

LEANDRO ROBERTO DE MACEDO

ABORDAGENS FREQUENTISTA E BAYESIANA PARA DESCRIÇÃO DAS
CURVAS DE ACÚMULO DE MATÉRIA SECA DE PLANTAS DE ALHO

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Estatística Aplicada e Biometria, para obtenção do título de Doctor Scientiae.

VIÇOSA
MINAS GERAIS - BRASIL
2015

Ficha catalográfica preparada pela Biblioteca Central da Universidade
Federal de Viçosa - Câmpus Viçosa

T

M141a Macedo, Leandro, 1983-
2015 Abordagens frequentista e bayesiana para descrição das
curvas de acúmulo de matéria seca de plantas de alho / Leandro
Macedo. – Viçosa, MG, 2015.
xi, 55f. : il. ; 29 cm.

Inclui apêndice.

Orientador: Paulo Roberto Cecon.

Tese (doutorado) - Universidade Federal de Viçosa.

Inclui bibliografia.

1. Análise de regressão. 2. Distribuição (Teoria da
probabilidade). 3. Análise por agrupamento. 4. Teoria bayesiana
de decisão estatística. 5. Alho. I. Universidade Federal de
Viçosa. Departamento de Estatística Aplicada e Biometria.
Programa de Pós-graduação em Estatística Aplicada e Biometria.
II. Título.

CDD 22.ed. 519.536

LEANDRO ROBERTO DE MACEDO

ABORDAGENS FREQUENTISTA E BAYESIANA PARA DESCRIÇÃO DAS
CURVAS DE ACÚMULO DE MATÉRIA SECA DE PLANTAS DE ALHO

Tese apresentada à Universidade Federal
de Viçosa, como parte das exigências do
Programa de Pós-Graduação em
Estatística Aplicada e Biometria, para
obtenção do título de *Doctor Scientiae*.

APROVADA: 3 de dezembro de 2015



Mario Puiatti



Willerson Custódio da Silva



Fabyano Fonseca e Silva
(Coorientador)



Moysés Nascimento
(Coorientador)



Paulo Roberto Cecon
(Orientador)

Aos meus pais, Tereza e Pedro
À minha esposa Jussara
dedico.

AGRADECIMENTOS

Primeiramente a Deus por ter me dado forças e perseverança ao longo dessa jornada.

À Universidade Federal de Viçosa, pela oportunidade de realizar a graduação, mestrado e este curso de doutorado.

À CAPES, pela concessão da bolsa de estudos.

À FAPEMIG, pelo financiamento do projeto de pesquisa.

Ao Departamento de Estatística Aplicada e Biometria pela oportunidade em realizar este curso.

Ao professor orientador e amigo, Paulo Roberto Cecon, pelas valiosas orientações, pela confiança e pela amizade.

Aos coorientadores e amigos, professor Fabyano Fonseca e Silva e professor Moysés Nascimento pelos ensinamentos, sugestões e contribuições para a realização deste trabalho.

Aos professores Mario Puiatti e Willerson, membros da banca, pelas sugestões para este trabalho.

Aos professores do Departamento de Estatística Aplicada e Biometria pelos conselhos e conhecimentos transmitidos durante o curso.

À Carla e a Anita pela prontidão em ajudar sempre que precisei.

À minha esposa Jussara que sempre acreditou em mim, e esteve ao meu lado me dando forças para enfrentar os desafios que a vida nos oferece.

À minha mãe Tereza, ao meu pai Pedro e meus irmãos José, Maria, Neuza e Valdeci que com afeto, amor e dedicação estiveram sempre ao meu lado torcendo por mim.

Aos amigos do doutorado pelo companheirismo, pelos momentos de estudo, descontração e pelo apoio.

Aos meus amigos do Departamento de Economia da Universidade Federal de Juiz de Fora, Campus Governador Valadares.

A todos aqueles que direta ou indiretamente contribuíram para a realização deste trabalho, muito obrigado.

BIOGRAFIA

LEANDRO ROBERTO DE MACEDO, filho de Tereza Joana Macedo e Pedro Silvério de Macedo, nasceu em Viçosa, Minas Gerais, em 27 de julho de 1983.

Em maio de 2006, ingressou no curso de Matemática na Universidade Federal de Viçosa, graduando-se em julho de 2011.

Em fevereiro de 2012, iniciou no Programa de Pós-graduação, a nível de Mestrado em Estatística Aplicada e Biometria, na Universidade Federal de Viçosa, defendendo a dissertação em 31 de julho de 2013.

Em agosto de 2013, iniciou o curso de Doutorado em Estatística Aplicada e Biometria, na Universidade Federal de Viçosa, submetendo-se à defesa em 3 de dezembro de 2015.

Em 21 de março de 2013 foi aprovado no concurso para professor do ensino básico, na Rede Estadual de Educação de Minas Gerais, onde trabalhou até junho de 2014.

Em junho de 2014 foi aprovado em concurso público para professor do magistério superior na Universidade Federal de Juiz de Fora, Campus Governador Valadares, no Departamento de Economia.

No dia 3 de dezembro de 2015 defendeu a tese de Doutorado pelo Programa de Pós-graduação em Estatística Aplicada e Biometria, na Universidade Federal de Viçosa.

SUMÁRIO

LISTA DE FIGURAS.....	vii
LISTA DE TABELAS.....	viii
RESUMO.....	x
ABSTRACT.....	xi
1 INTRODUÇÃO.....	1
2 REFERENCIAL TEÓRICO.....	2
2.1 Modelos de regressão não linear.....	2
2.2 Abordagem Bayesiana.....	4
2.2.1 Métodos de Monte Carlo via Cadeias de Markov (MCMC).....	6
2.2.2 Algoritmo de Metropolis-Hastings.....	6
2.2.3 Amostrador de Gibbs.....	7
2.2.4 Diagnósticos de convergência.....	8
2.3 Abordagem Frequentista.....	9
2.3.1 Método dos quadrados mínimos ordinários.....	9
2.3.2 Processos iterativos e método dos quadrados mínimos.....	11
2.4 Avaliadores da qualidade de ajuste.....	13
2.4.1 Percentual de convergência (C%).....	13
2.4.2 Quadrado médio do erro (QME).....	13
2.4.3 Coeficiente de determinação (R^2).....	14
2.4.4 Critério de informação de Akaike (AIC).....	14
2.4.5 Critério de Informação Bayesiano (BIC).....	14
2.4.6 Desvio médio absoluto (DMA).....	15
2.4.7 Deviance Information Criterion (DIC).....	15
2.5 Análise de agrupamento.....	16
2.6 Medidas de dissimilaridades.....	16
2.7 Determinação do número ótimo de grupos.....	17

3	MATERIAL E MÉTODOS	18
3.1	Descrição do experimento.....	18
3.2	Abordagem Bayesiana	20
3.3	Abordagem Frequentista	21
3.4	Avaliadores da qualidade de ajuste	24
3.5	Análise de agrupamento	24
3.6	Dissimilaridade entre os acessos	24
3.7	Determinação do número ótimo de grupos	25
3.8	Identidade de modelos.....	25
4	RESULTADOS E DISCUSSÃO.....	27
5	CONCLUSÕES	44
6	REFERÊNCIAS BIBLIOGRÁFICAS	45
	APÊNDICE.....	49
A-	Script das análises no software R	50
B-	Script das análises de agrupamento	54
C-	Figura 1A- Gráfico da densidade a posteriori para as estimativas $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$ e $\hat{\sigma}^2$ para os parâmetros do modelo Logístico.....	55

LISTA DE FIGURAS

Figura 1. Dendrograma obtido com o agrupamento das estimativas dos parâmetros β_1 e β_3 do modelo Logístico para a abordagem Frequentista com os 30 acessos de alho avaliados.....	32
Figura 2. Curvas de acúmulo de matéria seca total de planta do alho ajustada pelo modelo Logístico para a abordagem Frequentista para cada um dos dois grupos considerados.....	36
Figura 3. Dendrograma obtido com o agrupamento das estimativas do parâmetro β_1 e β_3 do modelo Logístico para a abordagem Bayesiana com os 30 acessos de alho avaliados.....	41
Figura 4. Curvas de acúmulo de matéria seca total de planta do alho ajustada pelo modelo Logístico para a abordagem Bayesiana para cada um dos três grupos considerados.....	44

LISTA DE TABELAS

Tabela 1. Modelos de regressão não linear para descrever curvas de crescimento.....	3
Tabela 2. Relação dos 30 acessos de alho registrados no Banco de Germoplasma de Hortaliças da Universidade Federal de Viçosa (BGH/UFV).....	19
Tabela 3. Estimativas dos parâmetros e soma de quadrados residuais nas iterações do algoritmo de Gauss-Newton para ajuste do modelo Logístico.....	24
Tabela 4. Médias do coeficiente de determinação (R^2), quadrado médio do resíduo (QMR), critério de informação de Akaike (AIC), critério de informação Bayesiano (BIC) e desvio médio absoluto dos resíduos (DMA) dos modelos apresentados na Tabela 1, para as médias da matéria seca total da planta (MSTP) dos 30 acessos avaliados, e a porcentagem de convergência de cada modelo.....	28
Tabela 5. Média, erro padrão e coeficiente de variação das estimativas $\hat{\beta}_1$, $\hat{\beta}_2$ e $\hat{\beta}_3$ dos parâmetros para os modelos apresentados na Tabela 1.....	30
Tabela 6. Estimativas dos parâmetros β_1 , β_2 e β_3 para o modelo Logístico para cada um dos 30 acessos considerados.....	31
Tabela 7. Estimativas dos parâmetros β_1 , β_2 e β_3 para o modelo Logístico para cada um dos três grupos de acessos.....	32
Tabela 8. Estimativas dos parâmetros do modelo completo (Ω) e dos modelos reduzidos (ω_1 , ω_2 e ω_3).....	33
Tabela 9. Hipóteses avaliadas, valores da estatística do teste F, número de graus de liberdade e nível descritivo do teste (p-valor) para a estimativa $\hat{\beta}_1$	34
Tabela 10. Hipóteses avaliadas, valores da estatística do teste F, número de graus de liberdade e nível descritivo do teste (p-valor).....	34
Tabela 11. Estimativas dos parâmetros do modelo completo (Ω) e do modelo reduzido (ω) para os grupos I e II.....	34
Tabela 12. Critérios de Geweke (p-valor) e Raftery & Lewis (fator de dependência-FD).....	37

Tabela 13. Valores do DIC para cada um dos sete modelos apresentados na Tabela 1.....	38
Tabela 14. Média, desvio padrão, coeficiente de variação, Erro de Monte Carlo e intervalo de máxima densidade a posteriori (HPD) para as estimativas dos parâmetros dos modelos ajustado aos 30 acessos de alho avaliados.....	39
Tabela 15. Estimativas dos parâmetros β_1 , β_2 e β_3 para o modelo Logístico para cada um dos 30 acessos avaliados.....	40
Tabela 16. Estimativas dos parâmetros β_1 , β_2 e β_3 para o modelo Logístico para cada um dos quatro grupos de acessos.....	41
Tabela 17. Média, desvio padrão e intervalo de máxima densidade a posteriori (HPD) para as diferenças entre as estimativas dos parâmetros β_1 e β_3 para cada grupo.....	42

RESUMO

MACEDO, Leandro Roberto de, D.Sc., Universidade Federal de Viçosa, dezembro de 2015. **Abordagens Frequentista e Bayesiana para descrição das curvas de acúmulo de matéria seca de plantas de alho.** Orientador: Paulo Roberto Cecon. Coorientadores: Fabyano Fonseca e Silva e Moysés Nascimento.

Este trabalho teve como objetivo identificar modelos de regressão não linear que melhor descrevem as curvas de acúmulo de matéria seca em acessos de alho ao longo do tempo (60, 90, 120 e 150 dias após o plantio) utilizando as abordagens Frequentista e Bayesiana. Objetivou-se também agrupar os acessos similares em cada abordagem com relação às estimativas dos parâmetros e validar este agrupamento via inferência para a igualdade desses parâmetros entre os grupos formados. Para tal estudo foram utilizados 30 acessos de alho registrados no Banco de Germoplasma de Hortaliças da Universidade Federal de Viçosa (BGH/UFV). Os modelos Logístico, Gompertz e Von Bertalanffy mostraram-se bons representantes para este tipo de estudo, sendo o modelo Logístico o que melhor se ajustou aos dados. Após a escolha do melhor modelo em cada uma das abordagens, as estimativas dos parâmetros das curvas provenientes do ajuste deste modelo foram submetidas a análise de agrupamento, em que as estimativas foram consideradas como variáveis. Para o agrupamento foi utilizado o algoritmo de Ward e a distância generalizada de Mahalanobis como medida de proximidade. O número ótimo de grupos, segundo o método de Mojena, foi de três para a abordagem Frequentista e quatro para a Bayesiana. A inferência sobre igualdade de parâmetros das curvas entre os grupos formados indicou que o método Bayesiano mostrou-se eficiente e caracterizou-se como uma ferramenta útil para o estudo das curvas de acúmulo de matéria seca em plantas de alho visto que não apresentou problemas de convergência e reportou estimativas com baixos desvios padrão a posteriori, além de determinar de forma mais efetiva o número de grupos.

ABSTRACT

MACEDO, Leandro Roberto de, D.Sc., Universidade Federal de Viçosa, December, 2015. **Frequentist and Bayesian approaches for description of the accumulation curves of dry garlic plants.** Adviser: Paulo Roberto Cecon. Co-advisers: Fabyano Fonseca e Silva and Moysés Nascimento.

This thesis aimed to identify nonlinear regression models that best describe dry matter accumulation curves in garlic accessions over time (60, 90, 120, and 150 days after planting). When doing so, frequentist and Bayesian technics of estimation were analyzed. It was also intended to cluster similar garlic accessions according to their estimated parameters in each estimation approach, and to validate such clustering by means of tests for the equality of parameters. Dataset comprised 30 garlic accessions belonging to the Vegetable Germplasm Bank of Universidade Federal de Viçosa (BGH/UFV). Our results showed that Logistic, Gompertz, and Von Bertalanffy models are well-suited for studies in this research area, while the Logistic model presented the best goodness-of-fit indicators in both approaches. Next, we applied Ward's clustering algorithm, with Mahalanobis' generalized distances, in order to group Logistic curve estimated parameters for each garlic accessions. The optimal number of groups, according to Mojenas' method, was three for the frequentist method, and four, when considering the Bayesian method. Finally, we were able to conclude that the Bayesian technic of estimation is well-suited for studies related to this one, since it has not presented convergence problems, has reported estimates with lower posterior standard deviations, and has discriminated in the most effective way the groups of garlic plants.

1 INTRODUÇÃO

A espécie *Allium sativum* L., conhecida popularmente como alho, é a quarta hortaliça mais importante do Brasil, sendo cultivada em grande parte das regiões brasileiras e muito usada como condimento no preparo das refeições, devido ao aroma e sabor que confere aos alimentos (Mota et al., 2006; Lucini, 2008).

O alho destaca-se quanto aos valores culinários, medicinais, econômicos e nutricionais, apresentando as seguintes propriedades medicinais: analgésico, anti-inflamatório, antisséptico, antibacteriano, antimicótico, antiviral, diurético, antioxidante, estimulante do sistema imunológico, controle da hipertensão, reduz o colesterol, vermífugo (contra amebas), e anticoagulante (Trani, 2009).

A planta exige baixas temperaturas para que ocorra a formação do bulbo, e fotoperíodos longos aceleram o processo. Geralmente, as temperaturas médias de 12,8 a 23,9°C favorecem o desenvolvimento normal do alho, mas os cultivares apresentam respostas diferentes quanto à temperatura e fotoperíodo, com variação de resposta destes às regiões de cultivo e épocas de plantio (Mueller et al., 1990).

O estudo da trajetória das curvas de crescimento de plantas é de grande importância para a realização de manejos mais adequados, onde as estimativas dos parâmetros das curvas provenientes de modelos de regressão não linear podem ser interpretadas como variáveis biológicas tais como peso assintótico e velocidade de crescimento e utilizadas principalmente na identificação de plantas com as qualidades mais desejadas. Segundo Pôrto et al. (2007), curvas de acúmulo de matéria seca e de nutrientes servem como base para a indicação da demanda e dos cuidados em cada etapa do desenvolvimento da planta.

Os modelos de regressão não linear têm se mostrado adequados para descrever estas curvas, tanto pelo enfoque Frequentista quanto Bayesiano, pois apresentam parâmetros que podem ser interpretados biologicamente, tais como peso assintótico e velocidade de crescimento (Filho et al. 2008). Em recentes trabalhos envolvendo ajustes de modelos de regressão não linear a acessos de alho, Puiatti et al. (2013) e Reis et al. (2014), encontraram ótimos ajustes para os modelos Logístico, Gompertz e Von Bertalanffy utilizando a abordagem Frequentista, porém estudos envolvendo ajustes de modelos de regressão não linear a acessos de alho via abordagem Bayesiana ainda são escassos.

Diante dos diferentes modelos de regressão não linear presentes na literatura, a escolha do modelo mais apropriado para descrever o comportamento das curvas de acúmulo de matéria seca em plantas de alho pode ser feita mediante informações

provenientes de avaliadores de qualidade de ajuste, os quais permitem indicar estatisticamente o melhor modelo.

Geralmente, em estudos de curvas de crescimento, o pesquisador tem interesse em comparar as estimativas dos parâmetros das curvas entre as diferentes populações, a fim de indicar para qual delas o processo de crescimento foi mais eficiente (Silveira et al., 2009). Uma maneira de se obter grupos homogêneos, de forma a reunir os indivíduos em um determinado número de grupos tais que exista grande homogeneidade dentro de cada grupo e heterogeneidade entre eles é através da técnica da análise de agrupamento (Johnson; Wichern, 1992). Assim, pode-se agrupar as variedades de alho de acordo com as características desejadas, como por exemplo taxa de crescimento e acúmulo de matéria seca e, posteriormente, validar este agrupamento via inferência para a igualdade de parâmetros entre os grupos formados.

Diante do exposto, este trabalho teve como objetivo identificar modelos de regressão não linear que melhor descrevem o acúmulo de matéria seca em acessos de alho, utilizando as abordagens Frequentista e Bayesiana e, também agrupar os acessos mais semelhantes em cada abordagem com relação às estimativas dos parâmetros dos modelos.

2 REFERENCIAL TEÓRICO

2.1 Modelos de regressão não linear

O estudo das curvas de crescimento de espécies vegetais, geralmente, é feito através de modelos de regressão não linear, os quais conseguem resumir em um pequeno conjunto de parâmetros com interpretação biológica, uma série de informações sobre todo o período de desenvolvimento da planta, (Filho et al., 2008). Um modelo é dito não linear se pelo menos uma das derivadas parciais da função esperança em relação ao parâmetro é função de parâmetros desconhecidos (Prudente, 2009).

Segundo Souza (1998), os modelos de regressão não linear podem ser escritos da seguinte forma:

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

em que X_i representa a i -ésima observação da variável independente x ; Y_i representa a i -ésima observação da variável dependente y , $f(x_i, \boldsymbol{\theta})$ é a função resposta; $\boldsymbol{\theta} = [\theta_1 \ \dots \ \theta_p]^t$ é um vetor de parâmetros p dimensional desconhecido e ε_i

representa o efeito do erro aleatório, supondo erros independentes com média zero e variância constante, tal que $\varepsilon_i \sim N(0, \sigma^2)$.

Os modelos de regressão não linear aplicados às curvas de crescimento têm por objetivo descrever uma trajetória assintótica da variável dependente peso, em função da variável independente tempo. Dentre os modelos mais utilizados para descrever curvas de crescimento de plantas ou de animais estão os modelos apresentados na Tabela 1.

Nesses modelos (Tabela 1), o parâmetro β_1 representa a estimativa do peso assintótico; o parâmetro β_2 não tem interpretação biológica (é uma constante matemática cuja função é ajustar o peso inicial em relação à origem) e o parâmetro β_3 representa a velocidade de crescimento.

Tabela 1. Modelos de regressão não linear para descrever curvas de crescimento

Modelos de Curvas de Crescimento	Referência
$y_i = \beta_1 e^{(-\beta_2 e^{-\beta_3 x_i})} + \varepsilon_i$	Gompertz (Gompertz, 1825)
$y_i = \frac{\beta_1}{1 + \beta_2 e^{(-\beta_3 x_i)}} + \varepsilon_i$	Logístico (Ratkowski, 1983)
$y_i = \beta_1 - \beta_2 e^{(-\beta_3 x_i)} + \varepsilon_i$	Meloun I (Meloun e Militky, 1996)
$y_i = \beta_1 - e^{(-\beta_2 - \beta_3 x_i)} + \varepsilon_i$	Meloun II (Meloun e Militky, 1996)
$y_i = \beta_1 (1 - \beta_2 e^{(-\beta_3 x_i)}) + \varepsilon_i$	Brody (1945)
$y_i = \beta_1 (1 - \beta_2 e^{(-\beta_3 x_i)})^3 + \varepsilon_i$	Von Bertalanffy (1957)
$y_i = \beta_1 (1 - e^{(\beta_3 \beta_2 - \beta_3 x_i)}) + \varepsilon_i$	Mitscherlich (1919)

Em recentes trabalhos, como os realizados por Puiatti et al. (2013) e Reis et al. (2014), ao ajustarem modelos de regressão não linear a acessos de alho, dentre eles os modelos Logístico, Gompertz e Von Bertalanffy, verificaram que todos estes se ajustaram bem aos dados, sendo o modelo Logístico o de melhor ajuste.

Segundo Reis (2014), os modelos mais usados para descrição de crescimento de plantas e animais são: Brody, Gompertz, Logístico, Richards, Mitscherlich, Weibull e

Von Bertalanffy. Maia et al. (2009) utilizaram os modelos de regressão não linear Brody, Gompertz, Logístico e Von Bertalanffy em estudo envolvendo bananeiras, encontrando ótimos ajustes.

Geralmente, a estimação dos parâmetros dos modelos de regressão não linear para descrição de curvas de crescimento tem sido realizada por meio de uma abordagem Frequentista. A estimação é fundamentada em processos iterativos, cujo procedimento visa a minimização da soma de quadrados dos resíduos. Porém, segundo Filho et al. (2008), quando se trata de ajustes individuais, os métodos iterativos, muitas vezes, produzem estimativas irrealistas para os parâmetros. Além disso, quando se trata de comparações de curvas provenientes de tratamentos diferentes, as distribuições dos estimadores dos parâmetros dos modelos de regressão não linear não seguem distribuições gaussianas. Assim, segundo Silva et al. (2005), torna-se complexo o processo de formulação de testes estatísticos, pois não serão atendidas pressuposições relacionadas à utilização da teoria assintótica.

Em alguns estudos envolvendo ajuste de modelos de regressão não linear (Blasco et al., 2003; Silva et al., 2005), a inferência Bayesiana foi utilizada com sucesso, pois reduziu o número de estimativas viesadas, mesmo na presença de poucas informações.

2.2 Abordagem Bayesiana

Um dos principais objetivos da estatística é realizar inferência sobre os parâmetros de um modelo. Na abordagem Frequentista, os parâmetros desconhecidos são considerados fixos e toda a análise é baseada nas informações contidas na amostra dos dados. Segundo Paulino et al. (2003), esta abordagem foi adotada de forma quase unânime pelos estatísticos durante a primeira metade do século XX.

Ao realizar inferência sobre os parâmetros de um modelo, a informação que se tem do parâmetro de interesse θ é de grande importância na estatística, porém o verdadeiro valor do parâmetro θ é desconhecido. Segundo Paulino et al. (2003), o que é desconhecido, neste caso o parâmetro θ , é incerto, e toda a incerteza deve ser quantificada em termos de probabilidade. A inferência Bayesiana consiste de uma informação a priori dos dados amostrais e do cálculo da densidade a posteriori dos parâmetros. A informação a priori é dada pela densidade de probabilidade $P(\theta)$, a qual expressa o conhecimento do pesquisador sobre os parâmetros a serem estimados.

A inferência Bayesiana trata o vetor de parâmetros desconhecidos como quantidades aleatórias e qualquer informação inicial sobre elas pode ser representada por

modelos probabilísticos para θ . Assim, a abordagem Bayesiana permite incorporar algum conhecimento sobre esses parâmetros antes que os dados tenham sido coletados, atribuindo assim distribuições de probabilidade. Essas distribuições podem ser obtidas por meio de análises anteriores, experiência do pesquisador na área em questão ou em revisões de literatura sobre o assunto que se deseja tratar. Deste modo, para se realizar uma inferência Bayesiana, é necessário modelar uma função densidade de probabilidade a priori $P(\theta)$, que combinada com a função de verossimilhança $L(y_1, \dots, y_n | \theta)$, por meio do teorema de Bayes, gera a função densidade de probabilidade a posteriori $P(\theta | Y_n)$.

A inferência Bayesiana consiste no cálculo da densidade a posteriori dos parâmetros a partir uma informação a priori dos dados amostrais. Quando se tem pouca ou nenhuma informação para incorporar a priori considera-se uma priori não-informativa, como por exemplo, a priori de Jeffreys (Jeffreys, 1961).

A partir do momento que uma distribuição a priori é considerada, seja ela informativa ou não, e obtém-se a função de verossimilhança $L(y_1, \dots, y_n | \theta)$, que é a densidade conjunta dos dados, por meio do Teorema de Bayes, obtém-se a distribuição densidade a posteriori de θ , de forma que qualquer conclusão a respeito do parâmetro θ é realizada a partir da distribuição densidade a posteriori,

$$P(\theta | Y_n) = \frac{L(Y_n | \theta)P(\theta)}{\int L(Y_n | \theta)P(\theta)d\theta} \quad (2)$$

Sendo $Y_n = \{y_1, \dots, y_n\}$.

Em geral omitimos o denominador, uma vez que não depende de θ , e escrevemos a densidade a posteriori como:

$$P(\theta | Y) \propto L(Y_n | \theta)P(\theta) \quad (3)$$

ou seja, Posteriori \propto Verossimilhança x Priori, onde \propto representa proporcionalidade.

Deste modo, toda a inferência sobre o parâmetro θ é realizada por meio da distribuição densidade a posteriori $P(\theta | Y_n)$. Segundo Rosa (1998) para se inferir em relação a qualquer elemento de θ , deve-se integrar a distribuição a posteriori conjunta dos parâmetros, $P(\theta | Y_n)$, em relação a todos os outros parâmetros. Assim, se o interesse do pesquisador se concentra sobre determinado conjunto de θ , por exemplo, θ_1 , necessita-se a obtenção da distribuição $P(\theta_1 | Y)$, denominada de distribuição marginal, dada por:

$$P(\theta_1|Y) = \int_{\theta \neq \theta_1} P(\theta|Y) d\theta \quad (4)$$

A integração da distribuição conjunta a posteriori para a obtenção das marginais geralmente não é analítica, sendo necessário o uso de algoritmos iterativos especializados como o Gibbs Sampler e o Metropolis-Hastings. Estes algoritmos são denominados de algoritmos MCMC (Markov Chain - Monte Carlo). Para a utilização desses algoritmos, é necessário que se obtenha as distribuições condicionais completas para cada parâmetro, que são obtidas a partir das distribuições a posteriori.

2.2.1 Métodos de Monte Carlo via Cadeias de Markov (MCMC)

Os Métodos de Monte Carlo via Cadeias de Markov (MCMC) são de grande importância para a estatística Bayesiana. Estes métodos MCMC consistem em obter uma amostra das distribuições marginais a posteriori dos parâmetros de interesse por meio de um processo iterativo utilizando as distribuições de cada parâmetro condicionada aos demais parâmetros do modelo, que são denominadas distribuições condicionais completas a posteriori $P(\theta_i | \theta_1, \dots, \theta_{i-1}, \dots, \theta_{i+1}, \dots, \theta_p, \mathbf{y})$. Assim os valores gerados são considerados amostras aleatórias de uma determinada distribuição de probabilidade, caracterizando o método de simulação Monte Carlo. Dessa forma, tem-se uma ação conjunta dos métodos MCMC que fazem uma associação entre os algoritmos para simulação de distribuições e o método de Monte Carlo para aproximações de integrais, cujos principais algoritmos são o Metropolis-Hastings e o amostrador de Gibbs.

2.2.2 Algoritmo de Metropolis-Hastings

Este algoritmo permite gerar uma amostra da distribuição conjunta a posteriori, $P(\theta_1, \theta_2, \dots, \theta_k | x)$, a partir das distribuições condicionais completas. Tal algoritmo, é utilizado principalmente quando as distribuições condicionais completas a posteriori dos parâmetros não possuem forma fechada, ou seja não é possível amostrar de seus elementos.

Para a implementação deste algoritmo, é necessário considerar uma distribuição de interesse $q(\theta, \varphi)$. Seja t o estado atual da cadeia de Markov $(\theta^0, \theta^1, \dots, \theta^k, \dots)$, o algoritmo de Metropolis-Hastings gera um valor candidato φ para o próximo estado $(t+1)$

da cadeia, utilizando a distribuição proposta $q(\theta, \varphi)$; e este candidato pode ser aceito com uma probabilidade $\alpha(\theta^t, \varphi)$, em que:

$$\alpha(\theta^t, \varphi) = \min \left\{ 1, \frac{P(\varphi | \mathbf{y}) q(\theta^t | \varphi)}{P(\theta^t | \mathbf{y}) q(\varphi | \theta^t)} \right\}. \quad (5)$$

Se o valor φ for aceito, o próximo estado θ^{t+1} será igual a φ , caso contrário a cadeia não se moverá, ou seja, $\theta^{t+1} = \theta^t$.

Uma apresentação descritiva deste algoritmo é mostrada a seguir.

Descrição do algoritmo de Metropolis-Hastings:

I – Inicialize o contador de iterações $t=0$ e especifique os valores iniciais $\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_d^{(0)})$, onde d representa o número de parâmetros;

II – Gere um novo valor φ a partir da distribuição proposta $q(\theta, \varphi)$;

III – Calcule a probabilidade de aceitação $\alpha(\theta^t, \varphi)$ e gere $u \sim U(0,1)$;

IV - Se $u \leq \alpha$, aceite o novo valor e faça $\theta^{t+1} = \varphi$. Caso contrário, rejeite e faça $\theta^{t+1} = \theta^t$;

V – Incremente o contador de t para $t+1$ e volte ao passo II.

2.2.3 Amostrador de Gibbs

É um caso particular do Metropolis-Hastings que permite gerar uma amostra da distribuição conjunta a posteriori $P(\theta_1, \theta_2, \dots, \theta_k | \mathbf{y})$ a partir das distribuições condicionais completas a posteriori $P(\theta_i | \theta_1, \dots, \theta_{i-1}, \dots, \theta_{i+1}, \dots, \theta_p, \mathbf{y})$, desde que as distribuições condicionais completas possuam forma fechada de forma que seja fácil amostrar de seus elementos (Gelfand, 2000).

Descrição do algoritmo Gibbs Sampler (Amostrador de Gibbs):

I – Inicialize o contador de iterações da cadeia $t=0$;

II – Especifique os valores iniciais $\theta = (\theta_1^0, \theta_2^0, \dots, \theta_p^0)$;

III – Obtenha um novo valor de $\theta^t = (\theta_1^t, \theta_2^t, \dots, \theta_p^t)$ a partir de θ^{t+1} através da geração sucessiva de valores

$$\begin{aligned}\theta_1^t &\sim P(\theta_1 | \theta_2^{t-1}, \theta_3^{t-1}, \dots, \theta_p^{t-1}, \mathbf{y}) \\ \theta_2^t &\sim P(\theta_2 | \theta_1^t, \theta_3^{t-1}, \dots, \theta_p^{t-1}, \mathbf{y}) \\ &\dots \\ \theta_p^t &\sim P(\theta_p | \theta_1^t, \theta_2^t, \dots, \theta_{p-1}^t, \mathbf{y})\end{aligned}$$

IV - Incremente o contador de t para $t+1$ e retorne ao passo II até obter convergência.

2.2.4 Diagnósticos de convergência

Como os algoritmos MCMC são processos iterativos, são necessários métodos para a avaliação de suas convergências. Estes métodos podem ser classificados como formais e informais. Propostos por Gelfand & Smith (1990), os métodos informais são baseados em análises gráficas para a verificação da convergência. Tais métodos consistem em observar a trajetória da cadeia ao longo das iterações, assim, se após um período inicial a trajetória da cadeia apresentar certa estacionariedade, é possível concluir que a convergência foi atingida. Porém, a utilização apenas destes métodos informais não é recomendada, sendo necessária a utilização de alguns métodos formais. Dentre os métodos formais estão os propostos por Heidelberger & Welch (1983), Geweke (1992) e Raftery & Lewis (1992).

O critério proposto por Heidelberger & Welch (1983) é baseado em testes estatísticos para avaliar a estacionariedade da amostra gerada (hipótese nula). Se a hipótese nula for rejeitada para um determinado valor, o teste é repetido depois de descartar os primeiros 10% das iterações. Se a hipótese for rejeitada novamente, descarta-se outros 10% após o descarte dos primeiros 10%. Deste modo, este processo repete-se até serem descartadas 50% das primeiras iterações. Se a hipótese nula for rejeitada novamente, é um indício de que será necessário maior número de iterações. Se a hipótese nula não for rejeitada, o número de iterações descartadas é indicado como o tamanho do burn-in (número de iterações descartadas a fim de evitar a influência dos valores iniciais).

O critério proposto por Geweke (1992), baseado em técnicas de análise espectral, fornece um diagnóstico para ausência de convergência. Este critério é baseado no teste de igualdade de média das primeiras n_a iterações e as últimas n_b iterações restantes. O autor sugere que sejam considerados $n_a=0,1n$ e $n_b=0,5n$, em que n é o tamanho da cadeia. Se a diferença padronizada entre as médias for significativa, existe evidência de ausência de convergência. Se o valor p do critério de Geweke for menor que um nível de significância pré-estabelecido, conclui-se que a cadeia não atingiu a convergência.

O critério proposto por Raftery & Lewis (1992) fornece estimativas para o número de iterações necessário para se obter a convergência. Tal método disponibiliza, também, estimativas do número de iterações iniciais que devem ser descartadas (burn-in) e da distância mínima de uma iteração à outra (thin). Outra saída importante deste método é o fator de dependência (FD), que é responsável pelo acréscimo multiplicativo ao número de iterações necessário para se atingir a convergência. Se este fator for maior que cinco, pode-se dizer que a convergência não foi obtida (Raftery & Lewis, 1992).

Nogueira et al. (2004) sugere o seguinte procedimento de forma combinada para o monitoramento da convergência:

- i- aplicar Raftery & Lewis em uma amostra-piloto e determinar o tamanho ideal da sequência;
- ii- monitorar a convergência das sequências nas proximidades do tamanho ideal pelo critério de Raftery & Lewis e por meio dos critérios de Gelman & Rubin e Geweke;
- iii- determinar o tamanho do burn-in pelo critério de Heidelberger & Welch.

2.3 Abordagem Frequentista

A abordagem Frequentista tradicional é fundamentada em processos iterativos, como o Algoritmo de Gauss-Newton, DUD e Algoritmo de Marquardt, devido à não-linearidade das variáveis (Filho et al., 2008). Tais procedimentos baseiam-se na minimização da soma de quadrados dos resíduos, que consiste em encontrar o melhor ajustamento para um conjunto de dados tentando minimizar a soma de quadrados dos resíduos entre a curva ajustada e os valores observados (Gallant, 1987).

2.3.1 Método dos quadrados mínimos ordinários

Matricialmente, o modelo de regressão não linear $\mathbf{y} = \mathbf{f}(\boldsymbol{\theta}) + \boldsymbol{\varepsilon}$, é dado por:

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \mathbf{f}(\boldsymbol{\theta}) = \begin{bmatrix} f(x_1, \boldsymbol{\theta}) \\ f(x_2, \boldsymbol{\theta}) \\ \vdots \\ f(x_n, \boldsymbol{\theta}) \end{bmatrix} \quad \text{e} \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}. \quad (6)$$

A soma dos quadrados dos erros aleatórios (SQE) deverá ser minimizada por $\boldsymbol{\theta}$. Deste modo, a função de mínimos quadrados para um modelo não linear é escrita da seguinte forma:

$$SQE(\boldsymbol{\theta}) = \sum_{i=1}^n [y_i - f(x_i, \boldsymbol{\theta})]^2, \quad (7)$$

que, em notação vetorial, pode ser escrita como:

$$\text{SQE}(\boldsymbol{\theta}) = \|\mathbf{y} - \mathbf{f}(\boldsymbol{\theta})\|^2, \quad (8)$$

$$\text{SQE}(\boldsymbol{\theta}) = [\mathbf{y} - \mathbf{f}(\boldsymbol{\theta})]' [\mathbf{y} - \mathbf{f}(\boldsymbol{\theta})], \quad (9)$$

em que $\|\mathbf{y} - \mathbf{f}(\boldsymbol{\theta})\|$ é a norma euclidiana de $\mathbf{y} - \mathbf{f}(\boldsymbol{\theta})$.

Segundo Souza (1998), em modelos de regressão não linear não se pode fazer afirmações gerais sobre as propriedades dos estimadores de quadrados mínimos, tais como não tendenciosidade e variância mínima, exceto para grandes amostras, os chamados resultados assintóticos.

Para uma melhor compreensão do processo de obtenção destes estimadores, utilizou-se a seguinte notação de diferenciação matricial:

$$\mathbf{f}(\boldsymbol{\theta}) = \begin{bmatrix} f_1(\boldsymbol{\theta}) \\ f_2(\boldsymbol{\theta}) \\ \vdots \\ f_n(\boldsymbol{\theta}) \end{bmatrix} \quad (10)$$

$$\mathbf{F}(\boldsymbol{\theta}) = \frac{\partial \mathbf{f}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} = \begin{bmatrix} \frac{\partial f_1(\boldsymbol{\theta})}{\partial \theta_1} & \frac{\partial f_1(\boldsymbol{\theta})}{\partial \theta_2} & \dots & \frac{\partial f_1(\boldsymbol{\theta})}{\partial \theta_p} \\ \frac{\partial f_2(\boldsymbol{\theta})}{\partial \theta_1} & \frac{\partial f_2(\boldsymbol{\theta})}{\partial \theta_2} & \dots & \frac{\partial f_2(\boldsymbol{\theta})}{\partial \theta_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n(\boldsymbol{\theta})}{\partial \theta_1} & \frac{\partial f_n(\boldsymbol{\theta})}{\partial \theta_2} & \dots & \frac{\partial f_n(\boldsymbol{\theta})}{\partial \theta_p} \end{bmatrix} \quad (11)$$

em que: $\mathbf{f}(\boldsymbol{\theta})$ é uma função vetor de dimensão $n \times 1$ do vetor de parâmetros $\boldsymbol{\theta}$, de dimensão $p \times 1$ e $\mathbf{F}(\boldsymbol{\theta})$ é a matriz Jacobiana de $\mathbf{f}(\boldsymbol{\theta})$. Dessa forma, o estimador de mínimos quadrados, $\hat{\boldsymbol{\theta}}$ deve satisfazer a equação:

$$\left. \frac{\partial \text{SQE}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \mathbf{0} \quad (12)$$

a qual representa a minimização de interesse.

Sendo,

$$\left. \frac{\partial \text{SQE}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \frac{\partial}{\partial \boldsymbol{\theta}'} [\mathbf{y} - \mathbf{f}(\boldsymbol{\theta})]' [\mathbf{y} - \mathbf{f}(\boldsymbol{\theta})] \quad (13)$$

$$= -2[\mathbf{y} - \mathbf{f}(\boldsymbol{\theta})]' \mathbf{F}(\boldsymbol{\theta}) \quad (14)$$

tem-se:

$$= \mathbf{F}'(\hat{\boldsymbol{\theta}}) \big|_{\mathbf{y} - \mathbf{f}(\hat{\boldsymbol{\theta}})} = \mathbf{0} \quad (15)$$

Portanto, o sistema de equações normais (SEN) para regressão não linear é dado por:

$$\begin{bmatrix} \frac{\partial f_1(\hat{\boldsymbol{\theta}})}{\partial \hat{\theta}_1} & \frac{\partial f_2(\hat{\boldsymbol{\theta}})}{\partial \hat{\theta}_1} & \dots & \frac{\partial f_n(\hat{\boldsymbol{\theta}})}{\partial \hat{\theta}_1} \\ \frac{\partial f_1(\hat{\boldsymbol{\theta}})}{\partial \hat{\theta}_2} & \frac{\partial f_2(\hat{\boldsymbol{\theta}})}{\partial \hat{\theta}_2} & \dots & \frac{\partial f_n(\hat{\boldsymbol{\theta}})}{\partial \hat{\theta}_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_1(\hat{\boldsymbol{\theta}})}{\partial \theta_p} & \frac{\partial f_2(\hat{\boldsymbol{\theta}})}{\partial \theta_p} & \dots & \frac{\partial f_n(\hat{\boldsymbol{\theta}})}{\partial \theta_p} \end{bmatrix} \cdot \left(\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} - \begin{bmatrix} f_1(\hat{\boldsymbol{\theta}}) \\ f_2(\hat{\boldsymbol{\theta}}) \\ \vdots \\ f_n(\hat{\boldsymbol{\theta}}) \end{bmatrix} \right) = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (16)$$

Geralmente, para estas equações que são não lineares em relação aos parâmetros, não existe uma solução explícita, sendo necessário processos iterativos para obter as estimativas dos parâmetros.

2.3.2 Processos iterativos e método dos quadrados mínimos

Segundo Regazzi (2003), um método amplamente usado em algoritmos computacionais para regressão não linear é a linearização da função não linear, conhecido por método de Gauss-Newton, o qual se resume ao seguinte procedimento:

supondo uma aproximação inicial $\hat{\boldsymbol{\theta}}_0$ para $\hat{\boldsymbol{\theta}}$ e, aproximando $\mathbf{f}(\boldsymbol{\theta})$ pelo ponto $\hat{\boldsymbol{\theta}}_0$ por uma expansão de Taylor de 1ª ordem, tem-se:

$$\mathbf{f}(x_i, \boldsymbol{\theta}) = \mathbf{f}(x_i, \hat{\boldsymbol{\theta}}_0) + \frac{\partial \mathbf{f}(x_i, \hat{\boldsymbol{\theta}}_0)}{\partial \boldsymbol{\theta}'} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_0) \quad (17)$$

que produzem a aproximação linear:

$$\mathbf{f}(\boldsymbol{\theta}) \cong \mathbf{f}(\hat{\boldsymbol{\theta}}_0) + \mathbf{F}(\hat{\boldsymbol{\theta}}_0) (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_0) \quad (18)$$

em que:

$$\mathbf{F}(\hat{\boldsymbol{\theta}}_0) = \frac{\partial \mathbf{f}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \bigg|_{\boldsymbol{\theta}' = \hat{\boldsymbol{\theta}}_0'} \quad (19)$$

Deste modo, o problema de minimizar a função (8) do item 2.3.1 da soma dos quadrados dos erros aleatórios SQE, torna-se minimizar a função $\mathbf{f}(\boldsymbol{\theta})$ associada à aproximação (18), dada por:

$$SQE(\theta) = \left\| \mathbf{y} - \mathbf{f}(\hat{\theta}_0) - \mathbf{F}(\hat{\theta}_0)(\theta - \hat{\theta}_0) \right\|^2 \quad (20)$$

Fazendo $\mathbf{y} - \mathbf{f}(\hat{\theta}_0) = \mathbf{E}_0$ e $\theta - \hat{\theta}_0 = \Delta\hat{\theta}_0$ em (5), tem-se que:

$$SQE(\theta) = \left\| \mathbf{E}_0 - \mathbf{F}(\hat{\theta}_0)\Delta\hat{\theta}_0 \right\|^2 \quad (21)$$

$$SQE(\theta) = \left[\mathbf{E}_0 - \mathbf{F}(\hat{\theta}_0)\Delta\hat{\theta}_0 \right]' \left[\mathbf{E}_0 - \mathbf{F}(\hat{\theta}_0)\Delta\hat{\theta}_0 \right] \quad (22)$$

$$SQE(\theta) = \mathbf{E}_0' \mathbf{E}_0 - \mathbf{E}_0' \mathbf{F}(\hat{\theta}_0) \Delta\hat{\theta}_0 - \Delta\hat{\theta}_0' \mathbf{F}'(\hat{\theta}_0) \mathbf{E}_0 + \Delta\hat{\theta}_0' \mathbf{F}'(\hat{\theta}_0) \mathbf{F}(\hat{\theta}_0) \Delta\hat{\theta}_0 \quad (23)$$

$$SQE(\theta) = \mathbf{E}_0' \mathbf{E}_0 - 2\Delta\hat{\theta}_0' \mathbf{F}'(\hat{\theta}_0) \mathbf{E}_0 + \Delta\hat{\theta}_0' \mathbf{F}'(\hat{\theta}_0) \mathbf{F}(\hat{\theta}_0) \Delta\hat{\theta}_0 \quad (24)$$

Derivando $SQE(\theta)$ em relação a $\Delta\hat{\theta}_0$ tem-se:

$$\frac{\partial SQE(\theta)}{\partial \Delta\hat{\theta}_0} = \mathbf{0} - 2\mathbf{F}'(\hat{\theta}_0) \mathbf{E}_0 + 2\mathbf{F}'(\hat{\theta}_0) \mathbf{F}(\hat{\theta}_0) \Delta\hat{\theta}_0 \quad (25)$$

$$= 2 \left[\mathbf{F}'(\hat{\theta}_0) \mathbf{F}(\hat{\theta}_0) \Delta\hat{\theta}_0 + \mathbf{F}'(\hat{\theta}_0) \mathbf{E}_0 \right] \quad (26)$$

Agora, igualando a derivada a zero, temos o Sistema de Equações Normais

$$\mathbf{F}'(\hat{\theta}_0) \mathbf{F}(\hat{\theta}_0) \Delta\hat{\theta}_0 = \mathbf{F}'(\hat{\theta}_0) \mathbf{E}_0 \quad (27)$$

onde \mathbf{E}_0 desempenha o papel de variável dependente \mathbf{y} e $\mathbf{F}'(\hat{\theta}_0)$ desempenha o papel da matriz X dos modelos lineares.

Assim, o valor de $\Delta\hat{\theta}_0$ que minimiza a SQE é dado por:

$$\Delta\hat{\theta}_0 = \left[\mathbf{F}'(\hat{\theta}_0) \mathbf{F}(\hat{\theta}_0) \right]^{-1} \mathbf{F}'(\hat{\theta}_0) \mathbf{E}_0 \quad (28)$$

Desde que $\mathbf{F}'(\hat{\theta}_0)$ possua posto coluna completo.

Como $\theta - \hat{\theta}_0 = \Delta\hat{\theta}_0$, pode-se definir um vetor $\hat{\theta}_1 = \hat{\theta}_0 + \Delta\hat{\theta}_0$ como estimativas atualizadas de θ . A partir deste procedimento, pode-se obter um outro conjunto de estimativas atualizadas $\hat{\theta}_2$ utilizando-se $\hat{\theta}_1$, e assim por diante.

Deste modo uma k-ésima iteração será:

$$\hat{\theta}_{k+1} = \hat{\theta}_k + \Delta\hat{\theta}_k \quad (29)$$

$$\hat{\theta}_{k+1} = \hat{\theta}_k + \left[\mathbf{F}'(\hat{\theta}_k) \mathbf{F}(\hat{\theta}_k) \right]^{-1} \mathbf{F}'(\hat{\theta}_k) \mathbf{E}_k \quad (30)$$

em que:

$$\mathbf{F}'(\hat{\theta}_k) = \left. \frac{\partial \mathbf{f}(\theta)}{\partial \theta'} \right|_{\theta=\hat{\theta}_k} \quad (31)$$

$$\mathbf{E}_k = \mathbf{y} - \mathbf{f}(\hat{\theta}_k) \quad (32)$$

$$\hat{\theta}_k = [\hat{\theta}_{1k} \quad \hat{\theta}_{2k} \quad \dots \quad \hat{\theta}_{pk}]' \quad (33)$$

Este processo iterativo prossegue até que a convergência seja atingida, isto é, quando:

$$\left| \frac{\hat{\theta}_{j(k+1)} - \hat{\theta}_{jk}}{\hat{\theta}_{jk}} \right| < \delta, \quad \text{para } j=1, 2, \dots, p, \quad (34)$$

em que δ é algum valor fixo bem pequeno.

2.4 Avaliadores da qualidade de ajuste

Devido às diferenças entre alguns modelos de regressão não linear, quando estes são ajustados a um mesmo conjunto de dados, torna-se necessário utilizar metodologias estatísticas com o intuito de compará-los e indicar o melhor modelo (Silveira et al., 2009).

Dentre os diversos avaliadores da qualidade de ajuste presentes nas literaturas para a abordagem Frequentista temos o percentual de convergência (C%), quadrado médio do erro (QME), coeficiente de determinação (R^2), critério de informação de Akaike (AIC), critério de informação Bayesiano (BIC) e Desvio Médio absoluto (DMA). Já para a abordagem Bayesiana, um avaliador muito utilizado é o Deviance Information Criterion (DIC).

2.4.1 Percentual de convergência (C%)

É usado quando se tem ajustes individuais, ou seja, quando se ajusta uma curva para dados de cada planta. Deste modo, é possível verificar qual modelo apresenta uma maior facilidade de convergência, que é dado pela porcentagem de ajustes que convergiram.

2.4.2 Quadrado médio do erro (QME)

Um dos critérios mais eficientes para testar a qualidade de ajuste de um modelo de regressão é o erro quadrático médio de predição (Chrobok et al., 2004). Representa a estimativa da variância residual e quanto menor o valor do QME, melhor será considerado o modelo. Sua expressão é dada por:

$$QME = \frac{\sum_{i=1}^n (\hat{Y}_i - \hat{Y}_i)^2}{n - p} \quad (35)$$

em que: n é o número de observações utilizadas para ajustar a curva; p é o número de parâmetros na função, Y_i é o valor observado da matéria seca total da observação i e \hat{Y}_i é o valor estimado da matéria seca total da observação i para o modelo analisado.

2.4.3 Coeficiente de determinação (R^2)

Quanto ao coeficiente de determinação para modelos de regressão não linear, existem diversas dúvidas. Por exemplo, segundo Ratkowsky (1990), o R^2 não tem nenhum significado óbvio e, portanto, não precisa ser calculado. Por outro lado, Souza (1998), afirma que a qualidade do ajustamento para o caso de regressão não linear pode ser medida pelo quadrado do coeficiente de correlação entre os valores observados e preditos.

O coeficiente de determinação pode ser calculado por:

$$R^2 = 1 - \frac{SQR}{SQT} \quad (36)$$

onde SQR é a soma dos quadrados residuais; e SQT é a soma de quadrados total. Quanto maior é o coeficiente de determinação calculado, melhor a qualidade do modelo ajustado.

2.4.4 Critério de informação de Akaike (AIC)

O Critério de informação de Akaike (Akaike, 1974) é uma estatística para comparação da qualidade de ajuste do modelo baseado no máximo da função de verossimilhança, e depende do número de observações e do número de parâmetros do modelo em questão. Este critério admite a existência de um modelo real, mas desconhecido, que descreve os dados. Assim, ele tenta escolher dentre um grupo de modelos avaliados aquele que minimiza a divergência de Kull-Leiber, que está relacionada à informação perdida por se usar o modelo aproximado ao invés do real (Sousa, 2012). Sua expressão é dada por:

$$AIC = 2(p) - 2\log L(\hat{\theta}) \quad (37)$$

em que: p é o número de parâmetros e $L(\hat{\theta})$ é o máximo da função de verossimilhança. Menores valores de AIC indicam um melhor ajuste (Akaike, 1974).

2.4.5 Critério de Informação Bayesiano (BIC)

Assim como no critério de informação de Akaike, o critério de informação Bayesiano (BIC), proposto por Schwarz (1978), é uma estatística para comparação da qualidade de ajuste do modelo baseado no máximo da função de verossimilhança, e

depende do número de observações e parâmetros do modelo analisado. Este critério maximiza a probabilidade de se identificar o verdadeiro modelo dentre os modelos avaliados (Sousa, 2012). Sua expressão é dada por:

$$\text{BIC} = p \log(n) - 2 \log L(\hat{\theta}) \quad (38)$$

em que: p é o número de parâmetros; n é o número de observações e $L(\hat{\theta})$ é o máximo da função de verossimilhança.

Menores valores de BIC indicam um melhor ajuste (Akaike, 1974).

2.4.6 Desvio médio absoluto (DMA)

Proposto por Sarmiento et al. (2006), o desvio médio absoluto dos resíduos (DMA), é calculado como o somatório dos desvios entre os valores observados e estimados, dividido pelo número de observações. Sua expressão é dada por:

$$\text{DMA} = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)}{n} \quad (39)$$

em que: n é o número de observações utilizadas para ajustar a curva; Y_i é o valor observado da matéria seca total da observação i e \hat{Y}_i é o valor estimado da matéria seca total da observação i para o modelo analisado.

Segundo Sarmiento et al. (2006), quanto menor o desvio médio absoluto dos resíduos, melhor o modelo ajustado.

2.4.7 Deviance Information Criterion (DIC)

O DIC (Deviance Information Criterion) é usual nos problemas bayesianos de seleção de modelos para os quais as amostras da distribuição a posteriori dos parâmetros dos modelos são obtidas por simulação de Monte Carlo via Cadeia de Markov (MCMC).

Define-se o desvio como:

$$D(\theta) = -2 \log L(\hat{\theta}) + C \quad (40)$$

em que: θ é o vetor de parâmetros desconhecidos do modelo; $L(\hat{\theta})$ é a função de verossimilhança e C é uma constante que anula em todos os cálculos que comparam diferentes modelos, portanto não precisa ser conhecida na comparação de dois modelos.

O critério DIC é dado por:

$$\text{DIC} = D(\hat{\theta}) + 2p_D \quad (41)$$

em que: $D(\hat{\theta})$ é o desvio avaliado na média a posteriori e p_D é o número efetivo de parâmetros no modelo, que é dado por $p_D = \bar{D} - D(\hat{\theta})$, em que $\bar{D} = E\{D(\hat{\theta})\}$ é o desvio médio a posteriori que mede a qualidade do ajuste do modelo aos dados. Quanto menor o valor para o DIC, melhor será o modelo.

2.5 Análise de agrupamento

A análise de agrupamento é uma técnica utilizada para obtenção de grupos semelhantes por meio de um esquema que permita classificar os indivíduos em determinado número de grupos, sendo que deve existir grande homogeneidade entre os indivíduos do mesmo grupo e heterogeneidade entre os grupos (Ferreira, 2011).

De acordo com Cruz et al. (2011), os métodos de agrupamento mais utilizados são os de otimização e os hierárquicos. Os métodos de otimização visam uma partição dos indivíduos de forma a otimizar (minimizar ou maximizar) alguma medida predefinida. Já os métodos hierárquicos agrupam os indivíduos por um processo que se repete em vários níveis, até que seja estabelecido um dendrograma (diagrama bidimensional com formato de árvore), ilustrando as partições ou fusões efetuadas em cada nível sucessivo do processo de agrupamento (Cruz et al., 2004). Os métodos hierárquicos são divididos em métodos aglomerativos e divisivos. Nos métodos aglomerativos, através de sucessivas fusões das n unidades, formam-se $n-1$, $n-2$, ..., $n-k$ grupos, até que todas as unidades estejam reunidas em um único grupo. Nos métodos divisivos, a idéia é partir de um único grupo e, através de divisões sucessivas, obter outros sub-grupos. Dentre os hierárquicos, citam-se o do vizinho mais próximo (Single Linkage Method); o do vizinho mais distante (Complete Linkage Method); o da ligação média (Average Linkage), ponderado ou não; e o proposto por Ward (1963).

2.6 Medidas de dissimilaridades

As medidas de dissimilaridades são de extrema importância em estudos de divergência genética para definição de progenitores a serem utilizados em programas de hibridações (Cruz et al., 2011) e podem ser definidas como critérios que quantificam a distância entre dois objetos. Quanto maior o valor mensurado, menos semelhantes são os objetos e quanto menor, mais semelhantes eles são (Ferreira, 2011).

De acordo com Cruz et al. (2004), dentre as distâncias mais utilizadas em estudos genéticos tem-se a distância euclidiana quadrática, distância euclidiana padronizada quadrática e a distância generalizada de Mahalanobis.

De acordo com Ferreira (2011), a maior parte dos métodos de agrupamento requer a obtenção de uma matriz de proximidades entre os indivíduos. O termo proximidade é utilizado para indicar ou similaridade ou dissimilaridade, que é medida pelas distâncias. A definição de qual medida de proximidade a ser adotada está diretamente ligada à natureza das variáveis a serem utilizadas na análise de agrupamento.

Sejam y_r e y_s vetores p -dimensionais de observações dos objetos r e s , para $r, s = 1, 2, \dots, n$. A distância quadrática entre os dois objetos r e s é dada por:

$$d_{rs}^2 = \|y_r - y_s\|_{\Psi}^2 = (y_r - y_s)^T \Psi (y_r - y_s), \quad (42)$$

sendo Ψ uma métrica de interesse

Segundo Ferreira (2011), de acordo com a métrica utilizada são definidas diferentes distâncias.

- 1) Se Ψ for a matriz identidade, a distância quadrática em questão é a euclidiana. Essa distância é mais apropriada para casos em que as variáveis mensuradas possuem escalas similares, pois caso contrário, variáveis com maior variabilidade irão ter um grande peso no cálculo das distâncias.
- 2) Se a matriz Ψ for dada por $\Psi = D^{-1} = \text{diag}\left(\frac{1}{S_{kk}}\right)$, então a distância quadrática é denotada por distância euclidiana padronizada quadrática, sendo S_{kk} a estimativa da variância da k -ésima variável ao longo da amostra de n objetos. Tal distância, conhecida também como distância de Karl Pearson, é indicada para variáveis mensuradas em diferentes escalas, porém não correlacionadas. Segundo Cruz et al. (2011) as razões para a padronização dos dados são: evitar que as unidades escolhidas para medir as variáveis afetem arbitrariamente a similaridade entre os indivíduos e, a padronização faz com que as variáveis contribuam igualmente na avaliação da similaridade entre indivíduos.
- 3) Se Ψ for dada por $\Psi = S^{-1}$, a distância quadrática corresponde à distância generalizada de Mahalanobis, a qual contorna tanto o problema de escala quanto leva em consideração a correlação entre as variáveis.

2.7 Determinação do número ótimo de grupos

A análise de agrupamento constitui uma das mais importantes técnicas de classificação multivariada, uma vez que permite classificar várias unidades amostrais em

um número reduzido de grupos contendo unidades semelhantes (Silveira, 2011). Porém, a determinação do número ótimo de grupos não é tão simples. Segundo Faria (2009) não há um critério objetivo que determine o ponto de corte do dendrograma.

O número de grupos nos métodos hierárquicos podem ser definidos por: razões inerentes ao pesquisador; análise visual das ramificações do dendrograma e critérios estatísticos, como: BBS (Between-group Sum of Squares) que é a soma de quadrados entre grupos para dois agrupamentos, e mede a homogeneidade de grupos unidos; a RSQ (Root-mean-square), que é a razão entre a soma de quadrados entre grupos e a soma de quadrados total, e mede a heterogeneidade do agrupamento; SPRSQ (Semi-partial R-Square) que mede a perda de homogeneidade por juntar dois grupos (Khattree & Naik, 2000), além do método de Mojena, baseado na maior amplitude das distâncias de junção dos grupos formados com o objetivo de determinar um número de grupos k que otimiza a qualidade do agrupamento dos dados (Mojena, 1977).

3 MATERIAL E MÉTODOS

3.1 Descrição do experimento

O experimento foi realizado a campo, no período de março a novembro de 2010, em área experimental pertencente ao Setor de Olericultura do Departamento de Fitotecnia da Universidade Federal de Viçosa (UFV), no município de Viçosa, Zona da Mata de Minas Gerais, cujas coordenadas geográficas são: 20° 45' de latitude Sul e 42° 51' de longitude Oeste, com altitude média de 650 m.

Foram avaliados 30 acessos de alho, registrados no Banco de Germoplasma de Hortaliças (BGH/UFV) (Tabela 2). O delineamento experimental utilizado foi o de blocos completos casualizados, com quatro repetições.

As parcelas (unidades experimentais) foram constituídas por quatro fileiras transversais de 1,0 m de comprimento, com plantas no espaçamento de 0,25 x 0,10 m, totalizando 40 plantas e foram consideradas como úteis as plantas das duas fileiras centrais.

As plantas colhidas foram submetidas ao processo de cura de campo e de galpão, por 3 e 60 dias, respectivamente. A avaliação de matéria seca total das plantas desses acessos foi realizada em quatro períodos, sendo: 60, 90, 120 e 150 dias após o plantio (DAP).

A relação dos 30 acessos de alho registrados no Banco de Germoplasma de Hortaliças da Universidade Federal de Viçosa (BGH/UFV), utilizados neste trabalho, com os respectivos nomes e origem, estão apresentados na Tabela 2.

Tabela 2. Relação dos 30 acessos de alho registrados no Banco de Germoplasma de Hortaliças da Universidade Federal de Viçosa (BGH/UFV) com respectivos nomes e origem

n	Acesso	Nome comum / variedade	Origem
1	492	-	Goianira - GO
2	4493	Cateto Roxo	Florestal – MG
3	4567	Cateto Roxo	Gouveia – MG
4	4662	-	-
5	4814	-	Santa Tereza – MG
6	4815	-	S. Leopoldina - ES
7	4818	-	Santa Tereza – MG
8	4820	-	Santa Tereza
9	4823	-	S. Leopoldina - ES
10	4825	Sapé	Rio Pantojas - ES
11	4830	Sapé	Rio Pasmoser
12	4832	Sapé	Rio Pasmoser
13	4835	-	Rio Pasmoser
14	4839	-	Cedrolândia N. Venécia -ES
15	4843	-	Itarana - ES
16	5366	Cultura de tecidos	Brasília - DF
17	5939	Amarante Aimorés	Viçosa - MG
18	5940	Amarante Novo Cruzeiro	Viçosa - MG
19	5943	Branco Mineiro	Viçosa - MG
20	5944	Cateto Amarantina	Viçosa - MG
21	5950	Chinês - 1	Viçosa - MG
22	5959	Gigante Roxão	EMCAPA
23	5960	-	Juréia
24	5964	-	Simonésia
25	6390	-	Araponga - MG
26	6396	-	-
27	7619	Amarante – EPAMIG	Espirito Santo – ES
28	7620	Amarante – EPAMIG	Espirito Santo – ES
29	7621	Amarante – EPAMIG	Espirito Santo – ES
30	7624	Amarante – EPAMIG	Espirito Santo – ES

3.2 Abordagem Bayesiana

Na abordagem Bayesiana, as estimativas dos parâmetros das curvas de crescimento foram obtidas por meio de um método hierárquico Bayesiano (Macedo et al., 2014), o qual foi ajustado separadamente para cada um dos sete modelos considerados na Tabela 1.

O primeiro estágio foi representado pela modelagem individual da forma da curva de crescimento de cada acesso, sendo este descrito por cada um dos modelos de crescimento considerados.

No segundo estágio da hierarquia, foram consideradas distribuições a priori para os parâmetros das distribuições assumidas no primeiro estágio, e no terceiro e último estágio, foram especificadas as distribuições a priori para os parâmetros das distribuições assumidas no segundo estágio (os hiperparâmetros).

A função de verossimilhança (distribuição conjunta dos dados amostrais) foi obtida, separadamente para cada modelo.

Como exemplo, para o modelo Logístico, a metodologia Bayesiana foi empregada mediante as seguintes especificações.

A distribuição dos dados amostrais foi dada por:

$$y_{ij} | \theta, \sigma_e^2 \sim N \left(A_i \left[1 + b_i \exp \left(- K_i t_{ij} \right) \right]^{-1}, \sigma_e^2 \right), \quad \theta = (A_i, b_i, K_i) \quad (43)$$

Em que A, b e K representam, respectivamente, os parâmetros β_1, β_2 e β_3 .

Portanto, a função de verossimilhança é:

$$L(y | \theta, \sigma_e^2) \propto \prod_{i=1}^N \prod_{j=1}^n \frac{1}{\sqrt{2\pi\sigma_e^2}} \exp \left\{ - \frac{\left[y_{ij} - A_i \left[1 + b_i \exp \left(- K_i t_{ij} \right) \right]^{-1} \right]^2}{2\sigma_e^2} \right\} \quad (44)$$

As distribuições a priori utilizadas foram:

$$A_i \sim N(\mu_A, \tau_A) \quad (45)$$

$$b_i \sim N(\mu_b, \tau_b) \quad (46)$$

$$k_i \sim N(\mu_k, \tau_k) \quad (47)$$

$$\sigma_e^2 \sim G(\mu_{\sigma_e^2}, \tau_{\sigma_e^2}) \quad (48)$$

Em que $\sigma_e^2 = \frac{1}{\tau_e}$, sendo σ_e^2 a variância do erro.

E no último estágio foram especificados os valores dos hiperparâmetros (parâmetros das distribuições assumidas nas distribuições anteriores)

$$\mu_A \sim N(\mu_{\mu_A}, \tau_{\tau_A}) \quad (49)$$

$$\mu_b \sim N(\mu_{\mu_b}, \tau_{\tau_b}) \quad (50)$$

$$\mu_k \sim N(\mu_{\mu_k}, \tau_{\tau_k}) \quad (51)$$

$$\tau_A \sim G(\alpha_A, \beta_A) \quad (52)$$

$$\tau_b \sim G(\alpha_b, \beta_b) \quad (53)$$

Os valores para os hiperparâmetros foram escolhidos com base em análises prévias e em revisões de literatura como Puiatti et al. (2013) e Reis et al. (2014).

A modelagem apresentada foi implementada no software R (R DEVELOPMENT CORE TEAM, 2014) através do pacote R2OpenBUGS, no qual foi implementado o algoritmo para o Amostrador de Gibbs e o Metropolis-Hastings. Para todos os modelos foram utilizadas 20.000 iterações, das quais as 1.000 primeiras iterações foram descartadas a fim de evitar os efeitos devido aos valores iniciais. Considerou-se, também, para assegurar a independência da amostra, um espaçamento entre os pontos amostrais (thin) de tamanho 4. Deste modo, obteve-se uma cadeia final com 4.000 observações para cada parâmetro. Para avaliar a convergência foi utilizado o teste de Geweke e o teste de Raftery & Lewis (1992) mediante o pacote Bayesian Output Analysis (BOA) do software R.

3.3 Abordagem Frequentista

Para a abordagem Frequentista tradicional os parâmetros dos modelos considerados na Tabela 1, utilizados para descrever o acúmulo de matéria seca para os 30 acessos de alho utilizados neste trabalho foram estimados por meio do método dos quadrados mínimos ordinários, com soluções obtidas por meio do processo iterativo de Gauss-Newton, cujo procedimento baseia-se na minimização da soma de quadrados dos resíduos.

Como exemplo, considerando o acesso de número 9, o modelo Logístico e utilizando o método de Gauss-Newton com os seguintes valores iniciais $\hat{\theta}_{1;0} = 20$, $\hat{\theta}_{2;0} = 5000$ e $\hat{\theta}_{3;0} = 0,08$, temos:

$$\mathbf{x} = \begin{bmatrix} 60 \\ 90 \\ 120 \\ 150 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} 1,08228 \\ 6,6375 \\ 21,1125 \\ 24,225 \end{bmatrix} \quad \text{e} \quad \hat{\boldsymbol{\theta}} = \begin{bmatrix} 20 \\ 5000 \\ 0,08 \end{bmatrix} \quad (54)$$

A partir desses valores e utilizando a equação (7), temos:

$$SQE(\hat{\theta}_0) = \sum_{i=1}^n [y_i - f(x_i, \theta)]^2 = 67,5228 \quad (55)$$

$$\text{em que: } f(x_i, \theta) = \begin{bmatrix} 0,4745 \\ 0,2257 \\ 14,9405 \\ 19,4039 \end{bmatrix} \quad (56)$$

Precisamos, agora, calcular as derivadas parciais do modelo Logístico em relação a cada parâmetro.

Como apresentado na Tabela 1, a equação do modelo Logístico é dada por:

$$f(x) = \frac{\beta_1}{1 + \beta_2 e^{-\beta_3 x}} \quad (57)$$

Calculando as derivadas parciais em relação a cada parâmetro temos:

$$\frac{\partial f(x, \beta_1, \beta_2, \beta_3)}{\partial \beta_1} = \frac{1}{1 + \beta_2 e^{-\beta_3 x}} \quad (58)$$

$$\frac{\partial f(x, \beta_1, \beta_2, \beta_3)}{\partial \beta_2} = \frac{-\beta_1 e^{-\beta_3 x}}{(1 + \beta_2 e^{-\beta_3 x})^2} \quad (59)$$

$$\frac{\partial f(x, \beta_1, \beta_2, \beta_3)}{\partial \beta_3} = \frac{-x \beta_1 \beta_2 e^{-\beta_3 x}}{(1 + \beta_2 e^{-\beta_3 x})^2} \quad (60)$$

Assim, para $x=60$ temos:

$$F_{1;1} = \frac{1}{1 + \beta_2 e^{-\beta_3 x}} = \frac{1}{1 + 5000 e^{-0,08 \cdot 60}} = 0,023726 \quad (61)$$

$$F_{1;2} = \frac{-\beta_1 e^{-\beta_3 x}}{(1 + \beta_2 e^{-\beta_3 x})^2} = \frac{-22 \cdot e^{-0,08 \cdot 60}}{(1 + 5000 e^{-0,08 \cdot 60})^2} = -0,00009265 \quad (62)$$

$$F_{1;3} = \frac{-x \beta_1 \beta_2 e^{-\beta_3 x}}{(1 + \beta_2 e^{-\beta_3 x})^2} = \frac{-60 \cdot 22 \cdot 5000 \cdot e^{-0,08 \cdot 60}}{(1 + 5000 e^{-0,08 \cdot 60})^2} = 27,79513 \quad (63)$$

para $x=90$ temos:

$$F_{2;1} = \frac{1}{1 + \beta_2 e^{-\beta_3 x}} = \frac{1}{1 + 5000 e^{-0,08 \cdot 90}} = 0,211286 \quad (64)$$

$$F_{2;2} = \frac{-\beta_1 e^{-\beta_3 x}}{(1 + \beta_2 e^{-\beta_3 x})^2} = \frac{-22 \cdot e^{-0,08 \cdot 90}}{(1 + 5000 e^{-0,08 \cdot 90})^2} = -0,0006666 \quad (65)$$

$$F_{2:3} = \frac{-x\beta_1\beta_2e^{-\beta_3x}}{(1+\beta_2e^{-\beta_3x})^2} = \frac{-60 \cdot 22 \cdot 5000 \cdot e^{-0,08 \cdot 90}}{(1+5000e^{-0,08 \cdot 90})^2} = 299,9593 \quad (66)$$

para $x=120$ temos:

$$F_{3:1} = \frac{1}{1+\beta_2e^{-\beta_3x}} = \frac{1}{1+5000e^{-0,08 \cdot 120}} = 0,747025 \quad (67)$$

$$F_{3:2} = \frac{-\beta_1e^{-\beta_3x}}{(1+\beta_2e^{-\beta_3x})^2} = \frac{-22 \cdot e^{-0,08 \cdot 120}}{(1+5000e^{-0,08 \cdot 120})^2} = -0,0007559 \quad (68)$$

$$F_{3:3} = \frac{-x\beta_1\beta_2e^{-\beta_3x}}{(1+\beta_2e^{-\beta_3x})^2} = \frac{-60 \cdot 22 \cdot 5000 \cdot e^{-0,08 \cdot 120}}{(1+5000e^{-0,08 \cdot 120})^2} = 453,549 \quad (69)$$

para $x=150$ temos:

$$F_{4:1} = \frac{1}{1+\beta_2e^{-\beta_3x}} = \frac{1}{1+5000e^{-0,08 \cdot 150}} = 0,970195 \quad (70)$$

$$F_{4:2} = \frac{-\beta_1e^{-\beta_3x}}{(1+\beta_2e^{-\beta_3x})^2} = \frac{-22 \cdot e^{-0,08 \cdot 150}}{(1+5000e^{-0,08 \cdot 150})^2} = -0,0001157 \quad (71)$$

$$F_{4:3} = \frac{-x\beta_1\beta_2e^{-\beta_3x}}{(1+\beta_2e^{-\beta_3x})^2} = \frac{-60 \cdot 22 \cdot 5000 \cdot e^{-0,08 \cdot 60}}{(1+5000e^{-0,08 \cdot 60})^2} = 86,75114 \quad (72)$$

A partir da matriz $\mathbf{F}(\hat{\theta}_0)$ das derivadas, calculamos o vetor de incrementos $\Delta\hat{\theta}_0$, que é dado por

$$\Delta\hat{\theta}_0 = \left[\mathbf{F}(\hat{\theta}_0)^t \mathbf{F}(\hat{\theta}_0) \right]^{-1} \mathbf{F}(\hat{\theta}_0)^t \mathbf{E}_0 = \begin{bmatrix} 4,39277 \\ 1628,338 \\ 0,008975 \end{bmatrix} \quad (73)$$

Portanto, a nova estimativa atualizada $\hat{\theta}_1$ é dada por:

$$\hat{\theta}_1 = \hat{\theta}_0 + \Delta\hat{\theta}_0 = \begin{bmatrix} 20 \\ 5000 \\ 0,08 \end{bmatrix} + \begin{bmatrix} 4,39277 \\ 1628,338 \\ 0,008975 \end{bmatrix} = \begin{bmatrix} 24,39277 \\ 6628,338 \\ 0,088975 \end{bmatrix} \quad (74)$$

Como a soma de quadrados residual $SQE(\hat{\theta}_1) = 1,105184$ é menor que $SQE(\hat{\theta}_0) = 67,52282$, considera-se o vetor $\hat{\theta}_1$ como sendo a estimativa atualizada de $\hat{\theta}_0$. A partir daí, uma nova iteração é efetuada considerando $\hat{\theta} = \hat{\theta}_1$ e assim até que o modelo atinja a convergência.

A Tabela 3 apresenta as estimativas atualizadas após algumas iterações:

Tabela 3. Estimativas dos parâmetros e soma de quadrados residuais nas iterações do algoritmo de Gauss-Newton para ajuste do modelo Logístico

k	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$S(\hat{\theta}_k)$
0	20	5000	0,08	67,52282
1	24,39277	6628,33800	0,08896	1,05018
2	24,60133	8573,13200	0,09033	0,33207
3	24,56699	9690,20100	0,09133	0,28112
4	24,56023	9927,38100	0,09150	0,27862
5	24,55859	9964,50900	0,09154	0,27861
6	24,55851	9967,24400	0,09154	0,27861
7	24,55850	9967,44100	0,09154	0,27861
8	24,55850	9967,45400	0,09154	0,27861
9	24,55850	9967,45500	0,09154	0,27861

3.4 Avaliadores da qualidade de ajuste

Os avaliadores da qualidade de ajuste de modelos de regressão não linear que foram utilizados no presente trabalho para a abordagem Frequentista foram: percentual de convergência (C %), quadrado médio do erro (QME), coeficiente de determinação (R^2), critério de informação de Akaike (AIC), critério de informação Bayesiano (BIC) e desvio médio absoluto (DMA). Já para a abordagem Bayesiana, foi utilizado o DIC (Deviance Information Criterion).

3.5 Análise de agrupamento

Depois de escolher um melhor modelo em cada abordagem, com o objetivo de agrupar os acessos mais semelhantes, as estimativas dos parâmetros foram agrupadas por meio do algoritmo de Ward. Proposto por Ward (1963), o método baseia-se na mudança de variação dentro dos grupos em formação e entre eles a cada passo do processo de agrupamento. Tal método tem como objetivo a minimização da soma de quadrados dentro dos grupos.

3.6 Dissimilaridade entre os acessos

A medida utilizada foi a Distância Generalizada de Mahalanobis, pelo fato de levar em consideração a existência de correlações entre os caracteres analisados e também devido às diferenças de escala entre as variáveis.

As distâncias de Mahalanobis foram obtidas por meio da expressão

$$d_{rs}^2 = \|y_r - y_s\|_{\Psi}^2 = (y_r - y_s)^T \Psi (y_r - y_s) \quad (75)$$

3.7 Determinação do número ótimo de grupos

Para determinar o número de grupos na análise de agrupamento foi utilizado o procedimento sugerido por Mojena (1977). Tal método baseia-se no tamanho relativo dos níveis de fusões no dendrograma para determinar o ponto de corte dos dendrogramas gerados pelos métodos hierárquicos, para definir o número de grupos.

A proposta é selecionar o número de grupos no passo j que, primeiramente, satisfizer a seguinte inequação:

$$\alpha_j > \theta_k \quad (76)$$

Em que α_j é o valor de distância do nível de fusão correspondente ao passo j ($j=1, 2, \dots, g-1$), e θ_k é o valor referencial de corte, dado por:

$$\theta_k = \bar{\alpha} + k\hat{\sigma}_{\alpha} \quad (77)$$

em que $\bar{\alpha}$ e $\hat{\sigma}_{\alpha}$ são respectivamente a média e o desvio padrão dos valores de α ; k é uma constante cujo valor adotado, conforme sugerido por Milligan e Cooper (1985) foi de $k = 1,25$ como regra de parada na definição do número de grupos.

3.8 Identidade de modelos

Tendo em vista a classificação dos acessos em grupos pela análise de agrupamento, foram estimadas equações para representar as curvas referentes a cada grupo de acesso. Dessa forma, o próximo passo consistiu em testar as hipóteses referentes a igualdade dos parâmetros destes modelos em relação aos grupos formados e, para a abordagem Frequentista, utilizou-se o método de identidade de modelos de regressão não-linear apresentado por Regazzi (2010).

Para ilustração de tal método, considerou-se o modelo Logístico indicado na Tabela 1 acrescido de uma variável indicadora (dummy) cujo objetivo é representar cada um dos i grupos em estudo, e consideremos $i=3$ grupos. Tal modelo, denominado a partir de agora de completo, é o seguinte:

$$y_{ik} = \sum_{k=1}^3 D_k \left[\frac{\beta_{1k}}{1 + 3\beta_{2k} e^{(-\beta_3 x_i)}} \right] + \varepsilon_i, \quad (78)$$

em que:

D_k é uma variável indicadora, tal que: $D_k = \begin{cases} 1 & \text{se a observação } y_{ik} \text{ pertence ao grupo } k; \\ 0 & \text{em caso contrário} \end{cases}$;

β_{1k}, β_{2k} e β_{3k} são os parâmetros do modelo para cada um dos 3 grupos, sendo $k=1, 2$ e 3 ;

ε_i é o termo de erro aleatório, $\varepsilon_i \sim N(0, \sigma^2)$.

A fim de verificar se as estimativas dos parâmetros para os grupos de acessos são iguais, foi realizado um teste F, no qual foi avaliada a hipótese de que o modelo reduzido ajustado para ambos os grupos é idêntico aos modelos completos ajustados. Assim:

$$F(H_0) = \frac{\frac{SQRR(\omega) - SQRR(\Omega)}{t(r-1)}}{\frac{SQRR(\Omega)}{N - H_p - H(r-1)}} \quad (79)$$

Onde: SQRR representa a soma de quadrados do resíduo da regressão (falta de ajustamento mais o resíduo) para dado modelo, Ω é espaço paramétrico para o modelo completo, ω é o espaço paramétrico para o modelo reduzido sob H_0 , t é o número de parâmetros a serem testados, e N é o número total de observações.

As hipóteses consideradas foram as seguintes:

$H_0^{(1)} : \beta_{1(1)} = \beta_{1(2)} = \beta_{1(3)} = \beta_1$ vs $H_A^{(1)} : \text{nem todos } \beta_{1k} \text{ são iguais}$

$H_0^{(2)} : \beta_{3(1)} = \beta_{3(2)} = \beta_{3(3)} = \beta_3$ vs $H_A^{(2)} : \text{nem todos } \beta_{3k} \text{ são iguais}$

$H_0^{(3)} : \beta_{1(1)} = \beta_{1(2)} = \beta_{1(3)} = \beta_1$ e $\beta_{3(1)} = \beta_{3(2)} = \beta_{3(3)} = \beta_3$ vs $H_A^{(3)} : \text{pelo menos uma igualdade é uma desigualdade.}$

Para este estudo foram consideradas apenas as estimativas com interpretação prática $\hat{\beta}_1$ e $\hat{\beta}_3$ e foi utilizado o seguinte esquema de comparação, o qual é ilustrado para o teste da hipótese $H_0^{(1)}$.

Consideremos o seguinte modelo completo:

$$y_i = D_1 \frac{\beta_{11}}{1 + \beta_{21} e^{(-\beta_{31} x_i)}} + D_2 \frac{\beta_{12}}{1 + \beta_{22} e^{(-\beta_{32} x_i)}} + D_3 \frac{\beta_{13}}{1 + \beta_{23} e^{(-\beta_{33} x_i)}} + \varepsilon_i \quad (80)$$

e também o seguinte modelo reduzido:

$$y_i = D_1 \frac{\beta_1}{1 + \beta_{21} e^{(-\beta_{31} x_i)}} + D_2 \frac{\beta_1}{1 + \beta_{22} e^{(-\beta_{32} x_i)}} + D_3 \frac{\beta_1}{1 + \beta_{23} e^{(-\beta_{33} x_i)}} + \varepsilon_i \quad (81)$$

no qual é assumido um único parâmetro β_1 para representar os 3 grupos considerados.

As somas de quadrados dos resíduos provenientes dos ajustes dos modelos (80) e (81) são representadas respectivamente por SQR_Q e SQR_ω .

Como estamos considerando três grupos nesta ilustração, a rejeição da hipótese $H_0^{(1)}$ em questão, apenas nos permite inferir que existe pelo menos uma diferença entre as estimativas do parâmetro β_1 , assim, uma vez constatada tal rejeição, deve-se realizar uma nova de sequência de comparações duas a duas para este parâmetro. Como exemplo para a comparação entre os grupos 1 e 2, temos os seguintes modelos completo e reduzido, dados respectivamente pelas expressões (82) e (83).

$$y_i = D_1 \frac{\beta_{11}}{1 + \beta_{21} e^{(-\beta_{31} x_i)}} + D_2 \frac{\beta_{12}}{1 + \beta_{22} e^{(-\beta_{32} x_i)}} + \varepsilon_i \quad (82)$$

$$y_i = D_1 \frac{\beta_1}{1 + \beta_{21} e^{(-\beta_{31} x_i)}} + D_2 \frac{\beta_1}{1 + \beta_{22} e^{(-\beta_{32} x_i)}} + \varepsilon_i \quad (83)$$

Para a abordagem Bayesiana, os testes referentes a igualdade dos parâmetros dos modelos em relação aos grupos formados foram realizados por meio de uma modelagem hierárquica Bayesiana, no qual as amostras das distribuições marginais a posteriori para as diferenças entre as estimativas dos parâmetros β_1 e β_3 para cada um dos grupos considerados foram obtidas com o intuito de verificar significância estatística por meio dos intervalos de credibilidade de máxima densidade a posterior (HPD) de 95%. Assim as diferenças apresentam-se como um parâmetro a mais no modelo, e permitem testar a hipótese de igualdade dos parâmetros mediante avaliação do HPD, pois se o valor zero estiver contido neste intervalo, conclui-se que os parâmetros das duas populações envolvidas no contraste são estatisticamente iguais. Esta metodologia foi apresentada por Silva et al. (2005) para comparar parâmetros de curvas de lactação de cabras referentes a duas diferentes populações, e posteriormente utilizada para comparar curvas de crescimento de gado Nelore de diferentes grupos genéticos (Silva et al., 2007).

4 RESULTADOS E DISCUSSÃO

4.1- Ajuste dos modelos de regressão não linear via abordagem Frequentista.

Na Tabela 4, estão apresentados a média e o desvio padrão dos avaliadores da qualidade de ajuste dos sete modelos apresentados na Tabela 1 para os 30 acessos de alho avaliados. Pode-se observar que a classificação dos modelos foi a mesma, considerando qualquer um dos avaliadores, ou seja, os modelos que tiveram os menores valores para o quadrado médio do erro (QME), critério de informação de Akaike (AIC), critério de

informação Bayesiano (BIC) e desvio médio absoluto dos resíduos (DMA), também foram os que apresentaram maior valor para o coeficiente de determinação (R^2).

Tabela 4. Médias do coeficiente de determinação (R^2), quadrado médio do resíduo (QMR), critério de informação de Akaike (AIC), critério de informação Bayesiano (BIC) e desvio médio absoluto dos resíduos (DMA) dos modelos apresentados na Tabela 1, para as médias da matéria seca total da planta (MSTP) dos 30 acessos de alho avaliados, e a respectiva porcentagem de convergência de cada modelo

Modelos	Parâmetros	R^2	QMR	AIC	BIC	DMA	CONV(%)
Brody	Média	0,9159	26,0962	26,2306	23,7758	2,2019	50
	sd	0,0244	14,3910	2,4637	2,4637	0,6277	-----
Gompertz	Média	0,9963	1,0304	13,2758	10,8210	0,2446	100
	sd	0,0022	0,5727	2,4647	2,4647	0,0700	-----
Logístico	Média	0,9979	0,6212	10,7817	8,3268	0,2372	100
	sd	0,0014	0,4709	3,1615	3,1615	0,0857	-----
Meloun I	Média	0,9135	27,0877	26,4030	23,9482	2,2476	46,67
	sd	0,0234	14,3926	2,4610	2,4610	0,6251	-----
Meloun II	Média	0,9135	27,0877	26,4030	23,9482	2,2476	46,67
	sd	0,0234	14,3926	2,4610	2,4610	0,6251	-----
Mitscherlich	Média	0,9135	27,0877	26,4030	23,9482	2,2476	46,67
	sd	0,0234	14,3926	2,4610	2,4610	0,6251	-----
Von	Média	0,9772	7,0592	20,8235	18,3686	1,2216	100
Bertalanffy	sd	0,0106	5,0048	2,5050	2,5050	0,4100	-----

O modelo que melhor se ajustou aos dados foi o Logístico, seguido em ordem decrescente dos modelos Gompertz, Von Bertalanffy, Brody, Meloun I, Mitscherlich e Meloun II, respectivamente. Este resultado, para o melhor modelo, está de acordo com o encontrado por Reis (2014), ao estudar modelos de regressão não linear para descrever o acúmulo de matéria seca de partes da planta do alho, onde constatou que o modelo que melhor se ajustou aos dados, para todas as partes da planta de alho, foi o modelo Logístico. Resultado semelhante também foi encontrado por Puiatti (2013), ao identificar e agrupar modelos de regressão não linear que melhor se ajustam na descrição do acúmulo de matéria seca total da planta do alho ao longo do tempo, sendo o Logístico o melhor

modelo dentre os modelos Brody, Gompertz Logístico, Meloun I, Meloun II, Meloun III, Mitscherlich e Von Bertalanffy.

Em vários trabalhos envolvendo modelos de regressão não linear para descrição de curvas de crescimento ou acúmulo de nutrientes, diversos autores concluíram que o modelo Logístico obteve ótimos ajustes. Como, por exemplo, Pôrto et al. (2007) encontrou ótimos ajustes para o modelo Logístico para descrever acúmulos de matéria seca da parte aérea do bulbo e da planta inteira da cultura de cebola. Maia et al. (2009) também identificou o modelo Logístico como sendo o de melhor ajuste em estudo envolvendo bananeiras.

Dentre os modelos trabalhados, os únicos modelos nos quais houve convergência para todos os acessos foram os modelos Logístico, Gompertz e Von Bertalanffy. Já para o modelo Brody, houve convergência para apenas metade dos acessos. Isto deve-se ao fato de que o modelo Brody não possui ponto de inflexão, o que faz com que ele deixe de ser um bom representante para este tipo de estudo, uma vez que o comportamento dos dados em estudo possui um formato sigmoidal. O ponto de inflexão indica a variação máxima de acúmulo de matéria seca da planta e que, em seguida, há um decréscimo na acumulação de matéria seca, não indicando que a planta para de acumular matéria seca, e sim que ganha menos matéria seca a cada unidade de tempo que passa a partir daquele ponto.

Na Tabela 5 estão apresentadas a média, erro padrão e coeficiente de variação das estimativas dos parâmetros para os modelos apresentados na Tabela 1. Com relação a estimativa $\hat{\beta}_1$ (peso assintótico), vemos que os modelos Brody, Meloun I, Meloun II e Mitscherlich não se mostraram bons representantes para curvas de acúmulo de matéria seca total desses acessos utilizados, uma vez que o valor dessas estimativas foi muito alta (acima de 140 g), ficando fora da realidade dos dados. Como citado anteriormente, o fato do modelo Brody não ser um bom representante pode ser pelo fato de não possuir o ponto de inflexão. Segundo Reis (2014), as estimativas $\hat{\beta}_1$ e $\hat{\beta}_3$ dos modelos Brody, Meloun I, Meloun II e Mitscherlich são, aproximadamente, iguais. Isso pode ser que ocorra devido ao fato de que esses quatro modelos são apenas reparametrizações diferentes de um mesmo modelo em que as mudanças nos valores das estimativas só acontecem para o parâmetro β_2 . Estes modelos foram os que tiveram a pior classificação quanto à qualidade de ajuste (Tabela 4). Por outro lado, os modelos Logístico, Gompertz e Von

Bertalanffy, mostraram-se bons representantes para este tipo de estudo, com as estimativas para o parâmetro β_1 bem próximas da realidade dos dados.

Tabela 5. Média, erro padrão e coeficiente de variação das estimativas $\hat{\beta}_1$, $\hat{\beta}_2$ e $\hat{\beta}_3$ dos parâmetros para os modelos apresentados na Tabela 1

Modelos	Parâmetros	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
Brody	Média	170,3897	1,1610	0,00244
	erro padrão	139,7598	0,1039	0,00149
	cv(%)	82,02	8,94	61,27
Gompertz	Média	21,0751	7884,12	0,0769
	erro padrão	3,5004	22999,74	0,0154
	cv(%)	16,60	291,72	20,00
Logístico	Média	20,8744	320524,52	0,104
	erro padrão	3,4443	1013021,58	0,0161
	cv(%)	16,49	316,05	15,54
Meloun I	Média	140,6217	158,13	0,00258
	erro padrão	82,0006	82,17	0,00143
	cv(%)	58,31	51,96	55,51
Meloun II	Média	140,6118	-4,8588	0,00273
	erro padrão	82,0016	0,5076	0,00138
	cv(%)	58,31	-10,44	54,75
Mitscherlich	Média	140,6178	59,9116	0,00258
	erro padrão	81,9911	1,3775	0,00143
	cv(%)	58,30	2,29	55,51
Von Bertalanffy	Média	24,2014	12,3512	0,0382
	erro padrão	4,2003	1,9198	0,00257
	cv(%)	17,35	15,54	6,73

A estimativa do parâmetro β_2 foi bem diferente para todos os modelos, sendo o menor valor de -4,8588 para o modelo Meloun II e o maior valor de 320524,522 para o modelo Logístico: porém isto não representa problema de ajuste, uma vez que tal parâmetro não tem interpretação biológica, é apenas um parâmetro de locação.

Os modelos Gompertz, Von Bertalanffy, Brody, Meloun I, Mitscherlich e Meloun II tiveram os valores para a estimativa do parâmetro β_3 (velocidade de crescimento) bem próximos, sendo estes, os menores valores. O modelo Logístico foi o modelo com maior valor para tal a estimativa, mostrando uma forma sigmoideal bem mais definida.

4.1.1- Agrupamento dos acessos pelas estimativas dos parâmetros.

Na Tabela 6 estão apresentadas as estimativas dos parâmetros β_1 , β_2 e β_3 do modelo que melhor se ajustou aos dados (modelo Logístico) para cada um dos 30 acessos considerados.

Tabela 6. Estimativas dos parâmetros β_1 , β_2 e β_3 para o modelo Logístico para cada um dos 30 acessos de alho considerados

n	Acesso	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	n	Acesso	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
1	492	18,60	53156,38	0,108	16	5366	22,52	4462961	0,152
2	4493	21,56	3563292	0,149	17	5939	18,85	18402,2	0,096
3	4567	26,77	358525,6	0,125	18	5940	16,60	62886,78	0,106
4	4662	23,20	44584,37	0,105	19	5943	20,90	11293,94	0,090
5	4814	18,10	46990,2	0,106	20	5944	21,69	144049	0,116
6	4815	20,13	77891,95	0,110	21	5950	21,86	45274,89	0,104
7	4818	23,96	23225,82	0,095	22	5959	18,42	29244,12	0,094
8	4820	19,20	14322,26	0,090	23	5960	22,53	61179,25	0,106
9	4823	24,55	9967,44	0,091	24	5964	19,35	105358,9	0,108
10	4825	20,41	13838,49	0,091	25	6390	27,35	138818,5	0,111
11	4830	16,82	19528,86	0,092	26	6396	20,29	98225,46	0,115
12	4832	19,57	15152,14	0,090	27	7619	11,43	6101,85	0,088
13	4835	23,59	25245,67	0,097	28	7620	19,87	100906,9	0,109
14	4839	28,49	19542,22	0,092	29	7621	20,19	6218,47	0,084
15	4843	21,46	21151,68	0,092	30	7624	17,82	18399,31	0,094

As estimativas $\hat{\beta}_1$ e $\hat{\beta}_3$ dos parâmetros do modelo Logístico foram agrupadas por meio do algoritmo de Ward, utilizando o método de Mojena (1977) com $k = 1,25$ para determinar o ponto de corte no dendrograma gerado (Figura 1), obtendo o número ótimo de três grupos, e o ponto de corte no dendrograma de 75,12, o que corresponde a 37,78% da distância máxima observada nos níveis de fusão. A estimativa $\hat{\beta}_2$ não foi considerada na análise de agrupamento, uma vez que não possui interpretação prática.

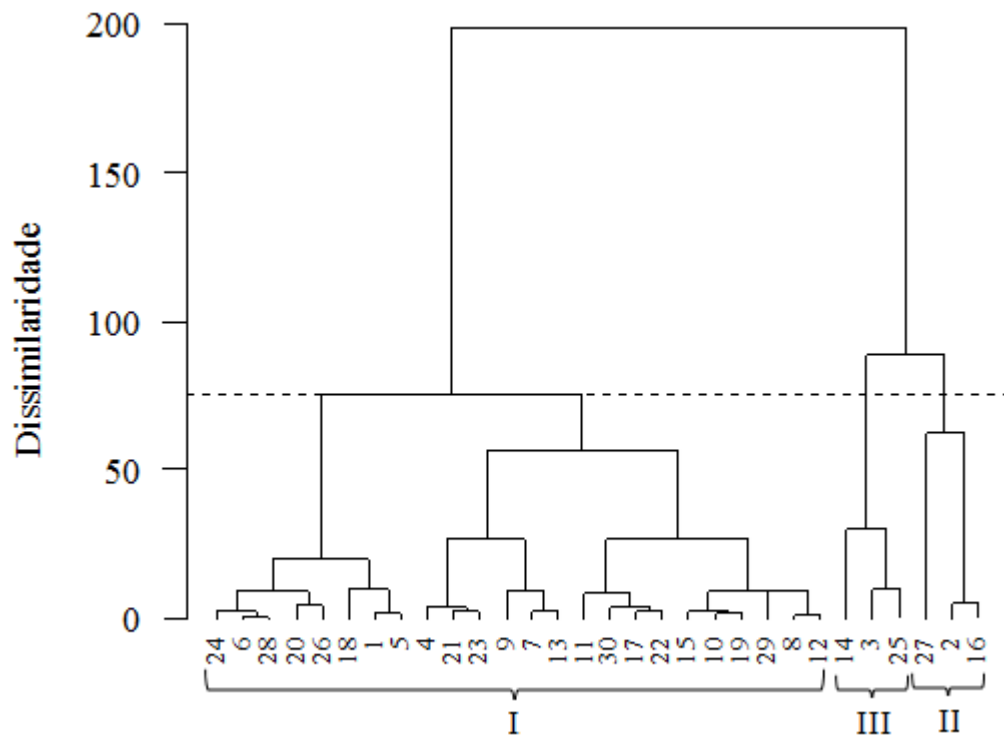


Figura 1. Dendrograma obtido com o agrupamento das estimativas dos parâmetros β_1 e β_3 do modelo Logístico para a abordagem Frequentista com os 30 acessos de alho avaliados.

Após a classificação dos acessos em grupos, e utilizando o modelo que melhor se ajustou aos dados (Logístico), com o objetivo de estudar o comportamento das curvas de cada grupo, ajustou-se uma curva para cada um dos grupos considerados. As estimativas dos parâmetros do modelo Logístico estão apresentadas na Tabela 7.

Tabela 7. Estimativas dos parâmetros β_1 , β_2 e β_3 para o modelo Logístico para cada um dos três grupos de acessos de alho

Grupo	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
I	20,3196	27442,4759	0,09859
II	18,4720	856629,5254	0,1364
III	27,4964	79138,0646	0,1075

Após a divisão dos acessos em grupos, a identidade de modelos mostrou que duas equações são suficientes para representar os acessos utilizados, sendo uma equação (84) para os grupos I e II e outra equação (85) para o grupo III. Para isso foram testadas as hipóteses de que os modelos reduzidos (ω) ajustados para os três grupos de acessos são idênticos ao modelo completo (Ω).

Os testes de identidade de modelos foram realizados sobre os parâmetros com interpretação biológica: o peso assintótico (β_1) e a velocidade de crescimento (β_2). As estimativas dos parâmetros do modelo sem restrições no espaço paramétrico (modelo completo Ω), e dos modelos reduzidos segundo as restrições ω_1 , ω_2 e ω_3 , estão representadas na Tabela 8.

Tabela 8. Estimativas dos parâmetros do modelo completo (Ω) e dos modelos reduzidos (ω_1 , ω_2 e ω_3)

Parâmetros	Modelo			
	Ω	ω_1	ω_2	ω_3
$\hat{\beta}_1$			21,1733	20,8417
$\hat{\beta}_{1(1)}$	20,2022	20,2268		
$\hat{\beta}_{1(2)}$	19,0658	19,4041		
$\hat{\beta}_{1(3)}$	27,7398	27,3068		
$\hat{\beta}_2$	40145,0158	39820,1714	27334,0095	38986,010
$\hat{\beta}_3$		0,1022		0,1021
$\hat{\beta}_{3(1)}$	0,1024		0,0970	
$\hat{\beta}_{3(2)}$	0,1041		0,0968	
$\hat{\beta}_{3(3)}$	0,1008		0,1046	

Como o teste de $H_0^{(3)}$ foi significativo (p-valor < 0,01) (Tabela 9), pode-se concluir que pelo menos uma igualdade é uma desigualdade. Assim, a equação comum, cujas estimativas estão apresentadas na Tabela 8 (modelo ω_3), não pode ser adotada para os três grupos. Entretanto, a hipótese de que o valor do parâmetro β_3 é igual para os três grupos não foi rejeitada (p-valor > 0,01)

Tabela 9. Hipóteses avaliadas, valores da estatística do teste F, número de graus de liberdade e nível descritivo do teste (p-valor)

Hipóteses	G. L.	Fcalc	p-valor
$H_0^{(1)} : \beta_{1(1)} = \beta_{1(2)} = \beta_{1(3)} = \beta_1$ vs $H_A^{(1)}$	1 e 113	26,6	$3,31 \times 10^{-10}$
$H_0^{(2)} : \beta_{3(1)} = \beta_{3(2)} = \beta_{3(3)} = \beta_3$ vs $H_A^{(2)}$	1 e 113	0,39	0,6724
$H_0^{(3)} : \beta_{1(1)} = \beta_{1(2)} = \beta_{1(3)} = \beta_1$ e $\beta_{3(1)} = \beta_{3(2)} = \beta_{3(3)} = \beta_3$ vs $H_A^{(3)}$	2 e 113	16,6	$9,85 \times 10^{-10}$

Assim, foi realizada uma nova comparação dois a dois para a estimativa do parâmetro β_1 (Tabela 10) para verificar quais são estatisticamente iguais.

Tabela 10. Hipóteses avaliadas, valores da estatística do teste F, número de graus de liberdade e nível descritivo do teste (p-valor) para a estimativa $\hat{\beta}_1$

Hipóteses	G. L.	F calc	p-valor
$H_0^{(1)} : \beta_{1(1)} = \beta_{1(2)} = \beta_1$ vs $H_A^{(1)}$	2 e 103	1,0439	0,3093
$H_0^{(2)} : \beta_{1(1)} = \beta_{1(3)} = \beta_1$ vs $H_A^{(2)}$	2 e 103	72,54	$1,433 \times 10^{-13}$
$H_0^{(3)} : \beta_{1(2)} = \beta_{1(3)} = \beta_1$ vs $H_A^{(3)}$	2 e 19	14,18	0,001308

Segundo as hipóteses avaliadas na Tabela 10, apenas as estimativas do peso assintótico dos grupos I e II são iguais (p-valor < 0,01). Assim, pode-se concluir que as equações dos grupos I e II não diferem significativamente. Deste modo, estes dois grupos podem ser representados pelo modelo reduzido ω_1 apresentado na Tabela 11.

Tabela 11. Estimativas dos parâmetros do modelo completo (Ω) e do modelo reduzido (ω) para os grupos I e II

Parâmetros	Modelo	
	Ω	ω
$\hat{\beta}_1$		20,1022
$\hat{\beta}_{1(1)}$	20,1980	
$\hat{\beta}_{1(2)}$	19,3766	
$\hat{\beta}_2$	35050,3640	35386,9918
$\hat{\beta}_3$	0,1013	0,1014

Logo, a equação comum, apresentada pelo modelo ω (Tabela 11), pode ser adotada para representar estes dois grupos de acessos.

O modelo escolhido é o com três parâmetros β_1 , β_2 e β_3 sendo a equação ajustada para representar os grupos I e II dada por:

$$\hat{y} = \frac{20,1022}{1 + 35386,99 e^{(-0,10x)}} \quad (84)$$

O modelo para representar o grupo III é o com três parâmetros β_1 , β_2 e β_3 , sendo a equação ajustada dada por:

$$\hat{y} = \frac{27,3068}{1 + 39820,1714 e^{(-0,10x)}} \quad (85)$$

Por meio destas duas equações ajustadas, pode-se concluir que a média do peso assintótico dos acessos pertencentes ao grupo III, que é de 27,3068 g, é superior aos demais acessos, que tem em média 20,1022 g, evidenciando que tais variedades possuem maior acúmulo de matéria seca aos 150 DAP. Quanto à estimativa do parâmetro β_3 , não houve diferenças significativas, o que mostra que, teoricamente, todas as variedades avaliadas possuem mesma taxa de crescimento.

Na Figura 2 estão representadas as curvas de acúmulo de matéria seca total de planta de alho ajustada pelo modelo Logístico para cada uma das duas equações ajustadas. Pode-se observar que, independentemente do tempo, a curva do grupo III é superior à curva dos grupos I e II. Nota-se, que as duas curvas atingem o peso assintótico por volta dos 130 DAP, ou seja, após esse tempo o peso tende à estabilização, mostrando ao pesquisador, que poderá colher as plantas aos 130 DAP sem perdas significativas de produtividade.

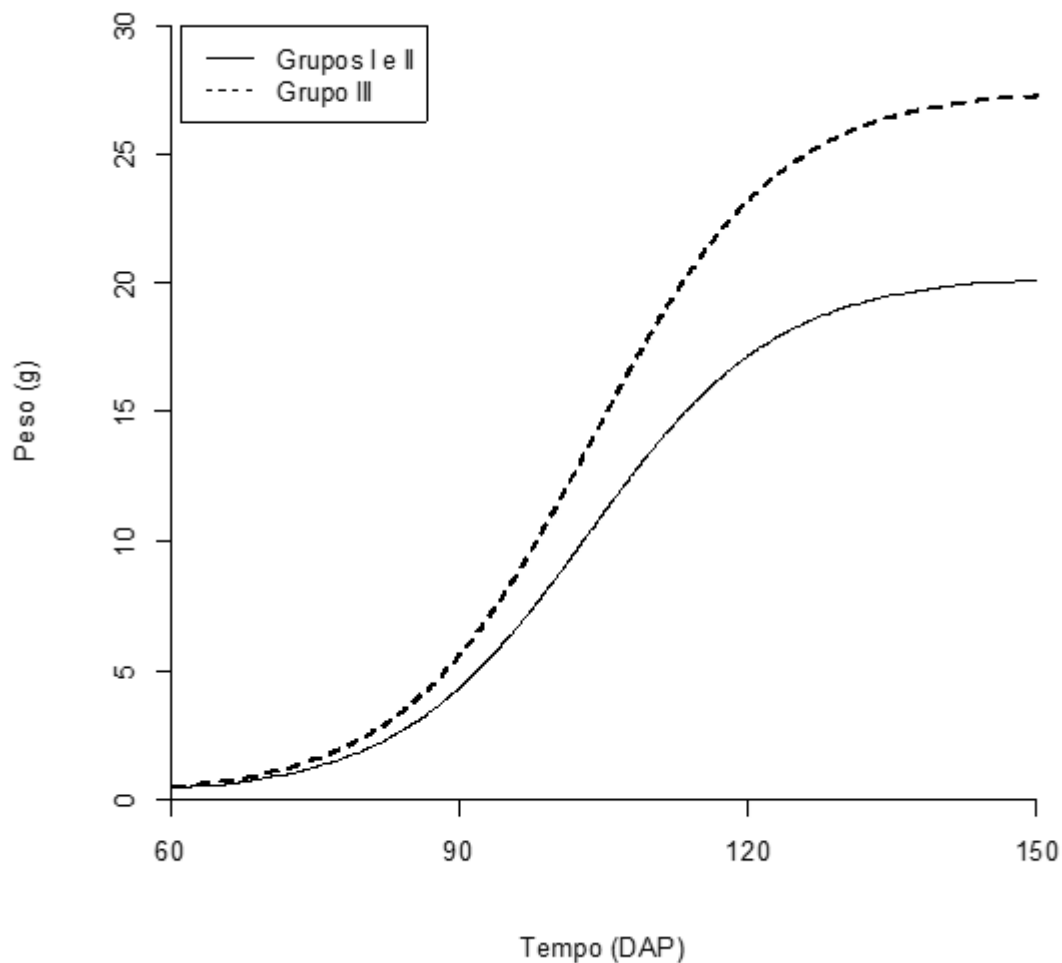


Figura 2. Curvas de acúmulo de matéria seca total de planta de alho ajustada pelo modelo Logístico para a abordagem Frequentista para cada um dos dois grupos considerados.

4.2- Ajuste dos modelos de regressão não linear via abordagem Bayesiana.

Por meio dos testes de convergência de Raftery & Lewis, Heidelberger & Welch e Geweke, pode-se dizer que a convergência foi atingida para todos os modelos, pois o período de aquecimento (burn in), o número total de iterações e o espaço entre as amostragens (thin) usados foram superiores ao mínimo recomendado pelo teste de Raftery & Lewis. Além disso, todos os valores para o fator de dependência (FD) do critério de Raftery & Lewis, que é responsável pelo acréscimo multiplicativo do número de iterações necessárias para a convergência, foram próximos de 1 e, de acordo com os autores, diz-se que a cadeia não atingiu a convergência se o fator de dependência for superior a cinco. O critério de Heidelberger & Welch demonstrou aceitação da hipótese nula de estacionariedade das cadeias, não sendo necessário maior número de iterações. Para o critério de Geweke (1992), todos os p-valores obtidos foram maiores que o nível de significância estabelecido de 0,05, não apresentando evidências contra a convergência

dos parâmetros. Na Tabela 12 estão representados os valores do critério de Geweke (p-valor) e Raftery & Lewis (fator de dependência-FD).

Tabela 12. Critérios de Geweke (valor p) e Raftery & Lewis (fator de dependência-FD)

Modelos	Parâmetros	Geweke (valor p)	Raftery & Lewis (FD)
Brody	$\hat{\beta}_1$	0,7471	0,9906
	$\hat{\beta}_2$	0,5405	0,9506
	$\hat{\beta}_3$	0,8426	0,9594
Gompertz	$\hat{\beta}_1$	0,3414	1,0080
	$\hat{\beta}_2$	0,5286	1,0109
	$\hat{\beta}_3$	0,6417	1,0371
Logístico	$\hat{\beta}_1$	0,8846	0,9506
	$\hat{\beta}_2$	0,4712	1,0109
	$\hat{\beta}_3$	0,8247	1,1390
Meloun I	$\hat{\beta}_1$	0,8728	0,9703
	$\hat{\beta}_2$	0,8459	1,0109
	$\hat{\beta}_3$	0,5991	1,0456
Meloun II	$\hat{\beta}_1$	0,6346	1,0320
	$\hat{\beta}_2$	0,3877	1,0456
	$\hat{\beta}_3$	0,4339	1,0407
Mitscherlich	$\hat{\beta}_1$	0,1527	0,9704
	$\hat{\beta}_2$	0,2805	1,0707
	$\hat{\beta}_3$	0,8486	1,1240
Von Bertalanffy	$\hat{\beta}_1$	0,1020	1,0285
	$\hat{\beta}_2$	0,9818	1,0109
	$\hat{\beta}_3$	0,3223	1,1660

Pelos valores do DIC (Deviance Information Criterion), apresentados na Tabela 13, observamos que o modelo Logístico foi o que melhor se ajustou aos dados, pois foi o modelo com menor valor para o DIC, concordando com o melhor modelo encontrado pela abordagem Frequentista no item 4.1 e também, com os autores Puitatti (2013) e Reis (2014). Depois do modelo Logístico, os modelos que melhor se ajustaram aos dados foram, respectivamente, os modelos Gompertz, Von Bertalanffy, Meloun II, Meloun I, Mitscherlich e Brody, diferindo da classificação encontrada pela abordagem Frequentista

apenas para os quatro últimos modelos. O modelo Brody também obteve para a abordagem Bayesiana o pior ajuste pois, como mencionado anteriormente, o fato deste modelo não possuir ponto de inflexão prejudica o ajuste.

Tabela 13. Valores do DIC para cada um dos sete modelos apresentados na Tabela 1

Modelos	Brody	Gompertz	Logístico	Meloun I	Meloun II	Mitscherlich	Von Bertalanffy
DIC	725	448,4	322,2	672,7	572,7	685,9	531,9

Na Tabela 14 estão representadas a média, desvio padrão, coeficiente de variação, Erro de Monte Carlo e intervalo de máxima densidade a posteriori (HPD) para as estimativas dos parâmetros dos modelos ajustado aos 30 acessos de alho avaliados, onde observa-se que todos os valores estimados tiveram menor coeficiente de variação do que os valores estimados pela abordagem Frequentista. Além disso, todas estimativas para o peso assintótico para um mesmo modelo na abordagem Frequentista superestimaram os valores encontrados pela abordagem Bayesiana.

Em relação a estimativa do parâmetro β_1 (peso assintótico), observamos que os modelos Brody, Meloun I, Meloun II e Mitscherlich, assim como na abordagem Frequentista, não se mostraram bons representantes para as curvas de acúmulo de matéria seca total desses acessos utilizados, uma vez que o valor dessas estimativas foi muito alta (acima de 34,9 g), ficando fora da realidade dos dados. Por outro lado, os modelos Logístico, Gompertz e Von Bertalanffy, mostraram-se bons representantes para este tipo de estudo, semelhante ao resultado encontrado pela abordagem Frequentista.

A estimativa do parâmetro β_2 teve uma grande variação entre os modelos, sendo o menor valor de -4,8615 para o modelo Meloun II e o maior valor de 144406,2 para o modelo Logístico. Porém, como mencionado anteriormente, isto não representa problema de ajuste.

Tabela 14. Média, desvio padrão, coeficiente de variação, Erro de Monte Carlo e intervalo de máxima densidade a posteriori (HPD) para as estimativas dos parâmetros dos modelos ajustado aos 30 acessos de alho avaliados

Modelos	Parâmetros	Média	Desvio padrão	CV(%)	Erro Monte Carlo	HPD	
						LI	LS
Brody	$\hat{\beta}_1$	34,9824	0,9812	2,80	0,0143	32,9600	36,8400
	$\hat{\beta}_2$	1,8635	0,0362	1,94	0,0006	1,7940	1,9360
	$\hat{\beta}_3$	0,0100	0,0001	0,94	0,000001	0,0098	0,0102
Gompertz	$\hat{\beta}_1$	19,6695	0,6556	3,33	0,0107	18,48	21,0100
	$\hat{\beta}_2$	130874,5	33248,3	25,40	450,23	71970	201600
	$\hat{\beta}_3$	0,1236	0,0032	2,62	0,00004	0,1177	0,1301
Logístico	$\hat{\beta}_1$	20,6847	1,0776	3,36	0,01599	18,53	22,72
	$\hat{\beta}_2$	144406,2	40218,1	31,99	634,1227	65840	223000
	$\hat{\beta}_3$	0,07001	0,0001	0,48	0,000002	0,06981	0,0702
Meloun I	$\hat{\beta}_1$	132,8127	20,2767	15,26	0,3174	94,4700	173,400
	$\hat{\beta}_2$	149,9596	20,6836	13,79	0,3241	108,4000	189,700
	$\hat{\beta}_3$	0,0020	0,00003	1,194	0,0000004	0,0020	0,0020
Meloun II	$\hat{\beta}_1$	114,4790	32,8594	28,70	0,4513	50,4000	179,60
	$\hat{\beta}_2$	-4,8615	0,1309	-2,69	0,0023	-5,1230	-4,6170
	$\hat{\beta}_3$	0,0027	0,00002	1,03	0,0000004	0,0026	0,0027
Mitscherlich	$\hat{\beta}_1$	91,7333	16,5554	18,04	0,2834	58,2000	123,600
	$\hat{\beta}_2$	59,5524	1,1950	2,00	0,0163	57,7200	61,8600
	$\hat{\beta}_3$	0,0030	0,0000	0,03	$\frac{0,0000000}{2}$	0,0030	0,0030
Von Bertalanffy	$\hat{\beta}_1$	24,1745	0,7687	3,07	0,0114	23,5100	26,4800
	$\hat{\beta}_2$	9,5419	0,7176	7,52	0,0105	8,0720	10,8600
	$\hat{\beta}_3$	0,0352	0,0002	0,59	0,000003	0,0348	0,0356

O modelo Gompertz obteve o maior valor para a estimativa do parâmetro β_3 (taxa de crescimento), enquanto que o modelo Meloun II teve o menor valor.

Comparando as estimativas dos parâmetros encontradas pela abordagem Frequentista com as encontradas pela abordagem Bayesiana, observamos que, para todos os modelos, as estimativas encontradas pela abordagem Frequentista foram superiores às

encontradas pela abordagem Bayesiana para um determinado modelo, sendo que nesta última, as estimativas tiveram menor coeficiente de variação.

4.2.1- Agrupamento dos acessos pelas estimativas dos parâmetros.

Na Tabela 15 estão apresentadas as estimativas dos parâmetros β_1 , β_2 e β_3 do modelo Logístico para cada um dos 30 acessos avaliados.

Tabela 15. Estimativas dos parâmetros β_1 , β_2 e β_3 para o modelo Logístico para cada um dos 30 acessos avaliados

n	Acesso	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	n	Acesso	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
1	492	19,64	262500	0,1213	16	5366	23,45	161600	0,1155
2	4493	22,6	235100	0,1161	17	5939	18,01	48530	0,1078
3	4567	26,2	332000	0,1237	18	5940	16,39	196000	0,119
4	4662	22,42	156600	0,1176	19	5943	21,57	298300	0,1201
5	4814	18,34	128100	0,1168	20	5944	19,12	173000	0,1226
6	4815	19,13	298400	0,1293	21	5950	20,75	54190	0,1093
7	4818	24,63	49920	0,1005	22	5959	17,46	160000	0,1111
8	4820	19,77	23470	0,09734	23	5960	20,99	130400	0,1136
9	4823	22,59	416100	0,1375	24	5964	20,14	173700	0,1081
10	4825	20,62	14010	0,09176	25	6390	27,16	94020	0,1094
11	4830	16,72	109800	0,1145	26	6396	19,46	164200	0,1221
12	4832	18,68	34360	0,09768	27	7619	10,59	87350	0,1131
13	4835	24,35	88180	0,1082	28	7620	20,15	109300	0,1101
14	4839	26,44	44050	0,1064	29	7621	19,79	1531	0,07126
15	4843	22,24	62820	0,09943	30	7624	18,16	262100	0,1153

As estimativas $\hat{\beta}_1$ e $\hat{\beta}_3$ dos parâmetros do modelo Logístico foram agrupadas por meio do algoritmo de Ward, utilizando o método de Mojena (1977) com $k = 1,25$ para determinar o ponto de corte nos dendrogramas gerados (Figura 3), obtendo o número ótimo de quatro grupos, e o ponto de corte no dendrograma de 66,37, o que corresponde a 47,08% da distância máxima observada nos níveis de fusão. Os grupos formados estão destacados na Figura 3, onde observa-se que o grupo I teve o maior número de acessos, enquanto o grupo IV foi constituído por apenas 2 acessos.

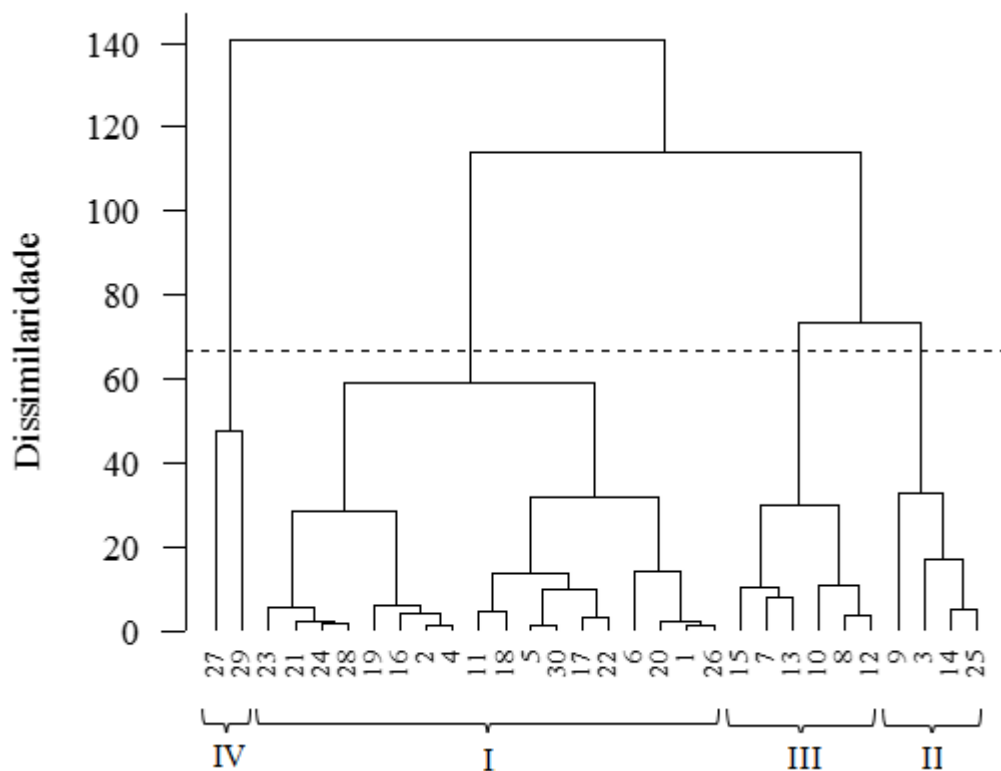


Figura 3. Dendrograma obtido com o agrupamento das estimativas do parâmetro β_1 e β_3 do modelo Logístico para a abordagem Bayesiana com os 30 acessos de alho avaliados.

Após a separação dos acessos em grupos, com o objetivo de estudar o comportamento das curvas de cada grupo, ajustou-se uma equação para cada um dos quatro grupos formados utilizando o modelo Logístico. As estimativas dos parâmetros estão apresentadas na Tabela 16, onde observa-se que as estimativas para os parâmetros β_1 e β_3 para o grupo I são bem próximas das estimativas para os parâmetros β_1 e β_3 para o grupo III.

Tabela 16. Estimativas dos parâmetros β_1 , β_2 e β_3 para o modelo Logístico para cada um dos quatro grupos de acessos

Grupo	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
I	20,1407	72690,22	0,1072
II	26,8186	151335,5855	0,1021
III	21,4016	17552,9877	0,0920
IV	15,4744	5770,5412	0,0853

Após realizar uma comparação entre as estimativas para os parâmetros β_1 e β_3 de cada grupo, obtiveram-se os resultados representados na Tabela 17.

Na Tabela 17 estão representados a média, o desvio padrão e o intervalo HPD para as diferenças entre as estimativas do peso assintótico entre todos os quatro grupos considerados na Tabela 16 e, também, as diferenças entre as estimativas da taxa de crescimento, onde observa-se que os grupos I e III não diferem estatisticamente um do outro, uma vez que o HPD contém o valor zero tanto para o peso assintótico quanto para a taxa de crescimento. As demais diferenças foram todas significativas ao nível de 95% pelo HPD.

Tabela 17. Média, desvio padrão e intervalo de máxima densidade a posteriori (HPD) para as diferenças entre as estimativas dos parâmetros β_1 e β_3 para cada grupo

Diferenças	Média	Desvio Padrão	HPD	
			LI	LS
$\beta_{1(1)} - \beta_{1(2)}$	-6,645	1,155	-8,93	-4,37
$\beta_{1(1)} - \beta_{1(3)}$	-0,973	0,858	-2,67	0,68
$\beta_{1(1)} - \beta_{1(4)}$	4,584	0,981	2,68	6,51
$\beta_{1(2)} - \beta_{1(3)}$	5,673	1,334	3,07	8,33
$\beta_{1(2)} - \beta_{1(4)}$	11,230	1,415	8,46	14,03
$\beta_{1(3)} - \beta_{1(4)}$	5,557	1,194	3,21	7,89
$\beta_{3(1)} - \beta_{3(2)}$	0,00522	0,000563	0,0042	0,0063
$\beta_{3(1)} - \beta_{3(3)}$	0,000367	0,000565	-0,0008	0,0014
$\beta_{3(1)} - \beta_{3(4)}$	0,0219	0,000411	0,02119	0,02277
$\beta_{3(2)} - \beta_{3(3)}$	-0,00485	0,000563	-0,0061	-0,0039
$\beta_{3(2)} - \beta_{3(4)}$	0,0167	0,000411	0,01598	0,01758
$\beta_{3(3)} - \beta_{3(4)}$	0,0216	0,000412	0,0208	0,02241

Portanto, os acessos podem ser classificados em três diferentes grupos, sendo os grupos I e III da Tabela 16 considerados como um único grupo e representados por uma única equação. Assim, a equação ajustada para representar os grupos I e III é dada por:

$$\hat{y} = \frac{20,4675}{1 + 54919,62 e^{(-0,1035x)}} \quad (86)$$

As equações ajustadas para representar os grupos II e IV são dadas, respectivamente pelas equações (87) e (88).

$$\hat{y} = \frac{26,8186}{1 + 151335,58 e^{(-0,1021x)}} \quad (87)$$

$$\hat{y} = \frac{15,4744}{1 + 17552,98 e^{(-0,092x)}} \quad (88)$$

Deste modo, pode-se concluir que a média do peso assintótico dos acessos pertencentes ao grupo II é superior aos demais grupos, o que mostra que tais variedades possuem maior acúmulo de matéria seca que os demais. Com exceção do acesso de número 9, os outros três acessos pertencentes ao grupo de maior peso assintótico para a abordagem Bayesiana foram os mesmos para a abordagem Frequentista. Já para o grupo de menor peso assintótico, os únicos acessos que se classificaram para este grupo em ambas as abordagens foi o acesso de número 27. O novo grupo formado pelos grupos I e III da Tabela 16 foi o que teve o maior número de acessos (24 acessos), e teve média de 20,46 g para o peso assintótico, valor esse bem próximo ao encontrado para o novo grupo da abordagem Frequentista formado pelos grupos I e II da Tabela 7, que foi de 20,10 g (grupo este que também teve o maior número de acessos). Quanto à taxa de crescimento, os grupos I e II obtiveram os maiores valores, já o grupo IV, o menor.

A abordagem Frequentista superestimou os valores dos parâmetros encontrados pela abordagem Bayesiana para os grupos formados, e a abordagem Bayesiana foi mais rigorosa, apontando maior número de grupos.

Na Figura 4 estão representadas as curvas de acúmulo de matéria seca total da planta de alho ajustadas pelo modelo Logístico para cada uma das três equações ajustadas, onde pode-se observar que, aproximadamente, a partir dos 125 DAP, a curva do grupo II é superior às demais curvas, enquanto que a do grupo IV é a menor dentre todas a partir dos 110 DAP. Isso evidencia que as variedades do grupo II são mais tardias que as dos grupos I, III e IV. Deste modo, se o produtor pretende por algum motivo colher as plantas antes dos 125 DAP, recomenda-se as variedades pertencentes aos grupos I e III; porém, se ele for colher após esta data, as variedades pertencentes ao grupo II irão lhe render maior peso, o que poderá resultar em maior lucro.

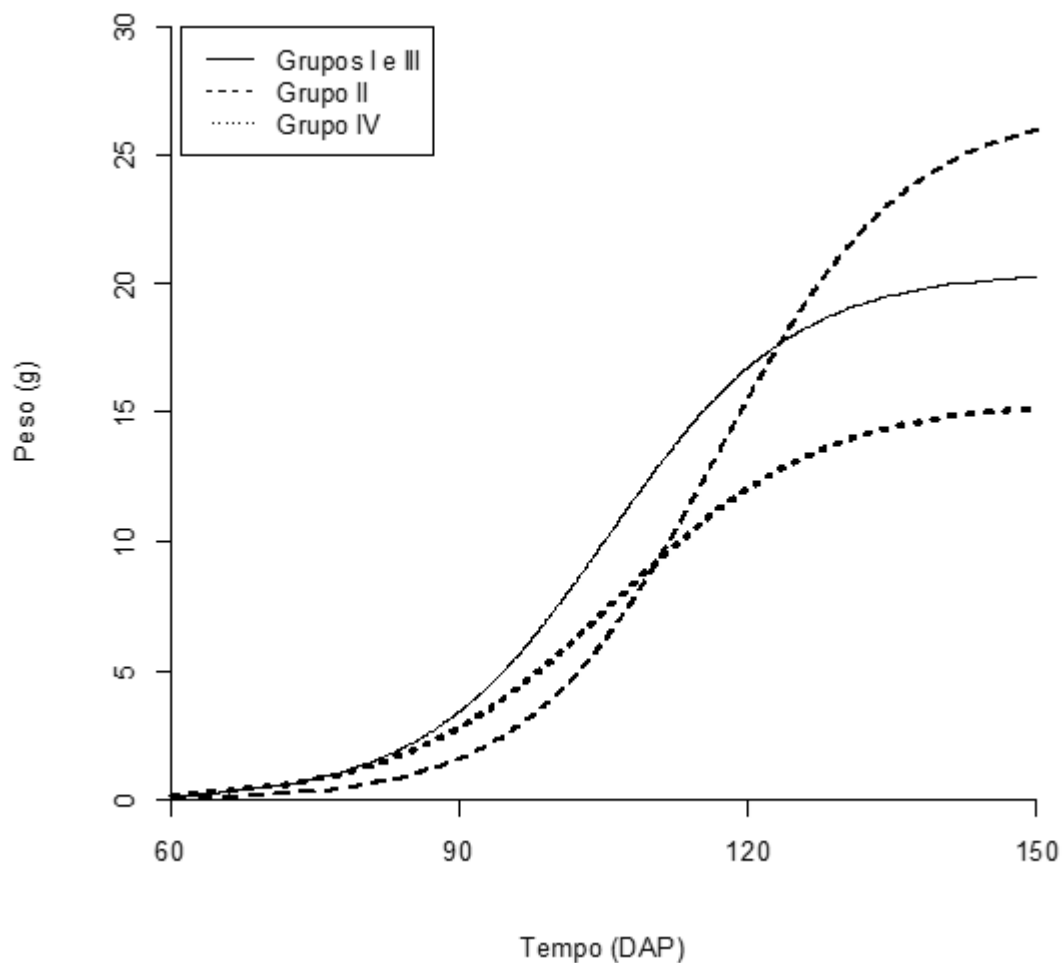


Figura 4. Curvas de acúmulo de matéria seca total de planta de alho ajustadas pelo modelo Logístico para a abordagem Bayesiana para cada um dos três grupos considerados.

5 CONCLUSÕES

O modelo Logístico foi o que melhor se ajustou aos dados de produção de matéria seca total de planta dos 30 acessos de alho utilizados neste trabalho.

A abordagem Bayesiana mostrou-se mais adequada por apresentar maior facilidade de convergência e por estimar parâmetros com menor variação, além de ser mais rigorosa quanto à classificação das variedades.

Os acessos 3, 14 e 25 apresentaram maior acúmulo de matéria seca aos 150 dias após o plantio em ambas abordagens, além destes, o acesso 9 também foi classificado para o grupo de maior acúmulo de matéria seca para a abordagem Bayesiana. Por outro lado, o acesso 27 apresentou um menor acúmulo de matéria seca aos 150 dias após o plantio em ambas abordagens e o acesso 29 também foi classificado para o grupo de menor acúmulo de matéria seca para a abordagem Bayesiana.

6 REFERÊNCIAS BIBLIOGRÁFICAS

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

BLASCO, A.; PILES, M.; VARONA, L. A Bayesian analysis of the effect of selection for growth rate on growth curves in rabbits. **Genetic Selection Evolution**, v.35, p.21-41, 2003.

BRODY, S. **Bioenergetics and Growth**. Reinhold Publishing, New York. 1945. 151p.

BUSSAB, W. DE O; MIAZAKI, E. S; ANDRADE, D. Introdução à análise de agrupamentos. São Paulo: **Associação Brasileira de Estatística**, 1990. 105p.

CHROBOK, V.; MELOUN, M.; SIMÁKOVÁ, E. Descriptive growth model of the height of stapes in the fetus: a histopathological study of the temporal bone. **Eur Arch Otorhinolaryngol**, v. 261, p.25-29, 2004.

CRUZ, C.D.; REGAZZI, A.J.; CARNEIRO, P.C.S. **Modelos Biométricos Aplicados ao Melhoramento Genético**. v. 1, 3 ed., Viçosa: UFV, 2004, 480p.

CRUZ, C.D.; FERREIRA, F.M.; PESSONI, L.A. **Biometria aplicada ao estudo da divergência genética**. Visconde do Rio Branco: Suprema, 2011, 620p.

FERREIRA, D.F. **Estatística Multivariada**. v. 1, 2 ed., Lavras: UFLA, 2011, 676 p. Filho, S.M.; Silva, F.F.; Carneiro, A.P.S.; Muniz, J. A. Abordagem Bayesiana das curvas de crescimento de duas cultivares de feijoeiro. **Ciência Rural**, v.38, n.6, p.1516-1521, 2008.

GALLANT, A. R. **Nonlinear statistical models**. [S.l.]: John Wiley & Sons, 1987.

GELFAND, A.E. Gibbs sampling. **Journal of the American Statistical Association**, v. 95, p. 1300-1304, 2000.

GELMAN, A.; RUBIN, D.B. Inference from iterative simulation using multiple sequences. **Statistical Science**, Hayward, v.7, n.4, p.457-511, 1992.

GEWEKE, J. **Evaluating the accuracy of sampling-based approaches to calculating posterior moments**. New York: Oxford University, 1992, 631 p.

GOMPERTZ, B. On the nature of the function of the law of human mortality and on a new model of determining the value of life contingencies. **Philos. Trans. Roy. Soc. London**. V. 115, p.513-585. 1825.

HEIDELBERGER, P.; WELCH, P. Simulation run length control in the presence of an initial transient. **Operation Research**, Maryland, v. 31, n. 6, p. 1109-44, 1983.

JEFFREYS, H. **Theory of probability**. Oxford, UK: Clarendon Press, 1961. 470 p.

JOHNSON, R.A.; WICHERN, D.W. **Applied Multivariate Statistical Analysis**. New Jersey, USA: Englewood Cliffs, 1992, 642p.

- KHATTREE, R.; NAIK, D.N. **Multivariate data reduction and discrimination with SAS software**. New York: BBU and John Wiley Sons, 2000. 574p.
- LUCINI, M.A. Alho roxo no Brasil: um pouco da história dos números desse nobre. **Revista Nosso Alho**, v. 1, n. 1, p. 16-21, 2008.
- MACEDO, L.R.; SILVA, F. F.; CIRILLO, M.A.; NASCIMENTO, M.; PAIXÃO, D.M.; GUIMARÃES, S.F.; LOPES, P.S.; SANTOS, J.A.; AZEVEDO, C.F. Modelagem hierárquica Bayesiana na avaliação de curvas de crescimento de suínos genotipados para o gene halotano. **Ciência Rural**, v. 44, n. 10, p. 1853-1859, 2014.
- MAHALANOBIS, P.C. On the generalized distance in statistics. **Proceedings of the National Institute of Science of India**. v.2, n. 1, p. 49–55. 1936.
- MAIA, E.; SIQUEIRA, D. L.; SILVA, F. F.; PETERNELLI, L. A.; SALOMÃO, L. C. C. Método de comparação de modelos de regressão não-lineares em bananeiras. **Ciência Rural**, v.39, n.5, p.1380-1386, 2009.
- MELOUN, M.; MILITKÝ, J. **Sbírka úloh, Statistické zpracování experimentálních dat**. Pardubice: Universita Pardubice. 1996. 308p.
- MILLIGAN, G. W.; COOPER, M. C. An examination of procedures for determining the number of cluster in a data set. **Psychometrika**, v. 50, p. 159-179, 1985.
- MITSCHERLICH, E.A. Das Gesetz des Pflanzenwachstums. **Landwirtsch Jahrb** v.53, p. 167-182. 1919.
- MOJENA, R. Hierárquical grouping method and stopping rules: an evaluation. **Computer Journal**, v. 20, n. 4, p. 359-363, 1977.
- MOTA, J.H.; YURI, J. E.; RESENDE, G. M.; SOUZA, R. J. Similaridade genética de cultivares de alho pela comparação de caracteres morfológicos, físico-químicos, produtivos e moleculares. **Horticultura Brasileira**, v. 24, n. 2, p. 156-160, 2006.
- MUELLER, S.; BIASI, J.; MENEZES SOBRINHO, J.A.; MULLER, J.J.V. Comportamento de cultivares de alho, plantio de junho. **Pesquisa Agropecuária Brasileira**, v.25, n.11, p.1561-1567. 1990.
- PAULINO, C.D.; TURKMAN, M.A.A.; MURTEIRA, B. **Estatística Bayesiana**. Lisboa: Editora Fundação Calouste Gulbenkian, 2003. 444p.
- PÔRTO, D.R. de Q.; CECILIO FILHO, A.B.; MAY, A.; VARGAS, P.F. Acúmulo de macro nutrientes pela cultivar de cebola Superex estabelecida por semeadura direta. **Ciência Rural**, v.37, p.949-955, 2007.
- PRUDENTE, A. A. Modelos não-lineares de regressão: alguns aspectos de teoria assintótica. 2009. 109 p. Dissertação (Mestrado em Modelagem Estatístico e Computacional) - Universidade Federal Rural de Pernambuco, Recife, 2009.
- PUIATTI, G.A.; CECON, P. R.; NASCIMENTO, M.; PUIATI, M.; FINGER, F. L.; SILVA, A. R.; NASCIMENTO, A. C. C. Análise de agrupamento em seleção de modelos

- de regressão não lineares para descrever o acúmulo de matéria seca em plantas de alho. **Revista Brasileira de Biometria**. v. 31, n. 3, p. 337-351, 2013.
- R CORE TEAM (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- RATKOWSKI, D.A. **Nonlinear regression modeling: a unified practical approach**. Marcel Dekker, New York. 1983. 276p.
- RAFTERY, A.E.; LEWIS, S. How many iterations in the Gibbs sampler? In: BERNARDO, J.M. et al. (Eds.). **Bayesian statistics**. Oxford, USA: University Press, p. 763-773. 1992.
- REGAZZI, A. J.; SILVA, C. H. O. Testes para verificar a igualdade de parâmetros e a identidade de modelos de regressão não-linear em dados de experimento com delineamento em blocos casualizados. **Revista Ceres**, v. 57, n. 3, p. 315-320, 2010.
- REGAZZI, A.J.; SILVA, C.H.O. Teste para verificar a igualdade de parâmetros e a identidade de modelos de regressão não linear. I. dados no delineamento inteiramente casualizado. **Revista Matemática e Estatística**, v.22, n.3, p.33-45, 2004.
- REIS, R. M.; CECON, P. R.; PUIATI, M.; FINGER, F. L.; NASCIMENTO, M.; SILVA, F. F.; CARNEIRO, A. P. S.; SILVA, A. R. Modelos de regressão não linear aplicados a grupos de acessos de alho. **Horticultura Brasileira**. v. 32, p. 178-183, 2014.
- ROSA, G.J.M. **Análise Bayesiana de modelos mistos robustos via amostrador de Gibbs**. 1998. 57 f. Tese (Doutorado em Estatística e Experimentação Agronômica)- Universidade de São Paulo, Piracicaba, 1998.
- SARMENTO, J. L. R.; REGAZZI, A. J.; SOUZA, W. H.; TORRES, R. A.; BREDAS, F. C.; MENEZES, G. R. O. Estudo da curva de crescimento de ovinos Santa Inês. **Revista Brasileira de Zootecnia**, v. 35, n. 2, p. 435-442, 2006.
- SILVA, F. F.; MUNIZ, J. A.; AQUINO, L. H.; SÁFADI, T. Abordagem Bayesiana da curva de lactação de cabras Saanen de primeira e segunda ordem de parto. **Pesquisa agropecuária brasileira**, Brasília, v.40, n.1, p.27-33, 2005.
- SILVA, N.A.M.; MUNIZ, J. A.; SILVA, F.F.; AQUINO, L.H. de; GONÇALVES, T.M. Aplicação do método bayesiano na estimação de curva de crescimento em animais da raça Nelore. **Ceres**, v. 54, n.312, p. 191-198, 2007.
- SILVEIRA, F.G.; SILVA, F.F.; CARNEIRO, P.L.S.; MALHADO, C.H.M.; MUNIZ, J.A. Análise de agrupamento na seleção de modelos de regressão não lineares para curvas de crescimento de ovinos cruzados. **Ciência Rural**, v.41, n.4, p.692-698, 2011.
- SOUZA, G.S. **Introdução aos modelos de regressão linear e não-linear**. 1 ed. Brasília: Embrapa, 1998. 489p.
- SOUSA, I. F. **Ajuste de modelos não lineares na descrição de germinação de sementes de café (Coffea arabica L.)**. 2012. 72f. Dissertação (Mestrado em Genética e Melhoramento) – Universidade Federal de Lavras. Lavras, 2012.

TRANI P.E. 2009. Cultura do alho (*Allium sativum*): Diagnóstico e recomendações para seu cultivo no Estado de São Paulo. Artigo em Hypertexto. Disponível em: <http://www.infobibos.com/Artigos/2009_2/alho/index.htm>. Acesso em: 25/2/2015.

VON BERTALANFFY, L. Quantitative laws for metabolism and growth. **Q. Rev. Biol.** v.32, p.217–231, 1957.

WARD, J.H. Hierarchical grouping to optimize an objective function. **Journal of the American Statistical Association**, v.58, p.236-244, 1963.

APÊNDICE

A- Script das análises no software R

```
#### Ajustes dos modelos de regressão não linear para a abordagem Frequentista

setwd("E:\\Tese\\dados")
dados=read.table("dados.txt",h=T) #lendo dados longitudinais (peso x idade)
head(dados)
  id x    y
1  1 60 0.68990
2  1 90 3.61250
3  1 120 10.62500
4  1 150 16.71250
5  2 60 0.85209
6  2 90 4.45000
#ajuste individual das curvas de crescimento
dados1=data.frame(id=as.numeric(dados$id), x=as.numeric(dados$x),
y=as.numeric(dados$y))
require(nlme)
dados2=groupedData(y~x|id,dados1) #agrupando os dados por ID
logist_pop=nls(y ~ a/(1+b*exp(-k*x)), data=dados1, start = list(a=22,b=4000,k=0.07))
#ajustando o modelo Logisitco para a pop
logist_pop
logist_id=nlsList(y ~ a/(1+b*exp(-k*x)), data=dados2, start = coef(logist_pop))
#ajustando o modelo Logisitco para cada ID
logist_id
logist_coef=coef(logist_id) # arquivo com as estimativas
logist_coefconv=na.omit(logist_coef) # eliminando IDs que não convergiram
id_converg=as.numeric(as.numeric(rownames(logist_coefconv)))
saida=logist_coef
write.table(saida,"saida_Logístico.txt",quote=FALSE,row.names=FALSE)

##### Avaliadores da qualidade do ajuste #####

#####porcentagem de convergencia#####
conv_logist=nrow(logist_coefconv)/nrow(logist_coef)*100 # %conv
conv_logist

##### QME #####

res=data.frame(matrix(t(matrix(id_converg,length(id_converg),4)),length(id_converg)*
4,1),unlist(as.matrix(residuals(summary(logist_id)))))
colnames(res)=c("id","res")
res=res[order(res$id),]
pred=cbind(as.numeric(rownames(as.matrix(predict(logist_id)))),predict(logist_id))
colnames(pred)=c("id","pred")
pred=na.omit(pred)
pred=pred[order(pred[,1]),]
outmodel=data.frame(res,abs(res[,2]),pred[,1],(pred[,2]+res[,2]),rep(c(1,1,2,2),length(id
_converg)))
colnames(outmodel)=c("id","res","absres","pred","yobs","part")
head(outmodel)
apply(outmodel,2,mean)
```

```

id<-outmodel$id
varmean<-data.frame(tapply(outmodel$res,id,var))
colnames(varmean)<-c("varres")
mean_qme<-mean(varmean$varres)
sd_qme<-sqrt(var(varmean))
mean_qme
sd_qme

##### R2 #####
#loop
aux<-unique(outmodel$id) #cria um vetor com valores unicos de id
r2<-NULL # matrix que vai receber os resultados
for(i in 1:length(aux)){
idx<-aux[i]
temp<-outmodel[outmodel$id==idx,]
cor<-cor(temp$pred,temp$yobs)^2 # cria uma linha de cor para cada id
r2<-rbind(r2,data.frame(aux[i],cor)) #unindo as linhas da matrix criada r2
}
mean(r2$cor)
sd(r2$cor)

##### AIC #####
teta=3
nobs=4
AIC = as.matrix(nobs*log(varmean)+2*teta)
mean(AIC)
sd(AIC)

## Ajustes dos modelos de regressão não linear para a abordagem Bayesiana

library(R2OpenBUGS)
setwd("E:\\Tese\\dados\\Modelos")

y=read.table("pesos_30.txt")
head(y)
  y1    y2    y3    y4
1 0.85209 4.4500 16.6875 18.45
2 1.28779 3.5750 20.4875 21.50
3 1.28353 4.8250 24.3375 26.65
4 1.63888 5.0750 20.4375 22.95
5 0.86209 4.1300 16.0250 17.95
6 0.72728 4.1375 17.7375 20.00
x=read.table("DAP_30.txt")
head(x)
  x1 x2 x3 x4
1 60 90 120 150
2 60 90 120 150
3 60 90 120 150
4 60 90 120 150
5 60 90 120 150
6 60 90 120 150

```



```

Nid=nrow(y)
N=ncol(y)

# especificando o modelo

linemodel<-function()
{
for (i in 1:Nid)
{
for (j in 1:N)
{
Y[i,j]~dnorm(mu[i,j],tau)
mu[i,j]<-a[i]/(1+b[i]*exp(-c[i]*X[i,j])) # modelo Logístico
}
a[i]~dnorm(mua,taua) # priori para ai
b[i]~dnorm(mub,taub) # priori para bi
c[i]~dnorm(muc,tauc) # priori para ki
}
mua~dnorm(20,4) # priori para mua
mub~dnorm(4000,10000) # priori para mub
muc~dnorm(0.07,0.0002) # priori para muk

taua~dgamma(06.25,10) # priori para taua
sigma2a<-1/sqrt(taua)

taub~dgamma(1200,1000000000) # priori para taub
sigma2b<-1/sqrt(taub)

tauc~dgamma(40000,100) #priori para tauc
sigma2c<-1/sqrt(tauc)

tau~dgamma(6.25,10) #priori para tau
sigma2<-1/sqrt(tau)
}

linedata<-list(Y=as.matrix(y),X=as.matrix(x),Nid=30,N=4)

# valores iniciais para MCMC.
lineinits=function(){list(a=rep(20,Nid),b=rep(4000,Nid),c=rep(0.07,Nid),mua=20,mub=
4000,muc=0.07,taua=0.01,taub=0.00000001,tauc=25000000,tau=0.01)}

# especificando o modelo
parametros=c("a","b","c","mua","mub","muc","sigma2a","sigma2b","sigma2c","sigma2")

## Ajustando o modelo usando OpenBUGS

lineout<-bugs(linedata,lineinits,parameters,linemodel,
n.chains=1,n.iter=5000,n.burnin=1000,n.thin=5)
#parametros estimados
lineout$summary
saida_a=lineout$last.values[[1]]$a

```

```
saida_b=lineout$last.values[[1]]$b
saida_c=lineout$last.values[[1]]$c

saida=cbind(saida_a, saida_b, saida_c)
saida
write.table(saida,"saida_Mo_Logístico.txt",quote=FALSE, row.names=FALSE)
lineout$DIC
```

B- Script das análises de agrupamento

```
### Distância Generalizada de Mahalanobis
```

```
dados=read.table("mahalanobis.txt",header=F)  
D2=as.matrix(dados)  
Colnames(D2)=c(1:30)  
matrizD2=as.dist(m=D2, diag=FALSE, upper=FALSE)
```

```
k=1.25  
Ward=hclust(matrizD2, method="ward")
```

```
PC= mean(ward$height)+k*sd(ward$height)
```

```
par(family="serif",las=1,cex=1.2)  
plot(ward,hang=-1,cex=0.7,main=" ",sub=" ",xlab=" ",ylab="Dissimilaridade")  
abline(h=PC,v=NULL,col=1,lty=2)
```

C- **Figura 1A-** Gráfico da densidade a posteriori para as estimativas $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$ e $\hat{\sigma}^2$ para os parâmetros do modelo Logístico.

