



RESEARCH ARTICLE

Estimation of variance components for some production traits of Iranian Holstein dairy cattle using Bayesian and AI-REML methods

Sadegh Alijani*, Mehdi Jasouri, Nasrolah Pirany and Hossein Daghigh Kia

Department of Animal Sciences, Faculty of Agriculture, University of Tabriz, Tabriz, 5166614766, Iran

*Corresponding author: sad-ali@tabrizu.ac.ir

ARTICLE HISTORY

Received: November 30, 2011
Revised: February 18, 2012
Accepted: April 25, 2012

Key words:

AI-REML
BAGS
Iranian Holstein dairy cattle
Variance components

ABSTRACT

The used data set included the records of 131990 Iranian Holstein dairy cattle for first three lactations that were collected from 1981 to 2008 time period by Animal Breeding Center, Iran. The traits which were considered for 305 days of lactation included milk, fat and protein yield and percentages of milk fat and protein. Variance components were estimated using average information restricted maximum likelihood (AI-REML) algorithm using AIREMLF90 software under single trait and repeatability models and Bayesian method by using a Gibbs sampling technique (BAGS) and by MTGSAM and GIBBS3F90 software by same models. The linear statistical models of the analyses included herd-year-season and lactations as fixed effects, age at calving as covariate and animal and permanent environment as random effects. The ranges of heritability estimates for lactations 1 to 3 by animal single and repeatability models using AI-REML and BAGS methods were 0.19 to 0.29, 0.17 to 0.26, 0.20 to 0.25, 0.21 to 0.25 and 0.19 to 0.35 for milk, fat and protein yield and percentage of milk fat and protein, respectively. Repeatability estimates by using BAGS method were 0.44, 0.35, 0.43 and by AI-REML method the values were 0.43, 0.34, 0.39 for milk, fat and protein yield, respectively. The results showed that estimated genetic parameter values by AI-REML analyses for all traits and lactations in both models were smaller than BAGS method. In addition, estimated heritability values for later lactations were lower in comparison with the first lactation.

©2012 PVJ. All rights reserved

To Cite This Article: Alijani S, M Jasouri, N Pirany and HD Kia, 2012. Estimation of variance components for some production traits of Iranian Holstein dairy cattle using Bayesian and AI-REML methods. *Pak Vet J*, 32(4): 562-566.

INTRODUCTION

Genetic parameters that are essentials for most animal breeding programs are based on (co) variance components. Therefore, accurate estimation of these components will result in satisfaction of animal breeding programs. Based on this fact, estimation methods of (co) variance components have been greatly improved over the last four decades and development of these methods has been a constant objective of many animal breeding researchers.

In statistical science, two major approaches of conventional (or frequent) and Bayesian are used. In the mid-1980s, the maximum likelihood method was applied for estimation of (co) variance components and now the new and optimized version of Restricted Maximum Likelihood by average information (AI-REML) method is being used. This method assumes that records of traits are normally distributed (Misztal, 1994; Thompson and

Mantysaari, 1999). Bayesian approach provides a complete paradigm for both statistical inference and decision making under uncertainty. Bayesian methods solve many of the difficulties faced by conventional statistical methods and extend the applicability of statistical methods. It exploits the use of probabilistic models to formulate scientific problems. To use Bayesian statistics, there is computational difficulty and secondly, Bayesian methods require specifying prior probability distributions. Markov Chain Monte-Carlo (MCMC) methods were applied to overcome the computational difficulty, and interest in Bayesian methods was renewed (Kokate *et al.*, 2011). In the past decades the Bayesian method by using a Gibbs sampling technique (BAGS) has been mostly used for estimation of variance components in animal breeding (Wang *et al.*, 1994; Sorensen *et al.*, 1994; Van Tassel and Van Vleck, 1996; Magnabosco *et al.*, 1998 and 2000). In BAGS method, sampled variance

components are taken from the inverted chi - square or its multiple dimensions Wishart distribution (Misztal, 2008).

In Gibbs sampling algorithm three parameters should be determined prior to its application. The first one is the amount of total cycles of sampling to reach reliable estimation and the second one is the number of sampling that must be discarded before the produced sampled values to include in the true posterior distribution, because these sampled values are not still correct. The third one is the interval between sampled values to obtain independent samples from the posterior distribution. Compared to REML, Bayesian methods have the advantage of allowing the inclusion of prior knowledge about unknown parameters in the analysis. Additionally, the Bayes theorem provides a solution for the finite sample size problem, since an exact a posteriori distribution exists for each large or small data set from which inferences can be drawn (Misztal, 2008; Borquis *et al.*, 2010; Leuenberger and Wegmann, 2010). When a large data set is analyzed, a priori information tends to be overwhelmed by the likelihood function in the establishment of the posterior distribution. In this case, parameter estimates are close to those obtained by methods based on likelihood functions. However, this may not be true when the sample size is limited because the maximum likelihood procedure only possesses well-defined properties when the sample size is large enough (Kuhner, 2009; Guillemaud, 2010; Fong *et al.*, 2010). In traditional dairy cattle breeding, first lactation records or, by assuming correlation of unity among yield in different lactations, repeated records of more lactations were used in genetic evaluation. In the first case it is assumed that first lactation records can provide adequate information about later lactation records by considering that same genes influence on first and later lactations (Albuquerque *et al.*, 1996). By improving in computing power and statistical modeling all lactation records are used in genetic evaluation and this has resulted in increasing accuracy of evaluation of dairy animals.

The main aim of this research was estimation of variance components of milk production traits in the first three lactations of Iranian Holstein dairy cattle using single trait and repeatability animal model by AI-REML and BAGS methods and determining Gibbs sampling parameters for applying BAGS method.

MATERIALS AND METHODS

Following editing data, the data set used in this research consisted of 254833 field records of first three lactations of Holstein cows that were collected from 1981 to 2008 in Animal Breeding Center of Iran. Total number of animals in the pedigree was 608478. Milk (MY), fat (FY) and protein yield (PY) and percentage of milk fat (FP) and protein (PP) was considered as traits analyzed for 305 days of lactation. All of the traits were adjusted to 305 days and two times milking. First lactation yield was necessary for all cows and additional lactations were used only if the previous one was recorded. The data distributions by traits are shown in Table 1.

In this study, contemporary groups were defined by a combination of herd, year and season of calving (HYS). Therefore, after editing data, each level of HYS contained at least 15 animals.

In the second step of data preparation, ages of animals at calving were edited and for this purpose, age range (days from birthday) of 660 to 1200 days for first lactation, 990 to 2000 days for second lactation and 1350 to 2500 days for third lactations were considered. The data set editing was done by using SAS (2003) and RENUMF90 (Misztal, 1999b) and FOXPRO 9 (2006) software.

Table 1: Summary information of production traits

Traits	N	Max	Min	Mean	CV (%)
MY1 (kg)	131990	11000	3000	7315	18.99
MY2 (kg)	82070	14775	3000	8132	21.75
MY3 (kg)	40773	16166	3000	8492	22.70
FY1 (kg)	106400	351	125	237	20.10
FY2 (kg)	67040	500	125	267	24.15
FY3 (kg)	32177	597	130	282	24.95
FP1 (%)	104121	5.94	2.50	3.32	13.01
FP2 (%)	68041	5	1.50	3.28	15.63
FP3 (%)	32644	5	1.33	3.31	15.58
PY1 (kg)	77489	300	115	226	16.09
PY2 (kg)	52933	466	110	258	19.19
PY3 (kg)	26581	479	110	268	19.84
PP1 (%)	80131	3.85	2.30	3.07	7.64
PP2 (%)	53023	5	1.84	3.11	8.54
PP3 (%)	25920	5	1.66	3.09	8.66

MY, milk yield; FY, fat yield; FP, fat percentage; PY, protein yield; PP, protein percentage. 1, 2, 3 refers to lactations. N: number of recorded animals; Max: Maximum; Min: Minimum; CV: coefficient of variance.

The analyses were conducted using a single trait and repeatability linear animal model. For univariate analyses, three lactation records were analyzed individually using a single trait linear animal model as follows (in matrix notation):

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{u} + \mathbf{e} \quad \text{Where,}$$

\mathbf{y} is the observation vector, in this case is one of the lactation records.

\mathbf{b} is the fixed effect vector of the HYS contemporary groups effect and regression coefficient of covariate of age of the animal at calving coefficient.

\mathbf{u} is the additive random genetic effect vector of the animal.

\mathbf{e} is the residual random effect vector.

\mathbf{X} and \mathbf{Z} are the incidence matrices that relate the data to the fixed and random effects, respectively.

The structure of (co) variance matrix in random variables in this model is as follows:

$$\text{Var} \begin{bmatrix} \mathbf{y} \\ \mathbf{u} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{V} & \mathbf{Z}\mathbf{G} & \mathbf{R} \\ \mathbf{G}\mathbf{Z}' & \mathbf{G} & \mathbf{0} \\ \mathbf{R} & \mathbf{0} & \mathbf{R} \end{bmatrix} \quad \text{Where,}$$

$$\mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R}.$$

After considering $\mathbf{R} = \mathbf{I}\sigma_e^2$ and $\mathbf{G} = \mathbf{A}\sigma_u^2$ that \mathbf{A} is the matrix of additive relationship between individuals and \mathbf{I} is the identically matrix in these equations, the following MME (mixed model equation) was solved for obtaining BLUP prediction of animal random effect;

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{A}^{-1}\mathbf{k} \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \mathbf{b} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix} \quad \text{Where,}$$

$$\mathbf{k} = \frac{\sigma_e^2}{\sigma_u^2} = \frac{1-h^2}{h^2}$$

The milk production traits in three lactations of each animal were considered as repeated measurements;

therefore a repeatability animal model was applied. This model assumes a genetic correlation equal to one among the various lactation measurements. The following statistical model was used for these analyses:

$$y = Xb + Zu + Wpe + e$$

In which y , b , u , e , X and Z are defined as single trait model and be is the random effects vector of the permanent environment and W is the incidence matrix that relates the observations in y with this effect. In this model it was assumed that

$$\text{Var} \begin{bmatrix} u \\ pe \\ e \end{bmatrix} = \begin{bmatrix} A\sigma_u^2 & 0 & 0 \\ 0 & I\sigma_{pe}^2 & 0 \\ 0 & 0 & I\sigma_e^2 \end{bmatrix}$$

For REML method, variance components were estimated using AI-REML algorithm and by applying AIREMLF90 software (Misztal, 1999a) under single trait and repeatability model. Single traits analyzed by BAGS method using MTGSAM (Multiple Trait using Gibbs Sampling under Animal Model) software (Van tassel and Van Vleck, 1996) and for the repeatability animal model, GIBBS3F90 software (Misztal, 1999c) were used. Totally, 36 analyses were done in this research. In order to study different levels of Gibbs Sampling parameters, different strategies were applied, these strategies were used for estimation of (co) variance component for traits of 305 days milk yield (MY) and mature equivalent milk yield (ME). These strategies included different combinations of different values for each of Gibbs Sampling parameters constituting 100, 150 and 200 for thinning interval; 1000, 10000 and 30000 for bur-in period; and 100000, 150000 and 200000 for total chain length. Considering all combinations, finally 56 analyses were done.

RESULTS AND DISCUSSION

The result of applying different strategies of BAGS method showed that firstly thinning interval had no effect on the mean, mode or median of the posterior marginal distribution of genetic parameters. Secondly, all values of chain length and bur-in period had very little effect on the amount of variance components estimations in this research. These results are presented in Table 2. Based on these results in other BAGS analyses the values of 100000, 30000 and 200 were used for total chain length, bur-in period and thinning interval, respectively. The results of this BAGS analysis were reported as the results of this estimation method.

The estimated values of variance components, genetic parameters and standard errors of first three lactations of study traits that were calculated by BAGS and AI-REML method under single trait model are shown in Table 3. In Table 4, the repeatability model output of these methods is included. Some of posterior distributions of residual and additive genetic variances of different traits have been shown in Fig. 1.

Results of single trait analyses indicated that residual variance, additive genetic variance and heritability of first lactations were higher in comparison with other lactations. This might be caused by factors that do not influence the different lactations, such as management systems or environmental effects. The range of heritability in this study was obtained at 0.10 to 0.29. In both methods and in

comparison to other lactations and traits the highest heritability value was for protein percentage in first lactation in contrast to the lowest value of heritability for fat yield in third lactation.

Table 2: Variance components estimation for different values of Gibbs sampling parameters

Total cycle	ME			MY		
Burning Periods	1000	10000	30000	1000	10000	30000
σ_u^2	532980	532956	533150	403420	403403	403547
σ_e^2	1230274	1230284	1230143	926261	926267	926163
Total cycle	ME			MY		
150000	532503	532440	532629	403016	403015	403154
σ_u^2	1230621	1230657	1230519	926523	926548	926447
Total cycle	ME			MY		
100000	531551	531357	531371	402344	402201	402207
σ_u^2	1231296	1231418	1231401	927030	927120	927110

MY: milk 305 days yield; ME: mature equivalent milk yield; σ_u^2 , additive variance; σ_e^2 , residual variance.

The heritability values obtained by repeatability model were relatively higher than single traits model, which probably is due to permanent environmental effects; similar to the results obtained by other studies including Palacios *et al.* (2007). These results showed that the genetic correlations among lactations were not equal to one. In this study repeatability of MY trait was higher than FY and PY traits in both AI-REML and BAGS methods.

According to a previous report (Boldman and Freeman, 1990), genetic parameter estimates may change with production levels, phenotypic variance and methods of estimation. However, in this research, estimated genetic parameters are similar to other results from the literature. Meyer (1984) reported heritabilities of 0.34, 0.35 and 0.28 of milk yield; 0.32, 0.33, 0.23 of fat yield and 0.24, 0.29, 0.12 of protein yield in 1st, 2nd and 3rd lactation, respectively. Dedkova and Wolf (2001) reported heritability of 0.30, 0.28 and 0.30 of milk yield; 0.24, 0.25 and 0.25 of fat yield and 0.25, 0.25 and 0.27 of protein yield in 1st, 2nd and 3rd lactation, respectively. AL-Seaf *et al.* (2007) obtained heritability of milk yield in Holstein dairy cows of 0.18, 0.18 and 0.14 for first three lactations, respectively. Dematawewa and Berger (1998) applied a repeatability model for first three lactation records and for milk yield, fat yield and protein yield traits of Holstein dairy cows and reported the value of these trait heritability and repeatability as 0.196, 0.177 and 0.179; 0.424, 0.410 and 0.412, respectively.

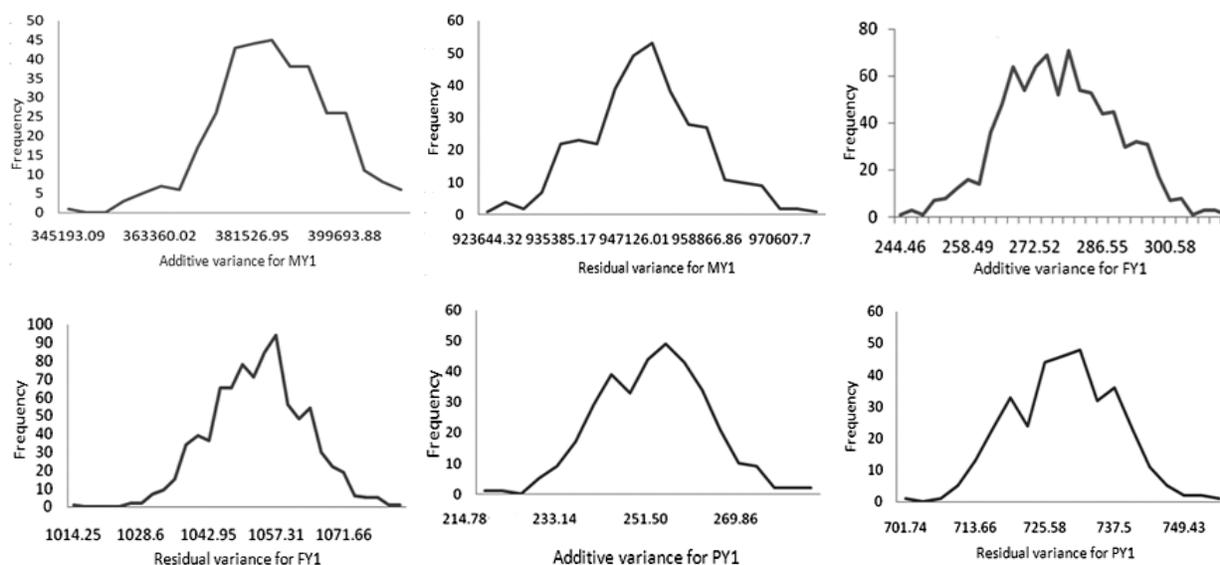
Generally, relative values of the estimated heritabilities in this study coincide with other reports such as Albuquerque *et al.* (1996) and Palacios *et al.* (2007), which concluded that the heritability of milk production traits in the first lactation was higher than of later lactations.

Conclusion: The results presented here show that the estimates of variance component and genetic parameter in BAGS method were higher than those by AI-REML method, for all traits and models. Additionally, estimates

Table 3: Estimated parameters for first three lactations of different traits calculated by BAGS and AI-REML method under single trait model

Traits	BAGS			AI-REML		
	σ_u^2	σ_e^2	h^2	σ_u^2	σ_e^2	h^2
MY1	383562±11252	947443±9114	0.29	348320±11640	977090±8845	0.26
MY2	481926±23787	1778344±20160	0.21	379750±20922	1856300±17829	0.17
MY3	542477±39099	2297765±32954	0.19	357780±31318	2445600±29621	0.13
FY1	276±12	1052±10	0.21	217±11	1097±9	0.17
FY2	445±25	2085±22	0.17	304±22	2186±20	0.13
FY3	556±50	2720±45	0.17	317±35	2912±36	0.10
FP1	0.0328±0.0044	0.104±0.0010	0.25	0.0267±0.0013	0.109±0.001	0.19
FP2	0.0400±0.0024	0.141±0.0019	0.23	0.0296±0.0019	0.151±0.002	0.16
FP3	0.0415±0.0033	0.154±0.0029	0.21	0.0269±0.0026	0.165±0.002	0.14
PY1	247±11	727±9	0.25	204±10	763±8	0.21
PY2	419±24	1418±20	0.23	334±21	1485±18	0.18
PY3	470±39	1854±34	0.20	335±32	1961±31	0.15
PP1	0.0106±0.0004	0.0210±0.0003	0.35	0.0090±0.0004	0.0220±0.0003	0.29
PP2	0.0121±0.0007	0.0376±0.0006	0.24	0.0085±0.0006	0.0405±0.0005	0.17
PP3	0.0102±0.0009	0.0422±0.0008	0.19	0.0071±0.0007	0.0444±0.0007	0.14

MY, milk yield; FY, fat yield; FP, fat percentage; PY, protein yield; PP, protein percentage; σ_u^2 = additive variance; σ_e^2 = residual variance; h^2 = heritability (1, 2, 3) refers to milk lactations.

**Fig. 1:** Posterior distributions of additive and residual variance for MY1, FY1 and PY1.**Table 4:** Estimated values of variance components, genetic parameters and standard errors for first three lactations of different traits calculated by BAGS and AI-REML methods under repeatability model

	BAGS			AI-REML		
	MY	FY	PY	MY	FY	PY
σ_u^2	609986±11004	598.67±6	444.73±9	550680±12941	581.75±15	386.19±12
σ_{pe}^2	320506±5146	207.44±12	305.65±9	305320±9467	199.53±11	234.55±9
σ_e^2	1205450±3546	1526±6	1016.80±5	1144400±4814	1494.1±7	974.2±5
h^2	0.29	0.26	0.25	0.27	0.26	0.24
r	0.44	0.35	0.43	0.43	0.34	0.39

MY, milk yield; FY, fat yield; PY, protein yield; σ_u^2 = additive variance; σ_{pe}^2 = permanent environment; σ_e^2 = residual variance; h^2 = heritability; r = repeatability.

of variance components using BAGS method in comparison of AI-REML method had lower standard errors in both models. However, computer processing time was greater in BAGS than in AI-REML.

Acknowledgment: Authors are grateful to the Information Technology Center of University of Tabriz and Animal Breeding Center of IRAN that allowed us to

conduct this study. The authors also wish to thank Igancy Misztal, Curt Van Tassell and Dale Van Vleck for allowing access to their software.

REFERENCES

Albuquerque LG, JF Keown and LD Van Vleck, 1996. Genetic parameters of milk, fat, and protein yield in the three lactations,

- using an animal model and restricted maximum likelihood. *Braz J Genet*, 191: 79-86.
- Al-Seaf A, JF Keown and LD Van Vleck, 2007. Genetic parameters for yield traits of cows treated or not treated with bovine somatotropin. *J Dairy Sci*, 90: 501-506.
- Borquis RRA, R Dipalo, FRA Neto, F Baldi, GMF de Camargo, LG de Albuquerque, L Zicarelli and H Tonhati, 2010. Genetic parameter estimates for buffalo milk yield, milk quality and mozzarella production and Bayesian inference analysis of their relationships. *Genet Mol Res*, 9: 1636-1644.
- Boldman KG and AE Freeman, 1990. Adjustment for heterogeneity of variance by herd production level in dairy cow and sire evaluation. *J Dairy Sci*, 73: 503-512.
- Dedkova L and J Wolf, 2001. Estimation of genetic parameters for milk production traits in Czech dairy cattle population. *Czech J Anim Sci*, 46: 298-307.
- Dematawewa CMB and PJ Berger, 1998. Genetic and phenotypic parameters for 305-day yield, fertility and survival in Holsteins. *J Dairy Sci*, 81: 2700-2709.
- Fong Y, H Rue and J Wakefield, 2010. Bayesian inference for generalized linear mixed models. *Biostatistics*, 11: 397-412.
- Guillemaud T, 2010. Inferring introduction routes of invasive species using approximate Bayesian computation on microsatellite data. *Heredity*, 104: 88-99.
- Jensen J, CS Wang, DA Sorensen and D Gianola, 1994. Bayesian inference on variance and covariance components for traits influenced by maternal and direct genetic effects, using the Gibbs sampler. *Acta Agric Scand Anim Sci*, 44: 193.
- Kokate LS, GR Gowane, MS Dige, GS Sonawane, C Mishra and RK Singh, 2011. Bayesian Statistics: Concepts and Applications in Animal Breeding – A Review. *J Adv Vet Res*, 1: 94-98.
- Kuhner MK, 2009. Coalescent genealogy samplers: windows into population history. *Trends Ecol Evol*, 24: 86-93.
- Leuenberger C and D Wegmann, 2010. Bayesian computation and model selection without likelihoods. *Genetics*, 184: 243-252.
- Magnabosco CDU, RB Lobo, A Reyes, EN Martin and T Famula, 1998. Bayesian inference on growth traits in Nelore cattle in Brasil. *Proc 6th World Congress on Genetics Applied to Livestock Production*, Guelph, Canada, 25: 556-559.
- Magnabosco CDU, RB Lobo and T Famula, 2000. Bayesian influence for genetic parameter estimation on growth traits in Nelore cattle in Brazil, using the Gibbs sampler. *J Anim Breed Genet*, 117: 169-188.
- Meyer K, 1984. Estimation of Genetic parameters for milk and fat yield for the first three lactations in British Friesian cows. *J Anim Prod*, 38: 313-322.
- Misztal ML, 1994. Comparison of software package in animal breeding. *Proc 5th World Congress on Genetics Applied to Livestock Production*, Guelph, Canada, 22: 3-10.
- Misztal ML, 1999a. REMLF90 Manual. Available from URL: <http://nce.ads.uga.edu/igancy/numpub/bilupf90/docs/remlf90.pdf>.
- Misztal ML, 1999b. RENUMF90 Manual. Available from URL: <http://nce.ads.uga.edu/igancy/numpub/bilupf90/docs/reunmf90.pdf>.
- Misztal ML, 1999c. GIBBS3F90 Manual. Available from URL: <http://nce.ads.uga.edu/igancy/numpub/bilupf90/docs/gibbs3f90.pdf>.
- Misztal ML, 2008. Reliable computing in estimation of variance component. *J Anim Breed Genet*, 125: 363-370.
- Palacios A, Lj Espinoza, GD Pena, GD Iglesias, LP Rafael and RF Almeida, 2007. Estimation of covariance components for the first four lactations in Holstein cattle according to different models. *J Zootec Trop*, 25: 9-18.
- SAS, 2003. SAS/STAT User's Guide. Version 9.1. SAS Institute Inc., Cary, NC, USA.
- Sorensen DA, CS Wang, J Jensen and D Gianola, 1994. Bayesian analysis of genetic change due to selection using Gibbs sampling. *Genet Sel Evol*, 26: 333-360.
- Thompson R and EA Mantysaari, 1999. Prospects for statistical methods in dairy cattle breeding. *Interbull Bull*, 20: 70-79.
- Van Tassel CP and LD Van Vleck, 1996. Multiple-trait Gibbs sampler for animal model: flexible program for Bayesian and likelihood based (Co) variance component inference. *J Anim Sci*, 74: 2586-2597.
- Wang CS, JJ Rutledge and D Gianola, 1993. Marginal inferences about variance components in a mixed linear model using Gibbs sampling. *Genet Sel Evol*, 25: 41-62.
- Wang CS, JJ Rutledge and D Gianola, 1994. Bayesian analysis of mixed linear model via Gibbs sampling with an application to litter size in Iberian pigs. *Genet Sel Evol*, 26: 91-102.