variável explicativa seleciona-se com um clique na caixa em branco.



No tipo de regressão pode-se encontrar homocedástico (Homoscedastic) ou heterocedástico (Heteroscedastic).



Esse tipo de regressão está associado a modelagem da variância. No caso, usa-se quando existe problema com a pressuposição de homogeneidade de variância residual. Além disso, pode-se escolher a função de variância.



A variável z que é selecionada de acordo com heterogeneidade.

The screenshot shows the 'smsn.nl' window with the following settings:

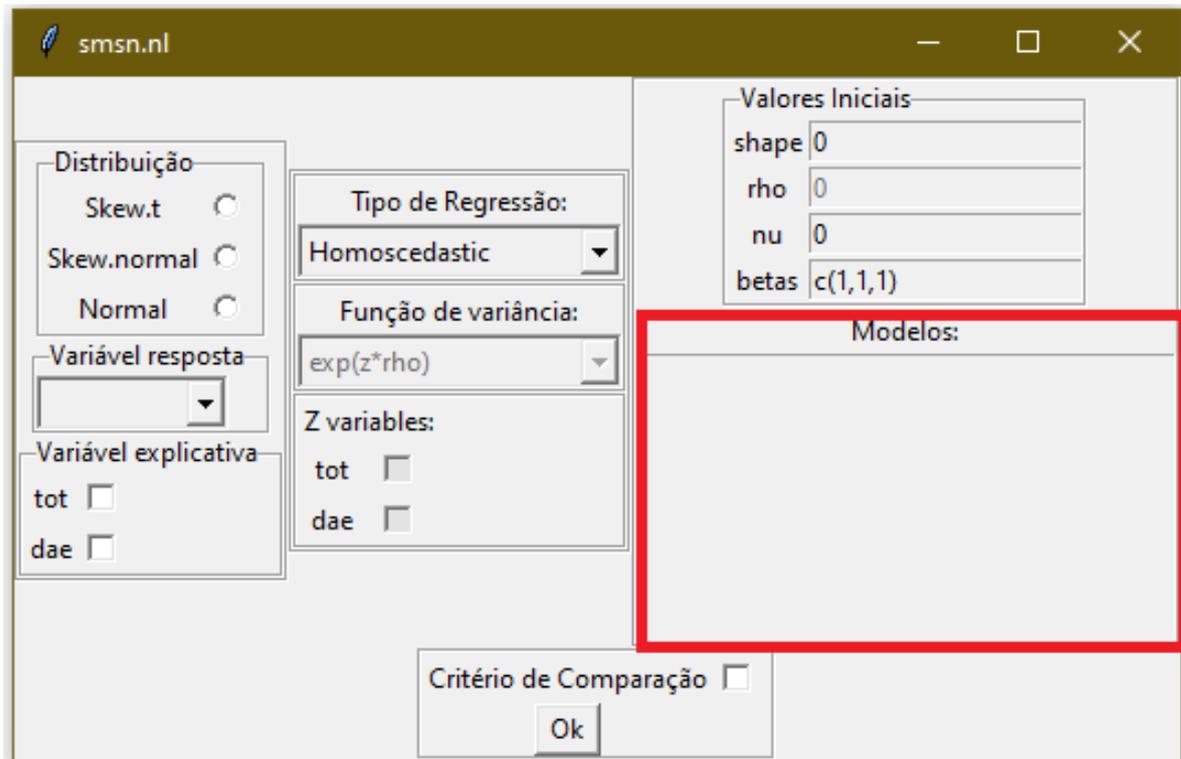
- Distribuição:** Skew.t, Skew.normal, Normal (radio buttons)
- Tipo de Regressão:** Homoscedastic (dropdown)
- Função de variância:** $\exp(z^*rho)$ (dropdown)
- Valores Iniciais:** shape: 0, rho: 0, nu: 0, betas: c(1,1,1)
- Z variables:** tot , dae (highlighted with a red box)
- Modelos:** (empty list)
- Crítério de Comparação:**
- Ok** button

Nos valores iniciais tem-se os valores para shape (assimetria), rho, nu (grau de liberdade da distribuição skew t) e os betas que estão associados aos modelos.

The screenshot shows the 'smsn.nl' window with the following settings:

- Distribuição:** Skew.t, Skew.normal, Normal (radio buttons)
- Tipo de Regressão:** Homoscedastic (dropdown)
- Função de variância:** $\exp(z^*rho)$ (dropdown)
- Valores Iniciais:** shape: 0, rho: 0, nu: 0, betas: c(1,1,1) (highlighted with a red box)
- Z variables:** tot , dae
- Modelos:** (empty list)
- Crítério de Comparação:**
- Ok** button

Na caixa de Modelos tem-se 1 para modelo Logístico e 2 para modelo Gompertz.



O Modelo 1 é representado pelo seguinte modelo:

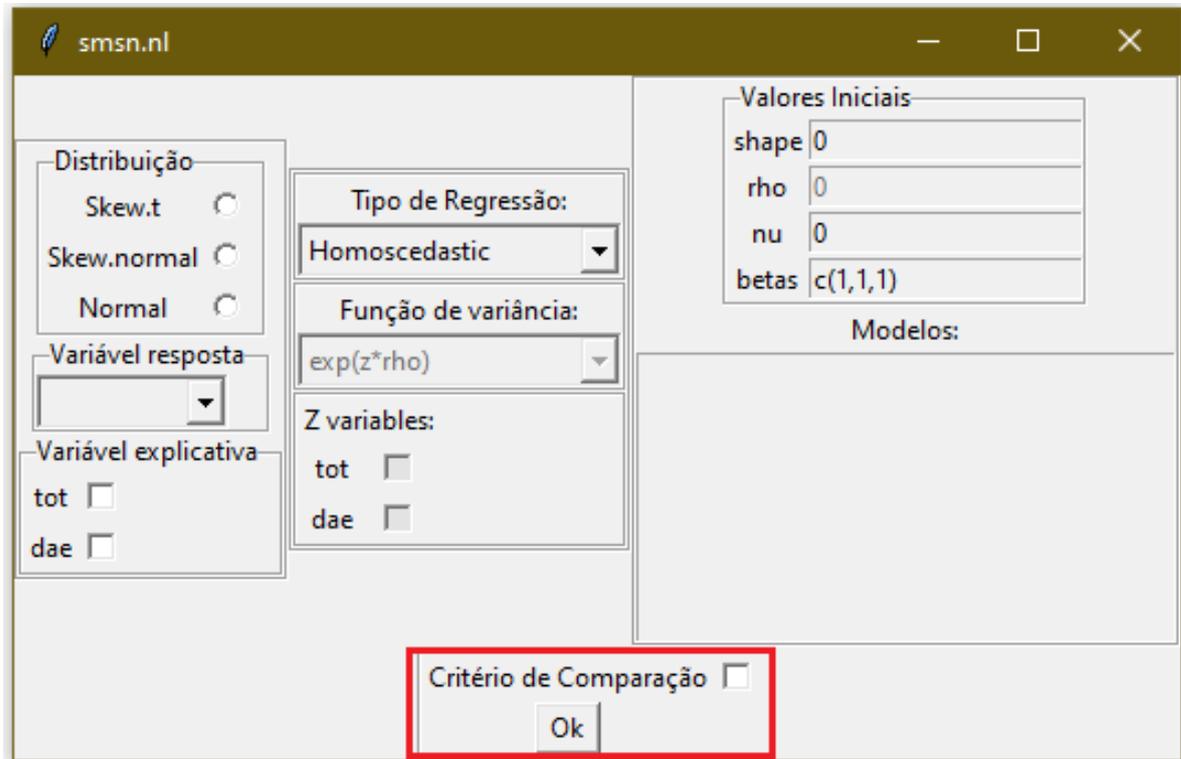
$$Y_i = \frac{a}{1 + e^{k(b-x_i)}} + \varepsilon_i$$

O Modelo 2 é dado por:

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

em que $\text{betas} = \{a, b, k\}$

No critério de comparação, pode-se ser selecionado a caixa em branco e finalizando a análise, clica-se no ok.



No caso, o critério de comparação é o AIC, BIC, EDC e loglik (log verossimilhança).

Um exemplo do resultado da análise pode ser vista na Figura 1 quando não se coloca os critérios de comparação.

Figura 1 – Resultado para análise com distribuição normal

```
> gsmsn.nl()
$betas
[1] 737,03078 43,49255 0,12231

$sigma2
[1] 49739

$shape
[1] 0

$iter
[1] 3

$res
[1] 0,094887 0,047627 0,103227 0,114963 0,139983 0,190023 -0,380706
[8] 0,061314 0,269813 -0,915958 0,571340 0,321141 -0,808012 0,242827
[15] 0,273407 -1,616851 0,954647 -0,118432 -0,256919 2,000439 2,903938
[22] -2,097117 -0,851679 -0,684879

$n
[1] 24

attr(,"class")
[1] "Normal"
```

Em que betas representa a estimativa do modelo Logístico em estudo. Os resultados considerando uma distribuição normal são iguais aos obtidos pela função nls, frequentemente usada para os modelos não lineares.

Figura 2 – Resultado para análise com distribuição normal considerando os critérios de comparação

```
> gsmsn.nl()
$betas
[1] 737,03078 43,49255 0,12231

$sigma2
[1] 49739

$shape
[1] 0

$loglik
[1] -163,83

$AIC
[1] 335,66

$BIC
[1] 340,37

$EDC
[1] 331,58

$iter
[1] 3

$res
[1] 0,094887 0,047627 0,103227 0,114963 0,139983 0,190023 -0,380706
[8] 0,061314 0,269813 -0,915958 0,571340 0,321141 -0,808012 0,242827
[15] 0,273407 -1,616851 0,954647 -0,118432 -0,256919 2,000439 2,903938
[22] -2,097117 -0,851679 -0,684879

$n
[1] 24

attr(,"class")
[1] "Normal"
```