Evaluation of lactation curve of low-yielding gir cows: a Bayesian approach

Joel Augusto Muniz[1*], Fabyano Fonseca e Silva[2], Marlon Henrique de Paiva Guedes[3] e Tarcísio de Moraes Gonçalves[4]

[1]Departamento de Ciências Exatas, Universidade Federal de Lavras, Cx. Postal 3037, Campus Universitário, Lavras, Minas Gerais, Brasil. [2]Departamento de Informática, Universidade Federal de Viçosa, Viçosa, Minas Gerais, Brasil. [3]Bolsista do CNPq e estudante do Curso de Graduação em Medicina Veterinária, Universidade Federal de Lavras, Lavras, Minas Gerais, Brasil. [4]Departamento de Zootecnia, Universidade Federal de Lavras, Minas Gerais, Brasil. *Autor para correspondência. E-mail: joamuniz@ufla.br

ABSTRACT. The Bayesian methodology was used to fit the Wood function for milk production of low-yielding Gir cows with five lactation length: shorter than 228 days (group 1), of 229 to 269 days (group 2), of 270 to 293 days (group 3), of 294 to 304 days (group 4) and longer than 305 days (group 5). The presented procedure provides precise estimates for parameters of lactation curve and permitted a direct comparison between lactations curves by evaluation of HPD interval for marginal posterior density of the difference between parameters. These results didn't show significance, accomplishing that the lactation length doesn't affect shape of curves.

Key words: Bayesian method, lactation curve, Gir cows.

RESUMO. Avaliação de curvas de lactação de vacas gir de baixa produção: um enfoque Bayesiano. O método Bayesiano foi utilizado para ajustasr a função de Wood a dados de produção de leite de vacas Gir de baixa produção divididas em cinco diferentes grupos de duração da lactação: menos que 228 dias (grupo 1), de 229 a 269 dias (grupo 2), de 270 a 293 dias (grupo 3), de 294 a 304 dias (grupo 4) and mais que 305 dias (grupo 5). O método utilizado proporcionou estimativas precisas e possibilitou uma comparação direta entre curvas de lactação provenientes dos diferentes grupos por meio da avaliação dos intervalos HPD´s. Os resultados mostraram que não existem diferenças entre as curvas em relação à duração da lactação.

Palavras-chave: metodologia Bayesiana, curvas de lactação, vacas Gir.

Introduction

At present, in modern milk production, a rigid control of the yield is determining factor for the success of the exploration. A practical and consistent form of reaching this control is through the analysis of lactation curve, which may be defined as the change in level of milk yield in the course of a lactation time (Cobuci *et al*., 1996).

Accurate description of a lactation curve is relevant to activities such as conducting feeding trials with lactating cattle, estimating total lactation yield from incomplete records, and forecasting herd performance on a monthly or individual cow basis (Martin and Sauvant, 2002).

The intensive feeding and the genetic improvement of high yielding cows led to greater performance during the initial days of lactation and to a faster descent from peak, which represent a less persistent lactation. Of this form, the shape of lactation curve can vary with respect to lactation length (Druet *et al*., 2003), but, for low yielding cows, for example Gir cows, the literature research don't present this information.

The Gir is one of the principal Zebu breeds in Brazil and is used for both beef and dairy production. The last present larger importance in the agro-pecuary national scene, due their tolerance to stress conditions and resistance to various tropical diseases (Gonçalves, 1994).

The wide variety of mathematical equations found in the literature can be regarded as alternatives in the sense that each fits some data sets more closely than the other. The form of the fitted curve must be sufficiently flexible to follow closely any trends in the data and to give a consistently good fit to the data. Nevertheless, the incomplete gamma function proposed by Wood (Wood, 1967) is the most widely used formula for describing lactation curve (Varona *et al*., 1998; Chang

et al., 2001; Groenewald and Viljoen, 2003; Silva et al., 2005). This model presents parameters related to level of initial yield, rate of increase to peak yield and rate of decline after peak.

Through functions of those parameters, others important characteristics of lactation curve may be obtained, such as peak milk yield, the permanence time at peak, total milk yield and lactation persistence (Groenewald and Viljöen, 2003).

The Wood equation to describe lactation curve is a nonlinear regression model. The frequentist method to fit this model is Least Square, which use iterative optimization procedures to compute the parameter estimates. The use of iterative procedures requires the user to provide starting values for the unknown parameters before the software can begin the optimization (Souza, 1998). The starting values must be reasonably close to the as yet unknown parameter estimates or the optimization procedure may not converge. In the practice is very difficult to obtain these values. Others disadvantages include a strong sensitivity to outliers, which can generate the biased estimates for parameters, and estimates with large confidence intervals, because the use of approximate distributions of parameters.

In relation to lactation curve comparison, when nonlinear regression is used, comparisons between curves are not possible because the sampling distribution of the parameters is not known, and approximate methods should also be used (Blasco et al., 2003).

Recent studies (Groenewald and Viljoen, 2003; Silva et al., 2005) showed that the Bayesian methodology has been used successfully in the lactation curve analysis. According to these authors, the use of Bayesian methods reduced the number of atypical curves in the presence of small number of observations and showed more precise interval estimation relative to that presented by frequentist methodology. Moreover, permit a direct comparison of curves coming from different treatments, via posterior distribution of the parameters difference.

In the context of Bayesian inference, the parameters of the model are viewed as random variables, according to the subjective concept of probability (O'Hagan, 1994). In this, a prior information on the parameters are utilized in association with the data (likelihood function), generating thus a posterior distribution, Posterior ∝ Likelihood x Prior (Box and Tiao, 1973). This form, the marginal distributions of the parameters, obtained from the posterior distribution integration, provide the Bayesian estimators.

The integration of the posterior distribution generally is not analytical and needs some specialized algorithms as Gibbs Sampler, which needs of the conditional distributions specification (Gelfand et al., 1990). The theory of this algorithm contemplates a method for multidimensional integration, the MCMC – "Markov Chain- Monte Carlo". This uses Monte Carlo simulation and Markov Chains correlations (Sorensen, 1996).

The objective of this paper is to illustrate the Bayesian analysis of lactation curve of low yielding Gir cows. We also study the comparison of the parameters estimates and two characteristics of economic interest (peak yield and persistence of lactation) between five classes of lactation length.

Material and methods

Wood´s model (Wood, 1967) assumes that the expected milk yield (kg day^{-1}) of an animal at time t (weeks) can be expressed over the lactation period as:

$$E[y_t] = t^b \exp(a + ct), \quad a > 0, 0 < b < 1 \text{ e } -1 < c < 0 \quad (1)$$

The parameters a, b and c are unknown and may differ from one animal to another. With the assumption of multiplicative errors in (1), and after a log transformation, the observation model can be written as:

$$Z_{ij} = ln(Y_{ij}) = a_i + b_i ln(t) + c_i t + \varepsilon_{ij} \quad (2)$$

where: $i = 1, 2,..., k$, k is the total number of animals, $j = 1,..., n$, n is the number of test day performances of the animal i , Y_{ij} is the daily milk yield (kg day^{-1}) of the ith animal at time t (weeks), a corresponds to the initial milk yield; b is the ascent rate to peak , c is the descent rate from peak. and ε_{ij} is the experimental random error.

The Wood's model parameters functions are given by: peak milk yield (η) and the persistence of lactation (φ), respectively:

$$\eta = (-b/c)^b \exp(a - b) \quad (3)$$

$$\varphi = -(b+1) ln(-c) \quad (4)$$

The data of milk yield of each cow (Z_i), considering the model presented by expression 2 and using the matrix notation, distribute according to a multivariate normal distribution with mean Xm_i and variance $\sigma^2 I_n$:

$$Z_i \sim N(Xm_i, \sigma^2 I_n) \quad (5)$$

where:

Z_i is the vector of milk yield of each animal i; X is the matrix (n x 3) of design with columns (1 $ln(t_j)$ t_j), m_i is the vector (3 x 1) of parameters a_i, b_i and c_i for each animal i; σ^2 is the residual variance and I_n is the identity matrix of order n.

This form, the likelihood function is given by:

$$f(\mathbf{Z}/a,b,c,\sigma^2) \propto \prod_{i=1}^{k}\prod_{j=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} exp\left\{-\frac{[Z_{ij}-(a_i+b_i ln(t_{ij})+ct_{ij})]^2}{2\sigma^2}\right\} \quad (6)$$

The parameters variation among the animals is assessed by specification of a distribution for m_i, which is a multivariate normal:

$$m_i \sim N(\mu,\Sigma) \quad (7)$$

where:

$\mu = (\alpha\ \beta\ \gamma)$ is the vector of the parameters (a, b and c) population mean and Σ is the covariance matrix (3 x 3).

The report of a *prior* uncertainty over all the parameters (σ^2, μ and Σ) involved in the previous stages, according to Groenewald *et al.* (1996), is give by:

$$f(\mu,\Sigma^{-1},\sigma^2) = f(\mu,\sigma^2)f(\Sigma^{-1}) \quad (8)$$

where:

$f(\mu,\sigma^2) \propto \frac{1}{\sigma^2}$ is the no informative *Jeffreys prior* (Jeffreys, 1961) and $f(\Sigma^{-1})$ corresponds to the inverse of the covariance matrix, which follows a three-dimensional Wishart distribution, $\Sigma^{-1} \sim W((\rho R)^{-1},\rho)$, R^{-1} is the mean of the distribution and ρ the degrees of freedom.

After the definitions of the *prior* distributions and likelihood function, is it possible to specify the conditional *posterior* distributions, which are required for the implementation of the Gibbs sampling algorithm. These conditionals are the following:

$$m_i/\mathbf{Z},\mu,\Sigma^{-1},\sigma^2 \sim N(D^{-1}(X'\mathbf{Z}_i/\sigma^2+\Sigma^{-1}\mu),D^{-1}); \quad (9)$$

$$\mu/\mathbf{Z},M,\Sigma^{-1},\sigma^2 \sim N(\bar{m},\tfrac{1}{k}\Sigma); \quad (10)$$

$$\Sigma^{-1}/\mathbf{Z},M,\mu,\sigma^2 \sim W((S+\rho R)^{-1},k+\rho); \quad (11)$$

$$\sigma^2/\mathbf{Z},M,\mu \sim IG(\tfrac{nk}{2},\tfrac{1}{2}(\sum_{i}^{k}(\mathbf{Z}_i - Xm_i)'(\mathbf{Z}_i - Xm_i))); \quad (12)$$

where:

$\mathbf{Z}=(\mathbf{Z}_1,\mathbf{Z}_2,...,\mathbf{Z}_k)';M=(m_1,m_2,...,m_k)';D=X'X/\sigma^2+\Sigma^{-1};\bar{m}=\tfrac{1}{k}\sum_{i}^{k}m_i;$

$S=\sum_{i}^{k}(m_i-\mu)(m_i-\mu)'.$

The notations N, W and IG in the expressions 9, 10, 11 e 12, represent, respectively, the Multivariate Normal, Wishart and Inverse Gamma distributions.

The IML procedure (Interactive Matrix Language) of the Statistical Analysis System-SAS® software (SAS, 1996) was used to implement the Gibbs Sampler algorithm and its convergence verified through Geweke's criterion (Geweke, 1992) available in the library BOA (Bayesian Output Analysis) of the R (R Development Core Team, 2006) statistical software. The analysis considered a number of 20,000 iterations in the Gibbs sampler, and the bur-in and the sampling interval were, respectively, 10,000 and 20 iterations.

The data used in this study were supplied by the Centro Nacional de Pesquisas em Gado de Leite – Embrapa/CNPGL (National Center of Dairy Cattle Research –Embrapa/NCDCR) and provides test day records, every twenty days, in the lactation period on the milk production to each group studied. These groups, each contained ninety animals, were determined by cows with lactations shorter than 228 days (group 1), of 229 to 269 days (group 2), of 270 to 293 days (group 3), of 294 to 304 days (group 4) and longer than 305 days (group 5).

Results and discussion

The means, standard deviations and Z-score of Geweke's, which represent the convergence criterion, of the posterior distributions of the population parameters a, b e c to each group studied are given in Table 1.

Table 1. Means, standard deviations (SD) and Z-score of Geweke's for the parameters lactation curves and residual variance of the five groups of lactation length.

Par.	Mean	SD	Z-score	Par.	Mean	SD	Z-score
a_1^*	1.1431	0.5054	0.4268	c_1	-0.0050	0.0020	-0.7500
a_2	1.1480	0.5097	0.9605	c_2	-0.0050	0.0021	0.1968
a_3	1.1502	0.5063	1.6259	c_3	-0.0050	0.0020	-0.7523
a_4	1.1569	0.5159	-0.8197	c_4	-0.0049	0.0021	0.5209
a_5	1.1518	0.5057	-1.0837	c_5	-0.0050	0.0020	0.2597
b_1	0.2787	0.1451	0.2732	σ^2_1	0.3028	0.0491	0.4851
b_2	0.2772	0.1487	-0.4198	σ^2_2	0.3032	0.0488	0.9455
b_3	0.2762	0.1454	-0.6515	σ^2_3	0.3014	0.0484	0.7819
b_4	0.2736	0.1483	-0.1019	σ^2_4	0.3018	0.0497	-1.6850
b_5	0.2752	0.1462	1.1163	σ^2_5	0.3019	0.0480	1.6816

*Index 1, 2, ..., 5 correspond to each group of lactation length; Par.: Parameter.

Results of convergence test presented in Table 1 were good, therefore none parameter show the z-score higher than 1.96 (Blasco *et al.*, 2003). These values indicate that the posterior means represented samples of the stationary distribution, as assumed by Baye´s theorem.

The observed posterior means (Table 1) of each parameter obtained by the Bayesian approach agree with previously reported estimates (Gonçalves, 1994), which used the frequentist methodology in the same data set. However, the parameters standard deviations and residual variance were reduced. The Bayesian procedure confirms its power in generate more precise estimate than frequentist methodology.

Tables 2 and 3 present, respectively, the differences for parameters estimates and production functions (peak and persistence) between the groups of lactation length.

Table 2. Means and high posterior density (HPD) intervals for the differences of parameters a, b and c between the five groups of lactation length.

Par.	Mean	HPD LI**	LS	Par.	Mean	HPD LI	LS
a_1-a_2*	-0.0049	-1.4424	1.3982	b_2-b_4	0.0035	-0.4133	0.4076
a_1-a_3	-0.0070	-1.4144	1.3905	b_2-b_5	0.0019	-0.4074	0.4102
a_1-a_4	-0.0013	-1.4431	1.4405	b_3-b_4	0.0026	-0.3885	0.4320
a_1-a_5	-0.0086	-1.4364	1.4277	b_3-b_5	0.0095	-0.4022	0.4085
a_2-a_3	-0.0002	-1.3920	1.4380	b_4-b_5	-0.0016	-0.4274	0.3913
a_2-a_4	-0.0088	-1.5467	1.3382	c_1-c_2	-0.00001	-0.0059	0.0054
a_2-a_5	-0.0037	-1.4172	1.4109	c_1-c_3	-0.00003	-0.0057	0.0056
a_3-a_4	-0.0066	-1.4326	1.4303	c_1-c_4	-0.00005	-0.0058	0.0056
a_3-a_5	-0.0001	-1.3822	1.4049	c_1-c_5	-0.00004	-0.0057	0.0056
a_4-a_5	0.0051	-1.3492	1.5285	c_2-c_3	-0.00001	-0.0058	0.0055
b_1-b_2	0.0015	-0.4015	0.4042	c_2-c_4	-0.00003	-0.0057	0.0059
b_1-b_3	0.0024	-0.4187	0.3926	c_2-c_5	-0.00003	-0.0056	0.0059
b_1-b_4	0.0050	-0.4032	0.4113	c_3-c_4	-0.00002	-0.0059	0.0054
b_1-b_5	0.0034	-0.4169	0.3969	c_3-c_5	-0.00001	-0.0059	0.0055
b_2-b_3	0.0096	-0.4041	0.4064	c_4-c_5	0.00006	-0.0060	0.0055

*Index 1, 2, ..., 5 correspond to each group of lactation length; **LL: Lower limit and UL: Upper limit; Par.: Parameter.

Table 3. Means and hight posterior density (HPD) intervals for the differences of persistence (φ) and peak (η) between the five groups.

Par.	Mean	HPD LI**	LS	Par.	Mean	HPD LI	LS
φ_1-φ_2*	0.0025	-1.5270	1.5549	η_1-η_2	-0.0062	-4.7105	10.7781
φ_1-φ_3	0.0036	-1.5831	1.5395	η_1-η_3	0.0096	-4.6606	4.3572
φ_1-φ_4	0.0033	-1.5485	1.5806	η_1-η_4	0.0142	-4.4970	4.7064
φ_1-φ_5	-0.0285	-1.5708	1.5226	η_1-η_5	-0.2472	-4.8699	4.5971
φ_2-φ_3	0.0011	-1.5614	1.5403	η_2-η_3	0.0158	4.6639	10.6869
φ_2-φ_4	0.0007	-1.5729	1.5345	η_2-η_4	0.0204	-4.6274	4.5033
φ_2-φ_5	-0.0310	-1.5719	1.5322	η_2-η_5	-0.2410	-4.7697	4.5767
φ_3-φ_4	-0.0003	-1.5556	1.6117	η_3-η_4	0.0046	-4.7449	10.8712
φ_3-φ_5	-0.0321	-1.5812	1.5342	η_3-η_5	-0.2568	-4.7106	4.6575
φ_4-φ_5	-0.0318	-1.6049	1.5377	η_4-η_5	-0.2614	-4.7111	10.9465

*Index 1, 2, ..., 5 correspond to each group of lactation length; **LL: Lower limit and UL: Upper limit; Par.: Parameter

The results in the Table 2 and 3 showed no significant difference in the lactation curve between five groups compared. This affirmation can be verified because the high posterior density (HPD) intervals includes zero.

Through these statistical results, it is possible to

conclude that low yielding Gir cows lactation curves are not influenced by the lactation length, disagreeing with the results obtained by Ramos (1984) and Gonçalves (1994), which used the frequentist methodology. Nevertheless, agree with Chowdhary and Barhat (1980), that studies lactations of the Nagauri and Malvi dairy cows, which present similar characteristics to the Brazilian Gir dairy cattle.

A possible cause of the disagreeing can be explained by the statistical methodology, because the Bayesian theory considers the exact distribution of the difference between the parameters values, while the frequentist considers approximate distributions through the asymptotic theory. Another possible reason is that these cited papers didn't use the Wood's function to describe lactation curves, so the parameters didn't present directly the same biological interpretation.

Figure 1 present the estimated lactation curves for the groups 1 (228 days) and 5 (305 days).
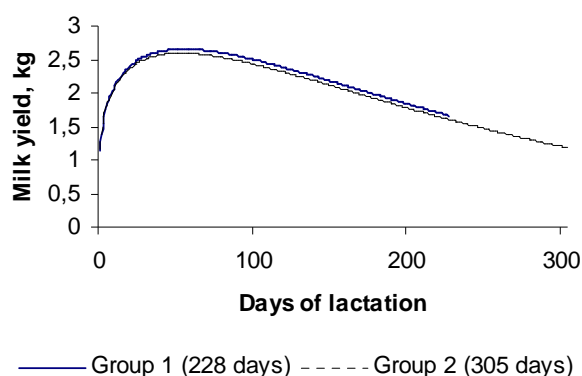
Figure 1. Lactation curves for Gir cows with lactation length of 228 and 305 days.

The estimated curves presented in Figure 1 ratify the no significance of the comparisons presented in Tables 2 and 3, because the plotted curves present practically the same nonlinear behavior.

Figure 2 shown the iterative process of the Gibbs Sampler algorithm and the posterior distribution for the difference between initial milk yield (parameter a) of the groups 1 (228 days) and 5 (305 days).

The iterative process (Figure 2a) illustrate the Gibbs Sampler convergence for the difference between parameter *a*, since discrepant peaks and trends in the chain no were found. The posterior distribution Figure 2b showed no significant difference in terms of probability density, since 95% HPD interval include zero.

(a)

(b)
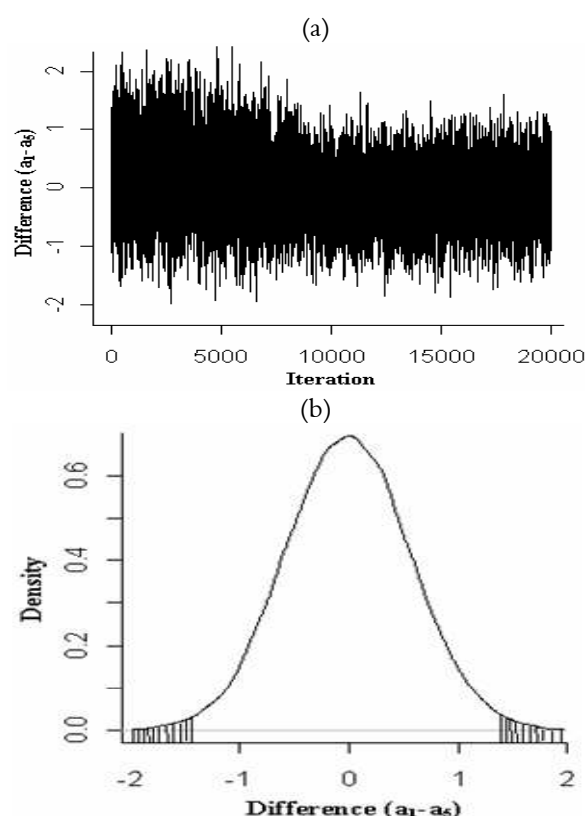
Figure 2. Gibbs Sampler iterations and posterior distribution for the difference between initial milk yield (parameter a): a_1 (228 days) and a_5 (305 days).

Conclusion

The presented Bayesian procedure provides precise estimates for parameters of the low-yielding Gir cow's lactation curve, which is presenting as a consistent method for analysis of nonlinear regression models.

The used methodology permitted a direct comparison between five lactation curves by evaluation of HPD interval for marginal posterior density of the difference between parameters. These results didn't show significance, accomplishing that the lactation length doesn't affect shape of curves.

References

BLASCO, A. *et al.* A Bayesian analysis of the effect of selection for growth rate on growth curves in rabbits. *Genet. Sel. Evol.*, Les Ulis, v. 35, p. 21-41, 2003.

BOX, G.E.P.; TIAO, G.C. *Bayesian inference in statistical analysis.* New York: John Wiley and Sons, 1973.

CHANG, Y.M. *et al.* Genetic variation of lactation curves in dairy sheep: a Bayesian analysis of Wood's function. *Liv. Prod. Sci.*, Paris, v. 71, n. 2, p. 241-251, 2001.

CHOWDHARY, M.S.; BARHAT, N.K. Age at first calving and its influence on important economic traits of Nagauri and Malvi cattle. *Indian Vet. J.*, Madras, v. 57, n. 10, p. 811-815, 1980.

COBUCI, J.A. *et al.* Aspectos genéticos e ambientais da curva de lactação de vacas da raça Guzerá. *Rev. Soc. Bras. Zootec.*, Viçosa, v. 30, n. 4, p. 1204-1211, 1996.

DRUET, T. *et al.* Modeling lactation curves and estimation of genetic parameters for first lactation test-day records of french holstein cows. *J. Dairy Sci.*, Savoy, v. 86, p. 2480-2490, 2003.

GELFAND, A.E. *et al.* Illustration of Bayesian inference in normal data models using Gibbs sampling. *J. Am. Stat. Assoc.*, Alexandria, v. 85, n. 410, p. 972-985, 1990.

GEWEKE, J. Evaluating the accurary of sampling-based approaches to the calculation of posterior moments. *In:* Bernardo, B. *et al.* (Ed.). *Bayesian Statistics 4.* Oxford: University Press, 1992. p. 169-194.

GONÇALVES, T.M. *Estudo da curva de lactação de vacas da raça Gir.* 1994. Dissertação (Mestrado em Zootecnia)– Universidade Federal de Viçosa, Viçosa, 1994.

GROENEWALD, P.C.N., VILJOEN, C.S. A Bayesian model for analysis of lactation curves of dairy goats. *J. Agr., Biol. Environ. Stat.*, Alexandria, v. 8, n. 1, p. 75-83, 2003.

GROENEWALD, P.C.N. *et al.* Application of Bayesian inference in the comparison of lactation curves of Merino ewes. *Anim. Sci.*, London, v. 62, p. 63-69, 1996.

JEFFREYS, H. *Theory of probability.* Oxford: Claredon Press, 1961.

MARTIN, O., SAUVANT, D. Metaanalysis of input/output kinetics in lactating dairy cows. *J. Dairy Sci.*, Savoy, v. 85, p. 3363-3381, 2002.

O'HAGAN, A. *Kendall's advanced theory of statistics:* Bayesian Inference. New York: Edward Arnold, 1994.

R DEVELOPMENT CORE TEAM *A language and environment for statistical computing.* Vienna: R Foundation for Statistical Computing, 2006.

RAMOS, A.A. *Estudo das características reprodutivas de zebuínos leiteiros da raça Gir.* 1994. Dissertação (Mestrado em Zootecnia)– Universidade Estadual Paulista, Botucatu, 1994.

SAS: *Users Guide:* statistics. 5. ed. Cary, 1996.

SILVA, F.F. *et al.* Análise Bayesiana da curva de lactação de cabras Saanen. *Pesq. Agropecu. Bras.,* Brasília, v. 40, n. 1, p. 27-33, 2005.

SORENSEN, D. *Gibbs sampling in quantitative genetic.* Copenhagen: Foulun, 1996.

SOUZA, G.S. *Introdução aos modelos de regressão linear e não linear.* Brasília: Embrapa-SEA, 1998.

VARONA, L. *et al.* Bayesian analysis of wood's lactation curve for Spanish dairy cows. *J. Dairy Sci.*, Savoy, v. 60, n. 5, p. 1469-1478, 1998.

WOOD, P.D.P. Algebraic model of the lactation curve in cattle. *Nature*, London, v. 216, p.164-165, 1967.