



Avaliação genética animal na presença de observações discrepantes

José Marques Carneiro Júnior¹, Giselle Mariano Lessa de Assis², Priscila Ferreira Wolter³, Williane Maria de Oliveira Martins⁴, Francisco Aloísio Cavalcante⁵

¹Pesquisador da Embrapa Acre. E-mail: marques@cpafac.embrapa.br

²Pesquisadora da Embrapa Acre. E-mail: giselle@cpafac.embrapa.br

³Graduanda de Ciências Biológicas da União Educacional do Norte. E-mail: priscilawolter18@yahoo.com.br

⁴Graduanda de Engenharia Agrônoma da Universidade Federal do Acre. E-mail: williane_martins@yahoo.com.br

⁵Pesquisador da Embrapa Acre. E-mail: aloisio@cpafac.embrapa.br

Resumo: O objetivo deste trabalho foi avaliar o efeito da presença de valores discrepantes na avaliação genética animal. Foram simuladas cinco estruturas de população com diferentes níveis de *outliers* (0%, 4%, 7%, 10% e 15%), herdabilidade média de 0,40 e três efeitos fixos para a formação dos grupos de contemporâneos. Foi utilizado o programa MTDFREML para predição dos valores genéticos e estimação dos efeitos fixos. As variáveis utilizadas para verificação da acurácia da avaliação foram o Quadrado Médio do Erro na predição dos valores genéticos (QME_VG), o Quadrado Médio do Erro na estimação dos grupos contemporâneos (QME_GC), a correlação de Spearman (CS) entre os valores genéticos verdadeiros e os preditos e as regressões entre: (a) as porcentagens de *outliers* e os quadrados médios dos erros e (2) as porcentagens de *outliers* e as correlações de Spearman. Verificou-se que quanto maior o nível de observações discrepantes, maior foi o QME_GC, o qual variou de 344,07 a 437,46 para 0% e 15% de *outliers*, respectivamente. Este fato foi reforçado pela obtenção de regressão significativa entre os níveis de observações discrepantes e o QME_VG e a Correlação de Spearman. A estimação dos efeitos de grupos de contemporâneos não foi prejudicada pelos níveis de *outliers*. Conclui-se que a presença de observações discrepantes na avaliação genética animal, utilizando a metodologia de modelos mistos, reduz a acurácia da predição dos valores genéticos e a classificação dos indivíduos.

Palavras-chave: análise de dados, avaliação genética, grupos de contemporâneos, melhoramento animal, simulação de dados, *outliers*

Animal genetic evaluation in the presence of discrepant observations

Abstract: The objective of this study was to evaluate the effect of discrepant values in animal genetic evaluation. Five population structures were simulated with different outliers levels (0%, 4%, 7%, 10% and 15%), average heritability of 0.40 and three fixed effects to contemporary groups formation. MTDFREML program was used for genetic values prediction and fixed effects estimation. The variables used for verification of accuracy evaluation were the Mean Square Error in breeding values prediction (QME_VG), the Mean Square Error in the contemporary groups estimation (QME_GC), the Spearman correlation (SC) between the true and predicted breeding values and regressions between: (a) the outliers percentage and the Mean Square of Errors and (b) the outliers percentage and Spearman correlation. It was found the higher was the discrepant observations level, the largest was the QME_GC, which ranged from 344.07 to 437.46, to 0% and 15% outliers, respectively. This fact was enhanced by obtaining of significant regression between the discrepant observations levels and QME_VG and Spearman correlations. The estimation of contemporary group effects was not affected by outliers levels. It is concluded that the presence of discrepant observations in animal genetic evaluation, using mixed models methodology, reduces the prediction breeding values accuracy and ranking of individuals.

Keywords: animal breeding, contemporary groups, data analysis, data simulation, genetic evaluation, outliers

Introdução

A avaliação genética animal utilizando modelos lineares mistos tem como pressuposição básica a distribuição normal, tanto para os resíduos quanto para os demais efeitos aleatórios, o que torna essas análises muito sensíveis à presença de valores discrepantes (*outliers*) nas observações (Rogers & Tukey, 1972). Considera-se como valor discrepante a observação cuja magnitude apresenta baixa probabilidade de pertencer à distribuição normal do conjunto de dados que se está trabalhando.

As possíveis correções para os *outliers* incluem a checagem e a correção de casos extremos dos dados; a retirada de casos extremos; a redefinição da população de interesse ou a reespecificação do

modelo. Barnett e Lewis (1995) definem *outliers* em um conjunto de dados como sendo as observações que parecem ser inconsistentes com o conjunto de dados remanescentes. Para Draper e Smith (1981), a simples exclusão de *outliers* pode não ser o procedimento correto e as regras propostas para rejeição de *outliers* devem incluir a reanálise sem essas observações as quais, dependendo das circunstâncias, podem ser portadoras de informações vitais dos indivíduos de uma população.

Dentro deste contexto, o objetivo do presente trabalho é avaliar o efeito da presença de *outliers* na predição dos valores genéticos e na estimação dos efeitos fixos.

Material e Métodos

Os dados utilizados neste trabalho foram simulados utilizando o software SAS (2000). Foram simuladas cinco estruturas de população com *outliers* (0%, 4%, 7%, 10% e 15%), com herdabilidade média de 0,40 e três efeitos fixos, que foram utilizados para formar os grupos de contemporâneos. A presença de *outliers* foi simulada inserindo-se valores discrepantes nos efeitos genéticos das vacas no processo de simulação, de acordo com a porcentagem desejada (Tabela 1).

Tabela 1 Porcentagens de *outliers*, números de *outliers*, número de repetições e tamanho da população simulada.

Porcentagens de <i>outliers</i>	Nº <i>Outliers</i>	NR	Tamanho da população
0	0	5	10.000
4	400	5	10.000
7	700	5	10.000
10	1.000	5	10.000
15	1.500	5	10.000

A estimação dos efeitos fixos, dos componentes de variâncias e a predição dos valores genéticos foram realizadas por meio do programa MTDFREML (*Multiple Trait Derivative-Free Restricted Maximum Likelihood*), descrito por Boldman et al. (1995).

Para verificar o efeito dos diferentes níveis de observações discrepantes, foi calculado o Quadrado Médio do Erro na predição dos valores genéticos (QME_VG), como a média dos quadrados das diferenças entre os valores genéticos verdadeiros e dos valores genéticos preditos e o Quadrado Médio do Erro na estimação dos GC (QME_GC), como a média dos quadrados das diferenças entre os valores dos grupos de contemporâneos verdadeiros e dos estimados, conforme Harvey (1979), citado por Schenkel (1998):

$$QME = \sum_{i=1}^n \frac{1}{n} (\hat{a}_i - a_i)^2,$$

em que QME é o quadrado médio do erro, n é o número de indivíduos, \hat{a}_i refere-se ao valor genético predito e a_i ao valor genético verdadeiro. Quanto mais próximo de zero maior é a acurácia da predição dos valores genéticos.

Foram calculadas também: (1) a correlação de Spearman (CS) entre os valores genéticos verdadeiros e os preditos e; (2) as regressões lineares: (a) entre as porcentagens de *outliers* e os quadrados médios dos erros dos valores genéticos (QME_VG) e (b) entre as porcentagens de *outliers* e os quadrados médios dos erros dos grupos de contemporâneos (QME_GC).

Resultados e Discussão

Na Tabela 2 encontram-se a média dos Quadrados Médios dos Erros para os valores genéticos preditos, a média dos Quadrados Médios dos Erros para os efeitos de grupos de contemporâneos e a correlação de Spearman entre os valores genéticos verdadeiros e preditos.

De modo geral, é possível verificar que quanto maior foi o nível de observações discrepantes, maior foi o Quadrado Médio do Erro na predição do valor genético, sendo de 344,07 para a estrutura de população sem *outliers* e 437,46 para a estrutura com 15% de observações discrepantes. Este fato foi reforçado pela análise de regressão entre QME_VG e as porcentagens *outliers* que foi significativa ao nível de 5% de probabilidade. A equação obtida foi $QME_VG = 361,61 + 5,89 \times POD$, onde POD é a porcentagem de observações discrepantes. Este resultado é um indicativo da perda de acurácia da predição dos valores na presença de observações discrepantes.

A menor acurácia na predição dos valores genéticos resultou em pior classificação dos indivíduos, sendo obtida correlação de Spearman de 0,78 para a estrutura de população sem a presença de *outliers* e 0,69 para o nível de 15% de observações discrepantes. A análise de regressão entre a CS e as

porcentagens de *outliers* foi significativa ao nível de 5% de probabilidade ($CS=0,7680-0,00514 \times \text{POD}$), demonstrando menor habilidade da metodologia em classificar os indivíduos na presença de observações discrepantes. A regressão entre QME_GC e a porcentagem de observações discrepantes foi não significativa.

Tabela 2 Médias dos Quadrados Médios dos Erros para os valores genéticos preditos (QME_VG), Médias dos Quadrados Médios dos Erros para efeitos dos grupos de contemporâneos (QME_GC) e correlação de Spearman (CS) entre valores genéticos verdadeiros e preditos e Desvio Padrão (DP) para os diferentes níveis de *outliers*.

Porcentagens de <i>outliers</i>	QME_VG ± DP	QME_GC ± DP	CS ± DP
0	344,07 ± 28,04	1.438,57 ± 514,01	0,78 ± 0,02
4	403,79 ± 19,23	1.623,40 ± 358,84	0,74 ± 0,03
7	403,25 ± 45,83	1.560,20 ± 571,17	0,73 ± 0,03
10	431,46 ± 37,92	1.417,40 ± 178,35	0,72 ± 0,02
15	437,46 ± 36,34	1.436,10 ± 199,96	0,69 ± 0,02

Conclusões

A presença de observações discrepantes na avaliação genética animal por meio da metodologia de modelos mistos reduz a acurácia da predição dos valores genéticos e prejudica a correta classificação dos indivíduos.

Literatura citada

- BOLDMAN, K.G., KRIESE, L.A., VAN VLECK, L.D. et al. **A manual for use of MTDFREML. A set of programs to obtain estimates of variances and covariances [DRAFT].** Lincoln: USDA/ARS, 1995.
- HENDERSON, C.R. **Selection index and expected genetic advance. In: STATISTICAL genetics and plant breeding**, Washington, DC: NAS/NRC, 1963.
- SAS Institute Inc. SAS/STAT: User guide. Cary: SAS Institute Inc., 2000. USA.
- SCHENKEL, F.S. Studies on effects of parental selection of estimation of genetic parameters an breeding values of metric traits. Ghelph: University of Ghelph, 1998. 191p. (Ph.d. Thesis) - University of Ghelph, 1998.