

Abordagem Bayesiana das curvas de crescimento de duas cultivares de feijoeiro

Bayesian approach in the growth curves of two cultivars of common bean

Sebastião Martins Filho^{1*} Fabyano Fonseca e Silva¹ Antonio Policarpo Souza Carneiro¹
Joel Augusto Muniz^{1†}

RESUMO

Neste trabalho foi utilizada a metodologia Bayesiana para ajustar o modelo não-linear logístico para dados de crescimento de duas cultivares de feijoeiro, "Neguinho" e "Carioca". O delineamento experimental utilizado foi o inteiramente casualizado, com vinte repetições, no esquema de parcelas subdivididas, sendo que os tratamentos principais foram constituídos pelas cultivares e as subparcelas foram constituídas por 17 períodos de avaliações, do plantio até aos 85 dias. A metodologia permitiu comparar as curvas de crescimentos sem utilizar a teoria assintótica e estes resultados mostraram um maior incremento em altura para a cultivar "Carioca".

Palavras-chave: *Phaseolus vulgaris*, inferência Bayesiana, modelos não-lineares, modelo logístico.

ABSTRACT

In this paper the Bayesian methodology was used to fit the logistic nonlinear model to growth data of two common bean cultivars, 'Neguinho' and 'Carioca'. The experiment was a split plot under a completely randomized design with twenty replicates, being the main treatments constituted by cultivars and the sub plots constituted by seventeen periods of evaluations, from planting to 85 days. The methodology allowed comparing the growth curves without using the asymptotic theory, and these results showed a larger height increment for the 'Carioca' cultivar.

Key words: *Phaseolus vulgaris*, Bayesian inference, nonlinear models, logistic model.

INTRODUÇÃO

Geralmente o estudo de curvas de crescimento de espécies vegetais tem sido conduzido

por meio de uma abordagem freqüentista, ajustando-se modelos não-lineares que buscam sintetizar as informações em poucas estimativas de parâmetros interpretáveis biologicamente. A estimação é fundamentada em processos iterativos, como o de Gauss-Newton, DUD e Algoritmo de Marquardt, devido à não-linearidade das variáveis. Usando-se estes procedimentos faz-se a minimização da soma de quadrados dos resíduos. Porém, quando se trata de ajustes individuais, ou seja, para várias unidades experimentais, de modelos matematicamente complexos ou se dispõe de poucas observações longitudinais, os métodos iterativos, muitas vezes, produzem estimativas irrealistas para os parâmetros. Isso pode levar à confecção de curvas de crescimento atípicas. Além disso, quando se trata de comparações de curvas provenientes de tratamentos diferentes, por exemplo, cultivares, na maioria das vezes, a distribuição dos estimadores dos parâmetros de modelos não-lineares não seguem distribuições gaussianas. Desse modo torna-se complexo o processo de formulação de testes estatísticos, pois não serão atendidas pressuposições relacionadas à utilização da teoria assintótica (SILVA et al., 2005).

Em recentes 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 utilizando poucas informações. A estimação por intervalo apresentou-se mais precisa em relação à obtida pela metodologia clássica e possibilitou

¹Departamento de Informática, Universidade Federal de Viçosa (UFV). Av. PH. Rolfs, 36570-000, Viçosa, MG, Brasil. E-mail: martinsfilho@ufv.br. *Autor para correspondência.

[†]Departamento de Ciências Exatas, Universidade Federal de Lavras (UFLA), Lavras, MG, Brasil.

a comparação estatística direta e simplificada entre curvas por meio da obtenção de uma distribuição para as diferenças entre parâmetros de duas curvas diferentes.

A teoria da inferência Bayesiana está fundamentada no teorema de Bayes, o qual é um resultado simples de probabilidade condicional. Para a sua utilização, é necessário especificar $p(\theta)$, que é conhecida como distribuição *a priori* do parâmetro θ . Tal distribuição é utilizada para representar probabilisticamente o conhecimento que se tem sobre θ antes dos dados serem obtidos. Outra exigência é a especificação de uma distribuição conjunta para os dados amostrais, denominada de função de verossimilhança, $L(\theta / y_1, \dots, y_n)$, a qual representa a informação sobre θ que foi obtida dos dados.

A expressão matemática do Teorema de Bayes é:

$$P(\theta | Y) = \frac{L(\theta | Y)P(\theta)}{\int L(\theta | Y)P(\theta)d\theta} \quad (1)$$

sendo $Y = \{y_1, y_2, \dots, y_n\}$.

Uma forma equivalente da expressão (1), visto que o denominador não depende de θ , é dada por: $P(\theta | Y) \propto L(\theta | Y)P(\theta)$ (2).

A expressão (2) é entendida como: “*Dist. Posteriori* \propto *Verossimilhança* \times *Dist. Priori*”.

A distribuição *a posteriori* de um parâmetro contém toda a informação probabilística a respeito do mesmo. Dessa forma, toda a inferência com respeito ao parâmetro é realizada por meio desta distribuição, pois é a partir de seus valores centrais que se têm as estimativas de interesse.

Para fazer inferências sobre qualquer elemento de θ , a distribuição *a posteriori* conjunta dos parâmetros, $p(\theta | Y)$, deve ser integrada em relação a todos os outros elementos que a constituem (ROSA, 1998). Assim, se o interesse do pesquisador se concentra sobre determinado conjunto de θ , por exemplo, θ_1 , tem-se a necessidade de obter a distribuição $p(\theta_1 | Y)$, dada por:

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

A integração da distribuição conjunta *a posteriori* para a obtenção das marginais geralmente não é analítica, necessitando de algoritmos iterativos especializados como o de Gibbs Sampler e o Metropolis Hastings, os quais são denominados de algoritmos MCMC (Markov Chain-Monte Carlo). Uma forma prática de implementar estes algoritmos é utilizar o software Winbugs (SPIEGELHALTER et al., 2007).

O objetivo do presente trabalho foi utilizar a metodologia Bayesiana para estimar os parâmetros do modelo de crescimento Logístico ajustado a dados longitudinais de altura de planta de duas cultivares de feijoeiro, “Neguinho” e “Carioca”. Esta metodologia foi usada com o intuito de obter distribuições *a posteriori* para a diferença dos parâmetros e assim realizar comparações diretas entre as curvas de crescimento das duas cultivares.

MATERIAL E MÉTODOS

O experimento foi conduzido em casa de vegetação utilizando vasos plásticos de 20kg, com solo Aluvial Eutrófico textura média da camada superficial (0,0 a 0,2m), peneirado em malhas de 4mm. Na preparação dos vasos, efetuou-se a adubação de plantio correspondente à dosagem de 250kg ha⁻¹ da fórmula 04-14-08 (N, P₂O₅ e K₂O) e 30 dias após a emergência das plantas foi feita uma adubação de cobertura com nitrogênio na forma de uréia, equivalente a 30kg ha⁻¹. O delineamento experimental utilizado foi o inteiramente casualizado com vinte repetições no esquema de parcelas subdivididas. Nas parcelas foram colocadas as cultivares de feijão (“Neguinho” e “Carioca”) e nas subparcelas foram coletados os dados de altura das plantas de cinco em cinco dias até aos 85 dias após o plantio, correspondendo assim a 17 observações longitudinais.

Para descrever o crescimento, utilizou-se o modelo não-linear de crescimento Logístico:

$y_{ij} = A_i [1 + b_i \exp(k_i t_{ij})]^{-1} + e_{ij}$, em que: y_{ij} é a altura da planta avaliada na repetição i na idade j , A_i é a altura máxima da planta, b_i é o parâmetro de integração (não se sabe a interpretação biológica) e k_i é a taxa de maturidade (quanto maior, maior a precocidade), e e_{ij} é o termo de erro aleatório, $e_{ij} \sim N(0, \sigma_e^2)$.

A metodologia Bayesiana foi empregada mediante as especificações seguintes.

A distribuição dos dados amostrais foi dada por:

$y_{ij} / \theta, \sigma_e^2 \sim N(A_i [1 + b_i \exp(-k_i t_{ij})]^{-1}, \sigma_e^2)$, $\theta = [A_i, b_i, k_i]$

Portanto, a função de verossimilhança é:

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

As distribuições *a priori* utilizadas foram: $\theta \sim N(\mu, \sigma_e^2)$, $\mu = [\mu_A, \mu_b, \mu_k]$, $\mu \sim \text{Uniforme}(\text{Linf}, \text{Lsup})$,

$\frac{1}{\sigma_e^2} \sim GI(\alpha, \beta)$, $\frac{1}{\sigma_{\theta_i}^2} \sim GI(\alpha_1, \beta_1)$, em que: *Linf*,

Lsup, α , β , α_1 , β_1 são os parâmetros das distribuições *a priori*, também chamados de hiperparâmetros.

Toda a análise foi conduzida no software Winbugs (SPIEGELHALTER et al., 2007), utilizando-se 25.000 iterações na implementação dos algoritmos MCMC para obter amostras convergentes a serem usadas na inferência dos parâmetros das distribuições marginais *a posteriori*. A rotina desenvolvida no presente trabalho é apresentada e comentada na tabela 1.

As amostras das distribuições marginais *a posteriori* para as diferenças dos parâmetros foram obtidas por meio da utilização das amostras geradas para as distribuições marginais dos parâmetros da curva representativa de cada população, ou seja, considerou-se a diferença como uma função, dada por: $f(x) = (\mu_{\text{cultivar}_1} - \mu_{\text{cultivar}_2})$

Assim, foi possível testar a hipótese de igualdade dos parâmetros mediante avaliação do intervalo de credibilidade para as diferenças, pois, se este vier a conter o valor zero, conclui-se que, em média, os parâmetros das duas populações envolvidas no contraste são iguais. Esta metodologia foi apresentada

por SILVA et al. (2005) para comparar parâmetros de curvas de lactação de cabras referentes a duas diferentes populações.

RESULTADOS E DISCUSSÃO

Na tabela 2 estão apresentadas as estimativas dos parâmetros do modelo de crescimento logístico ajustado aos dados das cultivares de feijoeiro “Neguinho” e “Carioca”. As curvas de crescimento representativas de cada cultivar estão apresentadas na figura 1.

As figuras 2, 3, 4 e 5 mostram as distribuições marginais dos parâmetros do modelo de crescimento Logístico, da variância residual e das diferenças entre os parâmetros A e K. De modo geral, essas figuras retratam de forma prática a utilização da metodologia Bayesiana, a qual está fundamentada em distribuições de probabilidade para estimar os parâmetros de interesse.

De acordo com os resultados apresentados na tabela 2, é possível verificar que, em média, o parâmetro A, ou seja, a altura máxima atingida pela planta, não apresentou diferença significativa entre as duas cultivares comparadas. O mesmo não foi

Tabela 1 - Código do software Winbugs para análise de curva de crescimento das cultivares de feijoeiro.

```

model {
  for (i in 1:K) { # K é o número de parcelas experimentais de cada cultivar
    for (j in 1:n) { # n é o número observações de altura no decorrer do tempo
      # cultivar 1: "neguinho"
      Y[i, j] ~ dnorm(eta[i, j], tauC) # Função de verossimilhança
      eta[i, j] <- phi[i, 1] / (1 + phi[i, 2] * exp(phi[i, 3] * x[j])) # Modelo Logístico
      # cultivar 2: "carioca"
      Y1[i, j] ~ dnorm(eta1[i, j], tauC1) # Função de verossimilhança
      eta1[i, j] <- phi1[i, 1] / (1 + phi1[i, 2] * exp(phi1[i, 3] * x[j])) # Modelo Logístico
    }
  }
  # prioris dos parâmetros do modelo Logístico
  phi[i, k] ~ dnorm(mu[k], tau[k]) # cultivar 1: "neguinho"
  phi1[i, k] ~ dnorm(mu1[k], tau1[k]) # cultivar 2: "carioca"
  # prioris para variância do erro
  tauC ~ dgamma(1.0, 1.0E-3) VarC <- 1 / tauC # cultivar 1: "neguinho"
  tauC1 ~ dgamma(1.0, 1.0E-3) VarC1 <- 1 / tauC1 # cultivar 2: "carioca"
  # prioris para a média e a variância dos parâmetros do modelo Logístico
  mu[1] ~ dunif(130, 190) mu[2] ~ dunif(10,400) mu[3] ~ dunif(0.05, 0.52)
  mu1[1] ~ dunif(130, 190) mu1[2] ~ dunif(10,400) mu1[3] ~ dunif(0.05, 0.52)
  tau[1] ~ dgamma(1, 1E-3) tau[2] ~ dgamma(1, 1E-3) tau[3] ~ dgamma(1, 1E-3)
  tau1[1] ~ dgamma(1, 1E-3) tau1[2] ~ dgamma(1.0, 1.0E-3) tau1[3] ~ dgamma(1, 1E-3)
  # distribuições a posteriori para as diferenças dos parâmetros
  dif[1] <- mu[1] - mu1[1]
  dif[2] <- mu[3] - mu1[3]
}

```

Tabela 2 – Estimativas dos parâmetros do modelo de crescimento logístico ajustado aos dados das cultivares de feijoeiro “Neguinho” e “Carioca”.

Cultivar	Parâmetros	Média	Desvio padrão	Erro Monte Carlo	Intervalo de credibilidade	
					2,5%	97,5%
“Neguinho”	A	163,2000	5,7620	0,0437	151,6000	174,7000
	b	42,9000	0,6322	0,0560	41,9400	43,8000
	K	0,0606	0,0113	0,0001	0,0502	0,0900
	σ^2	57,7300	4,5270	0,0582	46,6100	64,3600
“Carioca”	A	160,2000	17,4800	0,1866	131,4000	188,6000
	b	205,4000	113,2000	1,1460	20,2000	390,6000
	K	0,2871	0,1355	0,0012	0,0614	0,5083
	σ^2	0,0082	0,0694	0,0005	0,0003	0,0429
	diferença de A	3,2870	18,4100	0,0137	-28,6400	35,2900
diferença de k	-0,2255	0,1353	0,0009	-0,4494	-0,0013	

verificado para o parâmetro K, taxa de maturidade, pois se observou que o intervalo de credibilidade contém apenas valores negativos, o que permite afirmar que a cultivar “Carioca” apresentou maior precocidade que a cultivar “Neguinho”.

Informações a respeito de cultivares de feijoeiros mais precoces em relação ao crescimento são importantes para avaliar sistemas de plantio e recomendar adubações referentes a cada sistema (VIEIRA et al., 2005). Estes autores compararam a evolução do crescimento das cultivares Talismã e Ouro Negro sob os sistemas de plantio direto e convencional e concluíram que esta última apresentou maior eficiência de crescimento ao se utilizar plantio direto.

Além de a comparação entre os parâmetros das curvas, pode-se também inferir em relação à

precisão do ajuste do modelo não-linear aos dados de cada cultivar estudada. Dessa forma, os dados provenientes da cultivar “Carioca” apresentaram menor variabilidade em relação à forma da curva de crescimento, uma vez que a variância (σ^2) é menor que aquela apresentada pela outra cultivar.

CONCLUSÃO

A metodologia Bayesiana permitiu comparar curvas de crescimento de cultivares de feijoeiro de forma eficiente, sem utilizar aproximações assintóticas. Neste trabalho foi demonstrado que a cultivar “Carioca” apresenta um incremento em altura maior, ou seja, é mais precoce que a cultivar “Neguinho”. Assim esses dados podem ser explorados em experimentos futuros.

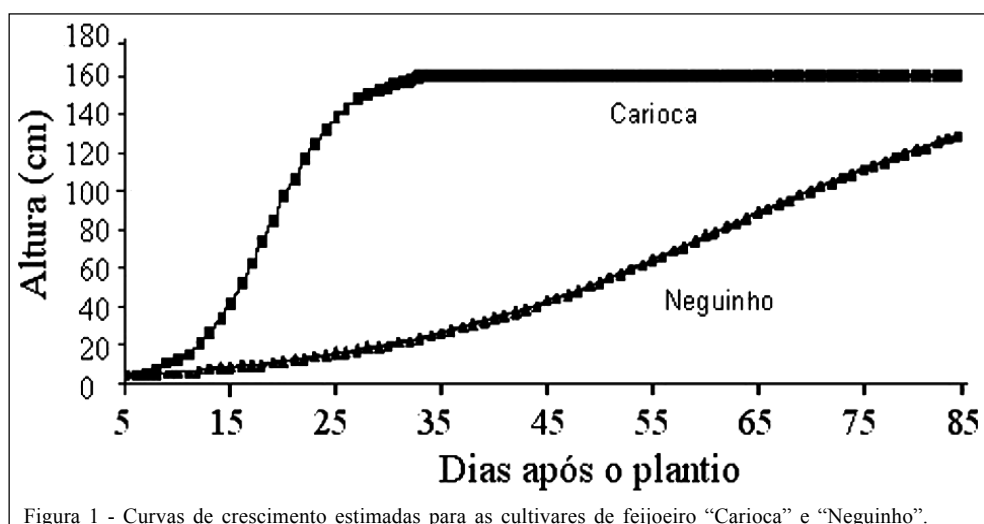


Figura 1 - Curvas de crescimento estimadas para as cultivares de feijoeiro “Carioca” e “Neguinho”.

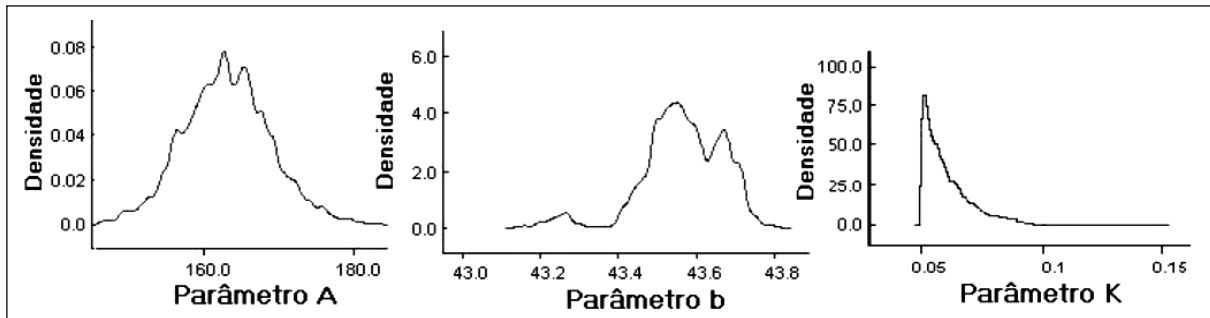


Figura 2 - Amostras das distribuições marginais dos parâmetros do modelo de crescimento logístico ajustado aos dados da cultivar de feijoeiro “Neguinho”.

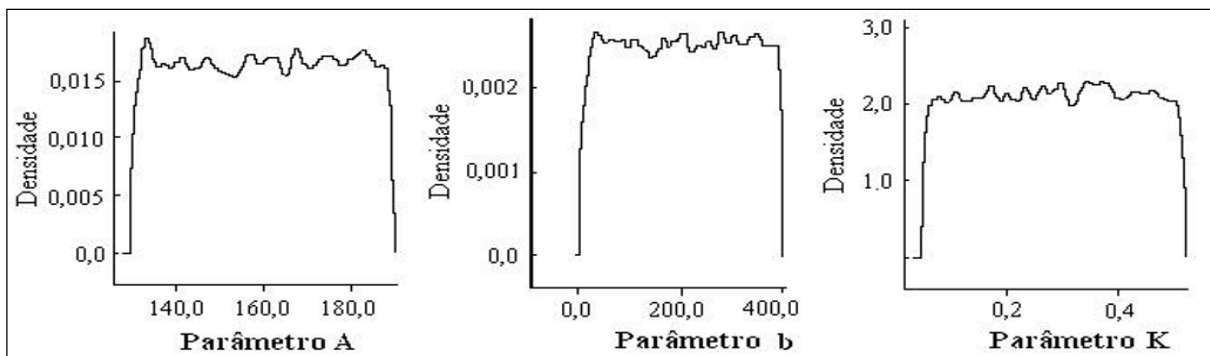


Figura 3 - Amostras das distribuições marginais dos parâmetros do modelo de crescimento logístico ajustado aos dados da cultivar de feijoeiro “Carioca”.

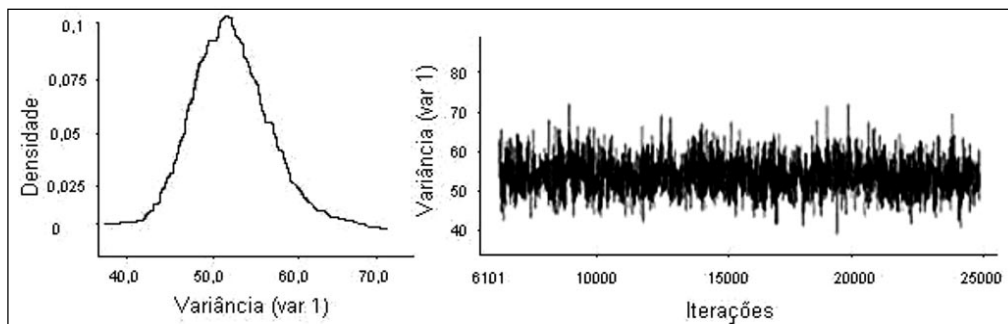


Figura 4 - Amostras da distribuição marginal e cadeia de valores da variância do erro para a cultivar de feijoeiro “Neguinho”.

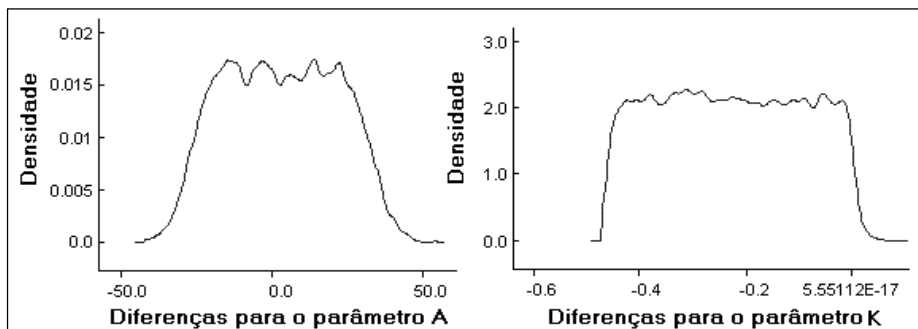


Figura 5 - Amostras das distribuições marginais das diferenças para os parâmetros A e K.

REFERÊNCIAS

- BLASCO, A. et al. Bayesian analysis of the effect of selection for growth rate on growth curves in rabbits. **Genetics Selection Evolution**, Les Ulis, v.35, n.1, p.21-41, 2003.
- ROSA, G.J.M. **Análise bayesiana de modelos lineares mistos robustos via amostrador de gibbs**. 1998. 57f. Tese (Doutorado em Agronomia) – Curso de Pós-graduação em Agronomia, Escola Superior de Agricultura Luiz de Queiroz, USP.
- SILVA, F.F. et al. 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.
- SPIEGELHALTER, D.J. et al. **WinBUGS Version 1.4.1 User Manual**. Cambridge: Medical Research Council Biostatistics Unit. Capturado em 20 mar. 2007. Online. Disponível na Internet <http://www.mrc-bsu.cam.ac.uk/bugs>.
- VIEIRA, N.M.B. et al. **Altura de planta do feijoeiro cvs. talismã e ouro negro em plantio direto e convencional**. In: CONGRESSO NACIONAL DE PESQUISA DE FEIJÃO, 8., 2005, Goiânia, GO. **Anais...** Santo Antônio de Goiás: Embrapa Arroz e Feijão, 2005. p.761-764.