



S B M A 2008

VII Simpósio Brasileiro de Melhoramento Animal

São Carlos, SP, 10 e 11 de julho de 2008

Comparação de modelos de regressão aleatória, para avaliação genética da produção de leite no dia do controle de vacas Guzerá

Luciana Salles de Freitas¹, Martinho de Almeida e Silva², Rui da Silva Verneque³,
 Glaucyana Gouvêa Santos⁴, Gerusa Salles Corrêa⁴, Bruno Dourado Valente⁴, Maria
 Gabriela Campolina Diniz Peixoto³, Ricardo Vieira Ventura⁴

¹Aluna de doutorado do departamento de Zootecnia – EV-UFMG, e-mail: lucianafreitas1002@hotmail.com

²Professor do departamento de Zootecnia – EV-UFMG – bolsista CNPq

³Pesquisador(a) Embrapa Gado de Leite - bolsista CNPq

⁴Aluno(a) de pós-graduação do departamento de Zootecnia – EV-UFMG

Resumo – Um total de 8276 observações de produção de leite no dia do controle foi utilizado para avaliação genética de 1021 vacas guzerá de primeira lactação por meio de modelos de regressão aleatória. Os modelos utilizados nas análises envolveram duas funções paramétricas (Wilmink, e Ali e Schaeffer) e polinômios de Legendre de terceiro, quarto e quinto grau. Na determinação do modelo que melhor se ajustava às observações, foram utilizados os critérios do logaritmo da função de verossimilhança, o critério de informação Akaike, o critério de informação Bayesiano e o maior requerimento de tempo para as análises dos dados. Os modelos que incluíram maior número de parâmetros foram os que melhor se ajustaram aos dados para avaliação genética de vacas guzerá utilizando ajustamento de modelos de regressão aleatória aos dados de produção de leite no dia do controle de vacas Guzerá de primeira lactação. O polinômio de Legendre de quarto grau, por apresentar menor demanda de tempo e não apresentar problemas de convergência na solução do sistema de equações, foi o mais indicado para avaliação genética da produção de leite no dia de controle de vacas Guzerá utilizando-se modelos de regressão aleatória.

Palavras-chave: Ajustamento de modelo, função paramétrica, polinômio de Legendre, Wilmink.

Comparison of random regression models for genetic evaluation of test day milk production of first lactation Guzerá cows

Abstract – A total of 8,276 records of test day milk production were used for genetic evaluation of 1,021 first lactations of Guzera cows using random regression models. The models used in the random regression analyses were two parametric models (Wilmink, and Ali and Schaeffer) and Legendre polynomials of 3rd, 4th, and 5th degrees. The criteria used to select models were the logarithm of the likelihood function, Akaike information criterion, Bayesian information criterion and the demanding computational time involved in

SP 39 53

P. 137

the analyses using random regression models for the genetic evaluation of test day milk production of first lactations of Guzera cows. The models including higher number of parameters showed the best goodness of fit. Fourth degree Legendre polynomial showed lower demand in computational time and no convergence problems in the solution of the equation system, and was the more appropriate for genetic evaluation of test day milk production of Guzera cows.

Key-words: Adjustment of model, Legendre polynomial, parametric function, Wilmink.

Introdução

Os modelos de regressão aleatória permitem ajustar funções suaves e parcimoniosas para prever os valores genéticos para a curva como um todo, o que proporciona a descrição de estrutura de covariância dos efeitos aleatórios ajustados no modelo para qualquer ponto da curva de produção considerada (Gilmour, 2006), porém a determinação do modelo que melhor se ajusta aos dados não está bem definida e tem sido muito estudada.

O objetivo deste trabalho foi avaliar entre modelos de regressão aleatória, utilizando duas funções paramétricas (Wilmink e Ali e Schaeffer) e os polinômio de Legendre de três graus diferentes, o que melhor se ajustava às avaliações de produção de leite no dia do controle de vacas da raça Guzera de primeira lactação.

Material e Métodos

Neste estudo foram utilizados cinco diferentes modelos de regressão aleatória para avaliação da produção de leite no dia no controle: duas funções paramétricas, Wilmink (1987) e Ali e Schaeffer (1987), e os polinômio de Legendre de terceiro, quarto e quinto graus.

As informações de 1021 primeiras lactações extraídas da base de dados do Programa Nacional de Melhoramento do Guzera para Leite, executado pela Embrapa Gado de Leite, em parceria com o Centro Brasileiro de Melhoramento do Guzera, totalizaram 8276 observações. A separação dos grupos contemporâneos foi feita de acordo com classes de rebanho-ano e épocas do controle, de abril a setembro (seca) e outubro a março (águas). O que permitiu formar 110 classes de efeito fixo classificatório. No modelo de regressão foram incluídos os efeitos fixos, como função dos dias em lactação, dos controles agrupados de acordo com cinco classes de idade da vaca (22 a 36 meses, 37 a 39, 40 a 42, 43 a 47 e 48 a 59 meses), e em duas épocas de parto (abril a setembro e outubro a março). Esses grupos foram combinados, constituindo-se 10 classes de idade-época do parto; pressupondo que cada classe pode apresentar um formato diferente para a curva de lactação. O arquivo de genealogia continha 5005 animais.

O seguinte modelo foi utilizado para obtenção dos coeficientes de regressão aleatória:

$$Y_{gijkl} = RAE_i + \sum_{m=1}^n \beta_{km} Z_{jlm} + \sum_{m=1}^n a_{jm} Z_{jlm} + \sum_{m=1}^n p_{jm} Z_{jlm} + e_{gijkl}; \text{ em que}$$

Y_{gijkl} é o controle l da vaca j no (t_j) período de dias em lactação, dentro da classe i (rebanho-ano-época do controle), pertencente à subclasse k (idade-época de parto); RAE_i efeito fixo de rebanho-ano-época do controle; e_{ijkl} é o efeito aleatório residual associado à Y_{ijkl} ; β_{km} é vetor que contém coeficientes de regressão dos efeitos fixos da produção de leite no dia do controle em função de t para cada subclasse k ; a_{jm} e p_{jm} são vetores contendo coeficientes de regressão aleatórios que descrevem a trajetória do efeito genético aditivo e

permanente de ambiente, respectivamente, durante o período de lactação para cada animal: n é o número de coeficientes da função utilizada e Z_{ilm} é um vetor de covariáveis, que representa os respectivos coeficientes das funções matemáticas utilizadas para descrever a curva de lactação.

Para verificação do modelo de melhor ajuste, foram utilizados a função de verossimilhança ($\ln L$), o Critério de Informação de Akaike (AIC) e o critério de informação Bayesiano de Schwarz (BIC) que são obtidos por:

$$AIC = -2\ln L + 2p \quad \text{e} \quad BIC = -2\ln L + p\ln(N-r); \text{ em que}$$

p é o número de parâmetros do modelo, N é o número total de observações e r é o posto da matriz de incidência dos efeitos fixos. Menores valores para AIC e para BIC indicam modelos mais adequados para ajuste dos dados.

O programa utilizado foi o REMLF90 (Misztal, 2001), que trabalha com a Maximização da Esperança da Função de Máxima Verossimilhança Restrita (EMREML) e com a aceleração da convergência, adotando-se como critério de convergência o valor de 10^{-9} . Assim foram estimadas as matrizes G e P e a σ_e^2 .

Os componentes de variância e covariância foram obtidos por meio de solução inicial do sistema adotando critério de convergência com baixa precisão (diferenças menores que 10^{-9}). Os resultados obtidos foram utilizados nas análises seguintes, como informação inicial de processamentos mais precisos.

Resultados e Discussão

No ajustamento da função de Ali e Schaeffer houve dificuldades para atingir o critério de convergência menor do que 10^{-9} provavelmente pelo número de parâmetros estimados, conforme já observado por Jamrozik et al. (1997), o que não foi observado para os demais modelos. Portanto, para melhor comparação entre ajustamentos a obtenção da solução para todos os sistemas foi realizada utilizando critério de convergência com estimativas de 10^{-9} .

A Tabela 1 apresenta o número de parâmetros estimados em cada função, e a qualidade do ajustamento apresentada em cada análise.

Tabela 1- Número de parâmetros estimados (NP), função de verossimilhança ($\log L$), critério de informação Akaike (AIC) e critério de informação Bayesiano (BIC) para os diferentes modelos de regressão aleatória.

Modelo	NP	$\log L$	AIC	BIC
Wilmink	13	-12607,347	25240,694	25331,7467
Ali e Schaeffer	31	-12021,0316	24104,0633	24321,1126
Legendre 3	19	-12279,2144	24596,4288	24729,4825
Legendre 4	31	-12082,0041	24226,0082	24443,0576
Legendre 5	46	-12042,0783	24176,1566	24498,1731

De acordo com os resultados obtidos, observa-se que, apesar dos critérios de informação AIC e BIC penalizarem modelos com maior número de parâmetros, houve melhor ajuste com o aumento do número de parâmetros. A função paramétrica de Ali e

Schaeffer apresentou o melhor ajuste, seguida do polinômio de Legendre de quinto grau para AIC e de quarto grau para BIC, mas com pequena diferença entre os dois.

A função de Wilmink apresentou maior demora na convergência, necessitando de número excessivo de rodadas para atingir o resultado final. O que gerou maior tempo nas análises computacionais, mesmo com menor número de parâmetros estimados. A função de Wilmink foi a que menos se ajustou aos dados de produção de leite no dia do controle.

Conforme Schaeffer (2004), polinômios ortogonais são os mais apropriados para estimativas de covariâncias em modelos de regressão aleatória, mas há sempre necessidade de se estimar a ordem do polinômio que deve ser utilizada em cada avaliação. Percebe-se neste estudo que os modelos que utilizaram polinômio de Legendre foram os que apresentaram menor demanda computacional e menor tempo de convergência. Os polinômios de Legendre de quarto e quinto grau apresentaram diferença pouco expressiva pelos critérios de informação AIC e BIC, então os autovalores encontrados foram observados. Os autovalores demonstraram melhor ajustamento para o polinômio de quarto grau, que apresentou o menor autovalor com grande diferença em relação ao maior ($0,5641 \times 10^{-5}$ e 1,858), o que não foi observado para o de quinto grau ($0,1126 \times 10^{-1}$ e 1,894), onde os autovalores estiveram muito próximos o que indica que o aumento de um grau no polinômio não estaria acarretando mudanças apreciáveis na análise.

Conclusões

A função paramétrica de Ali e Schaeffer apresentou dificuldades para convergência, mas foi a que melhor se ajustou aos dados de produção de leite no dia do controle de vacas da raça Guzerá.

Entre os modelos utilizados, o polinômio de Legendre de quarto grau foi o que proporcionou o segundo melhor ajuste aos dados de produção de leite no dia do controle e, em razão da pequena diferença de ajustamento em relação à função de Ali e Schaeffer, da diferença entre os autovalores e ao menor tempo de análise, pode ser indicado para avaliações genéticas para produção de leite no dia do controle em primeiras lactações de vacas Guzerá.

Literatura Citada

- ALI, T.E.; SCHAEFFER, L.R. Accounting for covariances among test day milk yields in dairy cows. **Canadian Journal Animal Science**, v.67, n.3, p.637-644, 1987.
- GILMOUR, A.R. Statistical models for multidimensional (longitudinal/spatial) data. In: WORLD CONGRESS ON GENETICS APPLIED TO LIVESTOCK PRODUCTION, 8., 2006, Belo Horizonte. **Proceedings...** Belo Horizonte, 2006. Disponível em: <http://www.wcgalp8.org.br/wcgalp8/articles/paper/25_339-773.pdf>. Acesso em: nov. 2007.
- JAMROZIK, J.; SCHAEFFER, L.R.; DEKKERS, J.C.M. Genetic evaluation of dairy cattle using test day yields and random regression model. **Journal Dairy Science**, v.80, n.6, p.1217-1226, 1997.
- MISZTAL, I. **REMLF90 Manual**, 2001. Disponível em: <<http://nce.ads.uga.edu/~ignacy/newprograms.html/>>. Acesso em: março 2007.
- SCHAEFFER, L.R. Application of random regression models in animal breeding. **Livestock Production Science**, v.86, p.35-45, 2004.
- WILMINK, J.B.M. Efficiency of selection for different cumulative milk, fat and protein yields in first lactation. **Livestock Production Science**, v.17, n.3, p.211-224, 1987.