



## Estudo de simulação de modelos lineares mistos com distribuição normal contaminada no melhoramento genético animal<sup>1</sup>

Idalmo Garcia Pereira<sup>2</sup>, Henrique Nunes de Oliveira<sup>3</sup>, Guilherme Jordão de Magalhães Rosa<sup>4</sup>

<sup>1</sup> Projeto financiado pela CAPES.

<sup>2</sup> Departamento de Zootecnia - FCA/UFVJM - Diamantina - MG, CEP: 39100-000.

<sup>3</sup> Departamento de Melhoramento e Nutrição Animal - FMVZ - Unesp - Botucatu - SP, CEP: 18618-000.

<sup>4</sup> Department of Dairy Science, University of Wisconsin - Madison, WS, USA.

**RESUMO** - Objetivou-se com esse trabalho comparar estimativas de componentes de variâncias obtidas por meio de modelos lineares mistos Gaussianos e Robustos, via Amostrador de Gibbs, em dados simulados. Foram simulados 50 arquivos de dados com 1.000 animais cada um, distribuídos em cinco gerações, em dois níveis de efeito fixo e três valores fenotípicos distintos para uma característica hipotética, com diferentes níveis de contaminação. Exceto para os dados sem contaminação, quando os modelos foram iguais, o modelo Robusto apresentou melhores estimativas da variância residual. As estimativas de herdabilidade foram semelhantes em todos os modelos, mas as análises de regressão mostraram que os valores genéticos preditos com uso do modelo Robusto foram mais próximos dos valores genéticos verdadeiros. Esses resultados sugerem que o modelo linear normal contaminado oferece uma alternativa flexível para estimação robusta em melhoramento genético animal.

Palavras-chave: componentes de variância, estimação robusta, inferência Bayesiana, valor genético

## Simulation study of linear mixed models with contaminated normal distribution in animal breeding

**ABSTRACT** - The objective of this study was to compare Gaussian and Robust linear mixed models for the estimation of variance components by REML and Gibbs Sampling, using data from fifty simulated populations consisting of 1,000 animals distributed in 5 generations. Two levels of fixed effect and three hypothetical phenotypic values for a trait, with different levels of contamination were used in the simulations. Additive and residual variance estimates were similar for both REML and Bayesian inference using the Gaussian and Robust model. The best estimates of residual variance in the presence of contaminants were obtained by the Robust model. Estimates of heritability were similar for all models, but regression analyses indicated that predicted genetic values obtained by the robust model were more similar to real breeding values. These results suggest that the contaminated normal linear model is a flexible alternative for robust estimation in animal breeding.

Key Words: Bayesian inference, breeding value, robust estimation, variance components

### Introdução

Para a maioria das características estudadas em melhoramento genético animal, utiliza-se o método da máxima verossimilhança restrita (REML) em modelos lineares Gaussianos para estimar os valores dos componentes de (co)variância. Essas estimativas são necessárias para obtenção de preditores dos valores genéticos dos animais para as características de interesse. O método REML é geralmente considerado um dos melhores métodos para estimar componentes de (co)variância em dados de melhoramento animal desbalanceados.

No melhoramento genético animal, para características quantitativas, cujas análises são realizadas por modelos

lineares mistos, supõe-se distribuição normal tanto dos resíduos quanto dos demais efeitos aleatórios, o que torna essas análises muito sensíveis à presença de valores discrepantes nas observações (Rogers & Tukey, 1972). Alternativas para diminuir este efeito são os modelos com distribuições leptocúrticas, que têm se mostrado eficientes e de fácil implementação para estimação robusta em análise de dados em diversas situações (Rosa et al., 2003).

A Inferência Bayesiana permite grande flexibilidade em seus modelos e pode ser utilizada para implementação de vários tipos de distribuição robusta. Essas distribuições não sofrem tanto a influência de valores discrepantes nas observações (*outliers*). Nesse contexto, algumas distribuições leptocúrticas, como a distribuição normal contamina-

da, têm sido utilizadas em estimação robusta (Rosa et al., 2003). Entretanto, são poucos os estudos sobre a utilização destas distribuições para estimação de componentes de (co)variâncias e parâmetros genéticos e predição dos valores genéticos de animais no melhoramento genético.

Um recurso eficiente para realização de comparações entre diferentes metodologias é a análise de dados simulados, a qual permite conhecer os verdadeiros parâmetros das distribuições. Por meio desse processo, é possível gerar dados fenotípicos a partir do valor genético hipotético real dos animais e possibilitar condições necessárias para a comparação das metodologias.

Objetivou-se com este trabalho comparar estimativas de componentes de variância e os valores genéticos preditos obtidos a partir de modelos lineares mistos Gaussianos e Robustos em dados simulados com diferentes níveis de valores discrepantes nas observações.

## Material e Métodos

Foram simulados 50 arquivos de dados amostrais de populações de 1.000 animais cada uma. Todos os animais apresentavam três registros, valores fenotípicos, para uma característica produtiva hipotética. Todo o processo de simulação foi realizado utilizando-se o procedimento IML do sistema SAS (1999), com a seguinte estrutura e simulação:

- foi gerado um conjunto de *pedigree* (animal, pai e mãe) em gerações discretas. Cada população de 1.000 animais foi composta de cinco gerações com um núcleo de dez pais e 50 mães. Todos os indivíduos da primeira geração não tinham ancestrais conhecidos e, nas gerações subsequentes, os pais foram escolhidos ao acaso, entre dez dos machos e 50 das fêmeas da geração anterior;

- o nível médio de endogamia em cada amostra foi de aproximadamente 5,5%;

- para os animais da primeira geração, foram inicialmente gerados valores genéticos e valores residuais, todos sob distribuições normais com média zero e variâncias iguais, respectivamente, a 36,0 e 64,0;

- posteriormente, os animais foram aleatoriamente distribuídos em dois grupos de efeito sistemático e, com probabilidade determinada por  $\phi$  (0,20 ou 0,40), poderiam ser alocados a uma população contaminante ou não-contaminante;

- um valor correspondente ao valor fenotípico dos animais foi formado a partir da soma do valor genético mais o valor residual e mais um valor fixo (510 para os animais atribuídos ao primeiro grupo de efeitos sistemáticos e 520 aos atribuídos ao segundo grupo de efeitos sistemáticos).

Esse primeiro valor fenotípico, identificado por VF, representou uma população com distribuição normal dos resíduos, sem contaminantes;

- mais dois valores fenotípicos identificados como VF1 e VF2 foram gerados de modo semelhante ao primeiro, porém, o valor residual dos animais alocados ao grupo de contaminantes foi dividido pela raiz de determinado valor  $\tau$  (entre 0,0625 e 0,125) antes de ser somado aos demais. Para VF1, o valor de  $\tau$  foi sempre maior que para VF2; e

- para os animais das gerações subsequentes, o procedimento adotado foi semelhante, no entanto, o valor genético dos animais foi gerado como a soma da média dos valores genéticos dos pais mais a segregação amostral mendeliana.

Desse modo, para cada população foram simulados três conjuntos de dados a serem analisados. O primeiro apresentava distribuição normal sem contaminantes e cada um dos outros dois apresentava proporção de contaminantes igual a  $\phi$  com variância residual igual àquela da população original dividida por  $\tau$ .

A população assim formada, com o agrupamento de duas populações com distribuição normal dos resíduos, porém com variâncias diferentes, caracteriza uma população normal contaminada, com formato leptocúrtico (cauda longa). Foram utilizadas combinações de dois valores de  $\phi$  e cinco valores de  $\tau$  para VF1 e para VF2, resultando em um total de dez conjuntos de parâmetros para cada um dos quais foram geradas cinco repetições (Tabela 1).

Para cada um dos arquivos de dados gerados, foram estimados os componentes de variância por máxima verossimilhança restrita. Estas estimativas foram usadas para obtenção dos preditores dos valores genéticos para as características estudadas. Utilizou-se o software *Multiple Trait Derivative Free Restricted Maximum Likelihood* (MTDFREML), descrito por Boldman et al. (1995).

Para efeito de comparação dos resultados obtidos pela estimação por máxima verossimilhança com os reais presentes nos arquivos de dados, uma vez que os fenótipos foram gerados a partir dos valores genéticos, com variâncias genética aditiva e residual conhecidas, forneceram-se como *priors* para a análise pelo MTDFREML, os valores reais das variâncias. Os *priors*, para cada um dos três valores fenotípicos da característica foram os mesmos, isto é, 36,0 para a variância genética aditiva ( $\sigma_a^2$ ) e 64,0 para a variância residual ( $\sigma_e^2$ ). O critério de convergência utilizado foi a variância da função de verossimilhança, estabelecida quando esta foi menor que  $10^{-9}$ . A cada convergência, o programa era reiniciado usando como *priors* os valores obtidos no processamento anterior até que ocorresse a estabilização das estimativas dos componentes de variância.

Tabela 1 - Conjunto de parâmetros utilizados para gerar os valores fenotípicos identificados como VF1 e VF2, com probabilidade de contaminantes e ponderação do valor residual determinado por  $\phi$  e  $\tau^*$ , respectivamente

Table 1 - Set of parameters used to generate the "phenotypic" values identified as VF1 and VF2, with probability of contaminants and weighting of the residual value determined by  $\phi$  and  $\tau^*$ , respectively

Conjunto de parâmetros Set of parameters	$\phi$			
	0,2		0,4	
	$\tau_1$	$\tau_2$	$\tau_1$	$\tau_2$
1	0,25	0,0625	—	—
2	0,21875	0,09375	—	—
3	0,1875	0,125	—	—
4	0,171875	0,140625	—	—
5	0,15625	0,1484375	—	—
6	—	—	0,25	0,0625
7	—	—	0,21875	0,09375
8	—	—	0,1875	0,125
9	—	—	0,171875	0,140625
10	—	—	0,15625	0,1484375

\*  $\tau_1$  = ponderação do valor residual para a VF1 e  $\tau_2$  = ponderação do valor residual para a VF2.

\*  $\tau_1$  = weight of the residual value for VF1 and  $\tau_2$  = weight of the residual value for VF2.

O modelo linear misto utilizado para descrever as observações, em termos matriciais, foi o seguinte:

$$\underset{\sim}{y} = \underset{\sim}{X}\underset{\sim}{b} + \underset{\sim}{Z}\underset{\sim}{a} + \underset{\sim}{e}$$

em que:  $\underset{\sim}{y}$  = vetor das observações;  $\underset{\sim}{X}$  = matriz de incidência dos efeitos fixos;  $\underset{\sim}{b}$  = vetor de efeitos fixos;  $\underset{\sim}{Z}$  = matriz de incidência dos efeitos aleatórios genéticos diretos;  $\underset{\sim}{a}$  = vetor dos efeitos aleatórios dos valores genéticos diretos do animal;  $\underset{\sim}{e}$  = vetor dos erros associados a cada observação.

Assumiu-se que os efeitos genéticos aditivos e residuais não eram correlacionados entre si e que  $E[\underset{\sim}{a}] = 0$  e  $E[\underset{\sim}{e}] = 0$ . Assim:

$$E[\underset{\sim}{y} | \underset{\sim}{a}] = \underset{\sim}{X}\underset{\sim}{b} + \underset{\sim}{Z}\underset{\sim}{a}$$

$$\text{Var} \begin{bmatrix} \underset{\sim}{a} \\ \underset{\sim}{e} \end{bmatrix} = \begin{bmatrix} A\mathbf{S}_a^2 & 0 \\ 0 & I\mathbf{S}_e^2 \end{bmatrix}$$

em que:  $\mathbf{S}_a^2$  = variância genética aditiva direta;  $A$  = matriz dos coeficientes de parentesco entre os animais;  $\mathbf{S}_e^2$  = variância residual;  $I$  = matriz identidade de ordem apropriada.

Para estimação dos componentes de variância e predição dos valores genéticos dos animais por Inferência

Bayesiana sob modelo Gaussiano, utilizou-se o *software Multiple Trait Gibbs Sampling in Animal Models* (MTGSAM), desenvolvido por Van Tassel & Van Vleck (1995), utilizado também na análise por máxima verossimilhança restrita. Este *software* foi desenvolvido considerando, no modelo animal simples utilizado neste trabalho, as seguintes distribuições *a priori* (Van Tassel & Van Vleck, 1995): para o conjunto de efeitos "fixos", uma distribuição não-informativa; para o conjunto de valores genéticos, uma distribuição normal multivariada com média 0 e matriz de variância e covariâncias igual ao produto da matriz de parentesco pela variância genética ( $A\sigma_a^2$ ); e, para as variâncias genética e residual, distribuições de qui-quadrado ( $\chi^2$ ) escalonadas invertidas com parâmetros  $\nu = -2$  e  $S^2 = 0$ , pois, de acordo com Carneiro Jr. et al. (2005), assumindo esses valores, tem-se uma distribuição uniforme, que indica que a probabilidade de ocorrência de qualquer valor no intervalo é a mesma, caracterizando assim uma *priori* não-informativa.

As variâncias utilizadas como hiperparâmetros das *priors* correspondem às estimativas fornecidas por meio de uma análise prévia realizada no MTDFREML. O período de descarte amostral (*thinning interval*) adotado foi de 1.000 ciclos, que representa o número de ciclos que devem ser descartados visando à eliminação de possíveis correlações entre valores gerados subsequentemente. O total de amostras geradas foi de 250 mil, portanto, eliminando as 1.000 primeiras observações (*burn-in*) para remover o efeito dos valores iniciais e considerando esse intervalo de descarte amostral, foram aproveitadas para análise 249 amostras das distribuições marginais *a posteriori*. Para avaliação da convergência das cadeias, utilizou-se o critério de avaliação gráfica recomendado por Gelman et al. (1995), que consiste em testar vários tamanhos de cadeias e, simultaneamente, observar se ocorre alguma alteração na trajetória do gráfico. Caso esta trajetória se mantenha constante, tem-se a obtenção da convergência.

Como alternativa robusta ao modelo Gaussiano utilizado nas análises anteriores, utilizou-se a distribuição normal contaminada para os resíduos. Nesse caso, cada elemento do vetor de resíduos foi definido como  $E_i = e_i / \sqrt{W_i}$ , em que  $e_i \stackrel{iid}{\sim} N(0, \mathbf{S}_e^2)$ ;  $W_i$  é uma variável aleatória com distribuição de probabilidade  $p(W_i/\nu)$ ; e  $\nu$  refere-se aos parâmetros de robustez (Rosa et al., 2003).

Na aplicação desse modelo, foram utilizadas as mesmas distribuições *a priori* consideradas para o modelo Gaussiano e, para os dois parâmetros de robustez ( $\phi$ ,  $\tau$ ), foi adotada a sugestão de distribuição uniforme (0, 1) e beta ( $a$ ,  $b$ ), respectivamente, como descrito por Rosa et al. (2003).

Para estas análises, foi desenvolvido um programa de computador específico em Fortran 77 utilizando-se como base um programa desenvolvido pelo pesquisador Daniel Sorensen, no Instituto Dinamarquês de Ciência Animal, para análise Gaussiana.

Todos os resultados obtidos para os dez conjuntos de parâmetros utilizando o método da máxima verossimilhança restrita e amostragem de Gibbs (contexto Bayesiano) em modelo animal Gaussiano foram comparados entre si e aos resultados obtidos pelo modelo Robusto, via amostragem de Gibbs.

Utilizaram-se como critério para comparação dos resultados as estimativas dos componentes de variância, dos parâmetros da normal contaminada ( $\tau$  e  $\phi$ ) e, principalmente, o coeficiente de determinação ( $R^2$ ) e o quadrado médio do resíduo da análise de regressão dos valores genéticos verdadeiros (simulados) sobre os preditos por REML e Inferência Bayesiana em modelo Gaussiano, e os preditos por Inferência Bayesiana em modelo Robusto.

## Resultados e Discussão

As estimativas médias dos parâmetros da distribuição normal contaminada  $\tau$  e  $\phi$  (Tabelas 2 a 6), em algumas situações, apresentaram valores que oscilaram entre um pouco abaixo e um pouco acima dos valores reais utilizados na simulação para geração dos arquivos de dados e alguns conjuntos de VF, nos quais não ocorre contaminação. Portanto, em vez de  $\tau = 1$  e/ou  $\phi = 0$ , obtiveram-se algumas estimativas intermediárias para ambos os parâmetros. A explicação para esta variação é que, em alguns dos cinco arquivos (repetições) usados para se obter a média das distribuições, ocorreram valores um pouco mais altos de  $\phi$  e  $\tau$  próximo de 1,0 (o que denota distribuição normal), enquanto outros apresentavam valores mais baixos de  $\tau$  e  $\phi$  próximos de zero (o que também denota distribuição normal). Desse modo, a média dos dois parâmetros em todas as repetições assumiu valores intermediários. Todavia, dentro do mesmo conjunto de dados (repetição), as estimativas de  $\tau$  e  $\phi$  garantiram o bom ajuste do modelo Robusto ao conjunto de dados. Em alguns casos, as estimativas de  $\tau$  e  $\phi$  diferenciaram um pouco dos valores esperados, que foram os utilizados na simulação, indicando, talvez, pior qualidade de ajuste do modelo como quando se observam os valores estimados de  $\tau$  para a VF1 (Tabelas 2 e 4), que diferem aproximadamente 8% do valor de  $\tau(0,25)$  considerado na simulação.

As estimativas da  $S_e^2$  obtidas por REML apresentaram valores mais elevados que os simulados para o resíduo da população-base. Como previsto, esses valores cresceram

de acordo com o grau de contaminação, aumentando da VF para a VF2 no mínimo 102% (Tabela 3, 4º conjunto) e no máximo de 618% (Tabela 4, 6º conjunto). Essa mesma tendência também foi observada na estimativa média obtida da distribuição *a posteriori* por Inferência Bayesiana. Na verdade, estas estimativas são de valores médios de variância residual das duas subpopulações que compõem a amostra no caso do modelo Gaussiano. Este modelo, tanto do ponto de vista Frequentista como do ponto de vista Bayesiano, não discrimina as subpopulações. Por outro lado, as estimativas de  $S_a^2$  obtidas por REML e Inferência Bayesiana em modelos Gaussianos foram menos influenciadas, o que poderia ser esperado, uma vez que a contaminação ocorreu apenas na variância residual. Desse modo, as estimativas de  $h^2$  foram menores para a característica com VF1 e VF2 em relação ao VF não contaminado.

Comparando também estimativas de componentes de variância obtidas por REML e Inferência Bayesiana em dados simulados na ausência de contaminação, Van Tassel et al. (1995) e Van Tassel & Van Vleck (1996) encontraram estimativas semelhantes para os componentes de variância entre os dois métodos. No entanto, Wright et al. (2000), trabalhando com dados de experimento de seleção de *Tribolium castaneum*, verificaram que a estimativa da variância genética aditiva obtida por REML diferenciava muito de sua média na distribuição *a posteriori* obtida por Inferência Bayesiana e se assemelhava mais à moda da mesma distribuição.

As estimativas da variância residual pelo modelo Robusto podem ser obtidas para as duas subpopulações: não-contaminante ( $S_e^2$ ) e contaminante ( $S_e^2/t$ ). Verifica-se que as estimativas de  $S_e^2$  obtidas pelo modelo Robusto foram próximas das reais utilizadas na simulação (Tabelas 2 a 6). As estimativas de  $S_e^2/t$  foram também próximas das geradas e aproximadamente 3 a 13 vezes maiores que a dos não-contaminantes, o que resultou em menores estimativas de herdabilidade ( $h^{2*}$ ) para essas subpopulações, prejudicando os valores genéticos preditos dos indivíduos contaminantes, conforme relatado por Strandén & Gianola (1999).

Os desvios-padrão aditivo e residual das distribuições *a posteriori* em modelo Gaussiano aumentaram de acordo com o aumento do grau de contaminação (Tabelas de 2 a 6). Os desvios-padrão aditivos para a VF foram sempre menores que os dos outros dois valores fenotípicos contendo contaminantes, independentemente do modelo utilizado. Esse resultado era esperado, pois, na ausência de contaminantes, os dados caracterizam-se como de apenas uma população, denotando uma única distribuição. Na ausência de contaminação (VF), os desvios-padrão aditivos

Tabela 2 - Resultados considerados para comparação dos modelos lineares mistos Gaussiano e Robustos, para o 1<sup>o</sup> ( $\phi = 0.2$ ;  $\tau_1 = 0.25$  e  $\tau_2 = 0.0625$ , respectivamente, para VF1 e VF2) e o 2<sup>o</sup> ( $\phi = 0.2$ ;  $\tau_1 = 0.21875$  e  $\tau_2 = 0.09375$ , respectivamente para VF1 e VF2) conjunto de parâmetros, com variâncias genética aditiva e residual médias iguais a 36,0 e 64,0, respectivamente  
 Table 2 - Results used for comparison of Gaussian and Robust mixed linear models, for the 1<sup>st</sup> ( $\phi = 0.2$ ;  $\tau_1 = 0.25$  and  $\tau_2 = 0.0625$  respectively, for VF1 and VF2) and 2<sup>nd</sup> set of parameters ( $\phi = 0.2$ ;  $\tau_1 = 0.21875$  and  $\tau_2 = 0.09375$ , respectively for VF1 and VF2), using averages for additive genetics and residual variances equal to 36.0 and 64.0, respectively

Modelo/Método <sup>a</sup> Model/Method <sup>a</sup>	Valor fenotípico <sup>b</sup> Phenotypic value <sup>b</sup>	R <sup>2</sup>				Estimativa <sup>d</sup> Estimate						Desvio-padrão <sup>e</sup> Standard deviation		
		Mínimo Minimum	Máximo Maximum	Médio Mean	QMR <sup>c</sup> MSE <sup>c</sup>	$\sigma^2_a$	$\sigma^2_e$	$\sigma^2_e/\tau$	h <sup>2</sup>	h <sup>2*</sup>	$\tau$	$\phi$	$\sigma_a$	$\sigma_e$
1 <sup>o</sup> conjunto 1 <sup>st</sup> set														
G/REML	VF	0,39	0,58	0,49	17,0029	31,67	66,35	—	0,32	—	—	—	—	—
	VF1	0,28	0,51	0,40	19,7917	30,95	105,39	—	0,23	—	—	—	—	—
	VF2	0,12	0,36	0,26	24,6615	33,38	258,62	—	0,12	—	—	—	—	—
G/Bayes	VF	0,39	0,58	0,48	17,0354	32,87	65,82	—	0,33	—	—	7,67	5,74	
	VF1	0,28	0,51	0,40	19,8042	33,36	104,61	—	0,24	—	—	9,64	8,11	
	VF2	0,13	0,36	0,26	24,6552	36,23	257,24	—	0,12	—	—	16,08	17,15	
R/Bayes	VF	0,39	0,58	0,49	16,9714	33,34	63,31	401,71	0,35	0,08	0,1576	0,0494	7,39	7,99
	VF1	0,29	0,53	0,42	19,2320	33,75	79,94	348,02	0,30	0,09	0,2297	0,1352	9,35	15,98
	VF2	0,27	0,50	0,38	20,5507	28,88	65,01	910,50	0,30	0,03	0,0714	0,2311	8,73	8,87
2 <sup>o</sup> conjunto 2 <sup>nd</sup> set														
G/REML	VF	0,50	0,56	0,53	17,2870	39,01	62,24	—	0,38	—	—	—	—	
	VF1	0,42	0,46	0,44	20,6537	40,73	105,98	—	0,28	—	—	—	—	
	VF2	0,33	0,37	0,36	23,8337	43,50	180,39	—	0,20	—	—	—	—	
G/Bayes	VF	0,50	0,56	0,53	17,4132	40,29	61,68	—	0,40	—	—	7,85	5,59	
	VF1	0,42	0,46	0,44	20,8139	42,59	105,00	—	0,29	—	—	10,31	8,38	
	VF2	0,33	0,36	0,35	24,0307	46,06	178,89	—	0,21	—	—	14,08	13,05	
R/Bayes	VF	0,51	0,56	0,53	17,3061	40,96	58,72	302,52	0,41	0,12	0,1941	0,0574	7,83	7,98
	VF1	0,43	0,50	0,46	19,8981	44,38	62,35	285,88	0,42	0,13	0,2181	0,2132	9,58	14,19
	VF2	0,42	0,49	0,45	20,3905	43,77	61,19	640,73	0,42	0,06	0,0955	0,2053	9,93	9,40

a G/REML = Gaussiano/Máxima Verossimilhança Restrita; G/Bayes = Gaussiano/Inferência Bayesiana; R/Bayes = Robusto (Normal Contaminada) / Inferência Bayesiana.  
 b VF = valor fenotípico sem contaminantes; VF1 = valor fenotípico com contaminação dada por  $\tau_1$ ; VF2 = valor fenotípico com contaminação dada por  $\tau_2$ .  
 c QMR = quadrado médio do resíduo da regressão dos valores genéticos verdadeiros sobre os valores genéticos preditos.  
 d  $\sigma^2_a$  = variância genética aditiva;  $\sigma^2_e$  = variância residual da população não-contaminante;  $\sigma^2_e/\tau$  = variância residual da população contaminante; h<sup>2</sup> = herdabilidade da população não-contaminante; h<sup>2\*</sup> = herdabilidade da população contaminante;  $\tau$  e  $\phi$  = parâmetros da distribuição normal contaminada.  
 e  $\sigma_a$  = desvio-padrão genético aditivo;  $\sigma_e$  = desvio-padrão residual.  
 a G/REML = Gaussian/Restricted Maximum Likelihood; G/Bayes = Gaussian/Bayesian Inference; R/Bayes = Robust (Contaminated Normal)/Bayesian Inference.  
 b VF = phenotypic value without contamination; VF1 = phenotypic value with contamination given for  $\tau_1$ ; VF2 = phenotypic value with contamination given for  $\tau_2$ .  
 c MSE = Mean square error of the regression of true genetic values on predicted genetic values.  
 d  $\sigma^2_a$  = additive genetic variance;  $\sigma^2_e$  = residual variance of the not contaminated population;  $\sigma^2_e/\tau$  = residual variance of the contaminated population; h<sup>2</sup> = heritability of the not contaminated population; h<sup>2\*</sup> = heritability of the contaminated population;  $\tau$  and  $\phi$  = parameters of contaminated normal distribution.  
 e  $\sigma_a$  = additive genetic standard deviation;  $\sigma_e$  = residual standard deviation.

Tabela 3 - Resultados considerados para comparação dos modelos lineares mistos Gaussianos e Robustos, para o 3<sup>o</sup> ( $\phi = 0,2$ ;  $\tau_1 = 0,1875$  e  $\tau_2 = 0,125$ , respectivamente, para VF1 e VF2) e o 4<sup>o</sup> ( $\phi = 0,2$ ;  $\tau_1 = 0,171875$  e  $\tau_2 = 0,140625$ , respectivamente, para VF1 e VF2) conjunto de parâmetros, com variâncias genética aditiva e residual médias iguais a 36,0 e 64,0, respectivamente  
 Table 3 - Results used for comparison of Gaussian and Robust mixed linear models, for the 3<sup>rd</sup> ( $\phi = 0,2$ ;  $\tau_1 = 0,1875$  and  $\tau_2 = 0,125$ , respectively for VF1 and VF2) and 4<sup>th</sup> sets of parameters ( $\phi = 0,2$ ;  $\tau_1 = 0,171875$  and  $\tau_2 = 0,140625$ , respectively for VF1 and VF2), using averages for additive genetic and residual variances equal to 36.0 and 64.0, respectively

Modelo/Método <sup>a</sup> Model/Method <sup>a</sup>	Valor fenotípico <sup>b</sup> Phenotypic value <sup>b</sup>	R <sup>2</sup>		Estimativa <sup>d</sup> Estimate						Desvio-padrão <sup>e</sup> Standard deviation			
		Mínimo Minimum	Máximo Maximum	Médio Mean	QMR <sup>c</sup> MSE <sup>c</sup>	$\sigma^2_a$	$\sigma^2_e$	$\sigma^2_e/\tau$	h <sup>2</sup> *	$\tau$	$\phi$	$\sigma_a$	$\sigma_e$
3 <sup>o</sup> conjunto 3 <sup>rd</sup> group													
G/REML	VF	0,46	0,63	0,53	17,1657	37,46	62,29	—	0,37	—	—	—	—
	VF1	0,35	0,54	0,42	21,1448	39,14	118,00	—	0,25	—	—	—	—
	VF2	0,31	0,51	0,38	22,5918	40,45	152,30	—	0,21	—	—	—	—
G/Bayes	VF	0,46	0,63	0,52	17,2247	38,69	61,75	—	0,38	—	—	7,99	5,60
	VF1	0,34	0,55	0,42	21,2268	41,09	116,96	—	0,26	—	—	10,58	8,88
	VF2	0,31	0,51	0,38	22,6839	42,73	151,03	—	0,22	—	—	12,15	10,93
R/Bayes	VF	0,46	0,63	0,53	17,1802	38,90	61,87	198,94	0,39	0,16	0,3110	0,0868	7,47
	VF1	0,37	0,56	0,44	20,2972	40,04	63,85	396,34	0,38	0,09	0,1611	0,1732	9,03
	VF2	0,36	0,55	0,43	20,6844	39,70	66,24	569,07	0,37	0,06	0,1164	0,1773	9,25
4 <sup>o</sup> conjunto 4 <sup>th</sup> group													
G/REML	VF	0,46	0,57	0,53	17,2999	34,26	63,81	—	0,35	—	—	—	—
	VF1	0,31	0,48	0,41	21,4382	35,44	115,19	—	0,24	—	—	—	—
	VF2	0,28	0,46	0,39	22,1909	35,64	129,12	—	0,21	—	—	—	—
G/Bayes	VF	0,46	0,57	0,52	17,3694	35,43	63,26	—	0,36	—	—	7,44	5,32
	VF1	0,30	0,48	0,41	21,5941	37,19	114,26	—	0,24	—	—	10,28	8,64
	VF2	0,27	0,46	0,39	22,3563	37,55	128,08	—	0,23	—	—	10,96	9,51
R/Bayes	VF	0,46	0,57	0,52	17,3150	36,61	58,04	318,38	0,39	0,10	0,1823	0,0903	7,66
	VF1	0,34	0,48	0,43	20,7853	36,23	66,73	358,57	0,35	0,09	0,1861	0,1909	8,65
	VF2	0,32	0,47	0,42	21,0407	34,70	68,79	428,86	0,34	0,07	0,1604	0,1866	8,79

<sup>a</sup> G/REML = Gaussiano/Máxima Verossimilhança Restrita; G/Bayes = Gaussiano/Inferência Bayesiana; R/Bayes = Robusto (Normal Contaminada) / Inferência Bayesiana.

<sup>b</sup> VF = valor fenotípico sem contaminantes; VF1 = valor fenotípico com contaminação dada por  $\tau_1$ ; VF2 = valor fenotípico com contaminação dada por  $\tau_2$ .

<sup>c</sup> QMR = quadrado médio do resíduo da regressão dos valores genéticos verdadeiros sobre os valores genéticos preditos.

<sup>d</sup>  $\sigma^2_a$  = variância genética aditiva;  $\sigma^2_e$  = variância residual da população não-contaminante;  $\sigma^2_e/\tau$  = variância residual da população contaminante;  $h^2$  = herdabilidade da população não-contaminante;  $h^2$  = herdabilidade da população contaminante;  $\tau$  e  $\phi$  = parâmetros da distribuição normal contaminada.

<sup>e</sup>  $\sigma_a$  = desvio-padrão genético aditivo;  $\sigma_e$  = desvio-padrão residual.

<sup>a</sup> G/REML = Gaussian/Restricted Maximum Likelihood; G/Bayes = Gaussian/Bayesian Inference; R/Bayes = Robust (Contaminated Normal) / Bayesian Inference.

<sup>b</sup> VF = phenotypic value without contamination; VF1 = phenotypic value with contamination given for  $\tau_1$ ; VF2 = phenotypic value with contamination given for  $\tau_2$ .

<sup>c</sup> MSE = Mean square error of the regression of true genetic values on predicted genetic values.

<sup>d</sup>  $\sigma^2_a$  = additive genetic variance;  $\sigma^2_e$  = residual variance of the not contaminated population;  $\sigma^2_e/\tau$  = residual variance of the contaminated population;  $h^2$  = heritability of the not contaminated population;  $h^2$  = heritability of the contaminated population;  $\tau$  and  $\phi$  = parameters of contaminated normal distribution.

<sup>e</sup>  $\sigma_a$  = additive genetic standard deviation;  $\sigma_e$  = residual standard deviation.



**Tabela 4 - Resultados considerados para comparação dos modelos lineares mistos Gaussianos e Robustos para o 5<sup>o</sup> ( $\phi = 0.2$ ;  $\tau_1 = 0.15625$  e  $\tau_2 = 0.1484375$ , respectivamente, para VF1 e VF2) e o 6<sup>o</sup> ( $\phi = 0.4$ ;  $\tau_1 = 0.25$  e  $\tau_2 = 0.0625$ , respectivamente, para VF1 e VF2) conjunto de parâmetros, com variâncias genética aditiva e residual médias iguais a 36,0 e 64,0, respectivamente**  
**Table 4 - Results used for comparison of Gaussian and Robust mixed linear models, for the 5<sup>th</sup> ( $\phi = 0.2$ ;  $\tau_1 = 0.15625$  and  $\tau_2 = 0.1484375$ , respectively for VF1 and VF2) and 6<sup>th</sup> sets of parameters ( $\phi = 0.4$ ;  $\tau_1 = 0.25$  and  $\tau_2 = 0.0625$ , respectively for VF1 and VF2), using averages for additive genetic and residual variances equal to 36.0 and 64.0, respectively**

Modelo/Método <sup>a</sup> Model/Method <sup>a</sup>	Valor fenotípico <sup>b</sup> Phenotypic value <sup>b</sup>	R <sup>2</sup>			Estimativa <sup>d</sup> Estimate					Desvio-padrão <sup>e</sup> Standard deviation					
		Mínimo Minimum	Máximo Maximum	Médio Mean	QMR <sup>c</sup> MSE <sup>c</sup>	$\sigma^2_a$	$\sigma^2_e$	$\sigma^2_e/\tau$	h <sup>2</sup>	h <sup>2*</sup>	$\tau$	$\phi$	$\sigma_a$	$\sigma_e$	
<b>5<sup>o</sup> conjunto</b> <i>5<sup>th</sup> group</i>															
G/REML	VF	0,45	0,56	0,51	17,2492	39,33	—	—	0,39	—	—	—	—	—	—
	VF1	0,33	0,45	0,38	21,7178	44,05	—	—	0,26	—	—	—	—	—	—
G/Bayes	VF	0,45	0,56	0,50	17,3204	40,47	—	—	0,40	—	—	—	8,21	5,68	
	VF1	0,33	0,45	0,38	21,8087	46,01	—	—	0,27	—	—	—	12,10	9,70	
R/Bayes	VF	0,45	0,56	0,51	17,2500	40,33	496,55	0,07	0,39	0,1245	0,0013	—	7,90	5,64	
	VF1	0,38	0,50	0,44	20,2256	44,18	361,44	0,11	0,42	0,1717	0,2215	—	10,65	12,43	
	VF2	0,37	0,49	0,42	20,2441	45,19	382,93	0,10	0,42	0,1611	0,2194	—	11,37	12,12	
<b>6<sup>o</sup> conjunto</b> <i>6<sup>th</sup> group</i>															
G/REML	VF	0,45	0,60	0,54	17,2223	37,00	—	—	0,37	—	—	—	—	—	
	VF1	0,32	0,48	0,41	22,0790	37,80	—	—	0,21	—	—	—	—	—	
G/Bayes	VF	0,45	0,60	0,53	17,3040	38,17	—	—	0,37	—	—	—	7,97	5,64	
	VF1	0,32	0,48	0,40	22,1855	39,84	—	—	0,22	—	—	—	11,82	10,37	
R/Bayes	VF	0,45	0,60	0,54	17,2344	38,89	297,18	0,11	0,39	0,2055	0,0522	—	20,82	26,90	
	VF1	0,35	0,49	0,42	21,6080	42,80	331,24	0,11	0,37	0,2247	0,3257	—	11,18	34,32	
	VF2	0,26	0,43	0,36	24,0095	37,20	1076,31	0,03	0,32	0,0722	0,3820	—	12,00	13,32	

<sup>a</sup> G/REML = Gaussiano/Máxima Verossimilhança Restrita; G/Bayes = Gaussiano/Inferência Bayesiana; R/Bayes = Robusto (Normal Contaminada) / Inferência Bayesiana.  
<sup>b</sup> VF = valor fenotípico sem contaminantes; VF1 = valor fenotípico com contaminação dada por  $\tau_1$ ; VF2 = valor fenotípico com contaminação dada por  $\tau_2$ .  
<sup>c</sup> QMR = quadrado médio do resíduo da regressão dos valores genéticos verdadeiros sobre os valores genéticos preditos.  
<sup>d</sup>  $\sigma^2_a$  = variância genética aditiva;  $\sigma^2_e$  = variância residual da população não-contaminante;  $\sigma^2_e/\tau$  = variância residual da população contaminante; h<sup>2</sup> = herdabilidade da população não-contaminante; h<sup>2\*</sup> = herdabilidade da população contaminante;  $\tau$  e  $\phi$  = parâmetros da distribuição normal contaminada.  
<sup>e</sup>  $\sigma_a$  = desvio-padrão genético aditivo;  $\sigma_e$  = desvio-padrão residual.  
<sup>f</sup> G/REML = Gaussian/Restricted Maximum Likelihood; G/Bayes = Gaussian/Bayesian Inference; R/Bayes = Robust (Contaminated Normal)/Bayesian Inference.  
<sup>g</sup> VF = phenotypic value without contamination; VF1 = phenotypic value with contamination given for  $\tau_1$ ; VF2 = phenotypic value with contamination given for  $\tau_2$ .  
<sup>h</sup> MSE = Mean square error of the regression of true genetic values on predicted genetic values.  
<sup>i</sup>  $\sigma^2_a$  = additive genetic variance;  $\sigma^2_e$  = residual variance of the not contaminated population; h<sup>2</sup> = heritability of the not contaminated population; h<sup>2\*</sup> = heritability of the contaminated population;  $\tau$  and  $\phi$  = parameters of contaminated normal distribution.  
<sup>j</sup>  $\sigma_a$  = additive genetic standard deviation;  $\sigma_e$  = residual standard deviation.

Tabela 5 - Resultados considerados para comparação dos modelos lineares mistos Gaussianos e Robustos, para o 7<sup>o</sup> ( $\phi = 0,4$ ;  $\tau_1 = 0,21875$  e  $\tau_2 = 0,09375$ , respectivamente, para VF1 e VF2) e o 8<sup>o</sup> ( $\phi = 0,4$ ;  $\tau_1 = 0,1875$  e  $\tau_2 = 0,125$ , respectivamente, para VF1 e VF2) conjuntos de parâmetros, com variâncias genética aditiva e residual médias iguais a 36,0 e 64,0, respectivamente  
 Table 5 - Results used for comparison of Gaussian and Robust mixed linear models, for the 7<sup>th</sup> ( $\phi = 0,4$ ;  $\tau_1 = 0,21875$  and  $\tau_2 = 0,09375$ , respectively for VF1 and VF2) and 8<sup>th</sup> sets of parameters ( $\phi = 0,4$ ;  $\tau_1 = 0,1875$  and  $\tau_2 = 0,125$ , respectively for VF1 and VF2), using averages for additive genetic and residual variances equal to 36.0 and 64.0, respectively

Modelo/Método <sup>a</sup> Model/Method <sup>a</sup>	Valor fenotípico <sup>b</sup> Phenotypic value <sup>b</sup>	R <sup>2</sup>		Estimativa <sup>d</sup> Estimate										Desvio-padrão <sup>e</sup> Standard deviation	
		Mínimo Minimum	Máximo Maximum	Médio Mean	QMR <sup>c</sup> MSE <sup>c</sup>	$\sigma_a^2$	$\sigma_e^2$	$\sigma_e^2/\tau$	h <sup>2</sup>	h <sup>2*</sup>	$\tau$	$\phi$	$\sigma_a$	$\sigma_e$	
7 <sup>o</sup> conjunto 7 <sup>th</sup> group															
G/REML	VF	0,47	0,54	0,51	17,5961	35,38	65,52	—	0,35	—	—	—	—	—	—
	VF1	0,32	0,39	0,36	23,0340	36,41	160,92	—	0,19	—	—	—	—	—	—
	VF2	0,22	0,29	0,26	26,8227	33,65	326,76	—	0,09	—	—	—	—	—	—
G/Bayes	VF	0,47	0,54	0,51	17,6853	36,71	64,89	—	0,36	—	—	—	7,73	5,51	
	VF1	0,31	0,39	0,36	23,1443	38,50	159,80	—	0,19	—	—	—	12,51	11,53	
	VF2	0,22	0,28	0,25	26,9321	36,52	325,15	—	0,10	—	—	—	18,09	20,94	
R/Bayes	VF	0,47	0,54	0,51	17,6073	36,68	63,82	421,25	0,36	0,08	0,1515	0,0156	7,50	7,00	
	VF1	0,32	0,40	0,37	22,5844	39,73	68,96	320,74	0,37	0,11	0,2150	0,4049	12,37	20,42	
	VF2	0,27	0,34	0,32	24,6556	36,46	69,96	681,87	0,34	0,05	0,1026	0,4163	12,96	14,01	
8 <sup>o</sup> conjunto 8 <sup>th</sup> group															
G/REML	VF	0,39	0,60	0,49	17,2900	36,30	63,90	—	0,36	—	—	—	—	—	
	VF1	0,17	0,45	0,30	23,8971	35,44	180,06	—	0,16	—	—	—	—	—	
	VF2	0,12	0,40	0,25	25,7211	35,83	250,81	—	0,13	—	—	—	—	—	
G/Bayes	VF	0,39	0,60	0,49	17,3399	37,40	63,48	—	0,37	—	—	—	8,31	5,80	
	VF1	0,17	0,44	0,30	23,9164	37,67	179,01	—	0,18	—	—	—	13,33	12,66	
	VF2	0,12	0,39	0,25	25,7420	38,65	249,38	—	0,14	—	—	—	15,88	16,68	
R/Bayes	VF	0,39	0,60	0,49	17,2750	37,77	63,49	462,75	0,37	0,07	0,1372	0,0014	8,06	5,94	
	VF1	0,25	0,47	0,33	22,7776	41,31	62,48	353,59	0,40	0,10	0,1767	0,4125	13,62	16,80	
	VF2	0,19	0,45	0,30	23,9759	40,00	65,77	506,70	0,38	0,07	0,1298	0,4147	11,93	14,01	

<sup>a</sup> G/REML = Gaussiano/Máxima Verossimilhança Restrita; G/Bayes = Gaussiano/Inferência Bayesiana; R/Bayes = Robusto (Normal Contaminada) / Inferência Bayesiana.

<sup>b</sup> VF = valor fenotípico sem contaminantes; VF1 = valor fenotípico com contaminação dada por  $\tau_1$ ; VF2 = valor fenotípico com contaminação dada por  $\tau_2$ .

<sup>c</sup> QMR = quadrado médio do resíduo da regressão dos valores genéticos verdadeiros sobre os valores genéticos preditos.

<sup>d</sup>  $\sigma_a^2$  = variância genética aditiva;  $\sigma_e^2$  = variância residual da população não-contaminante;  $\sigma_e^2/\tau$  = variância residual da população não-contaminante;  $h^2$  = herdabilidade da população contaminante;  $h^{2*}$  = herdabilidade da população não-contaminante;  $\tau$  e  $\phi$  = parâmetros da distribuição normal contaminada.

<sup>e</sup>  $\sigma_a$  = desvio-padrão genético aditivo;  $\sigma_e$  = desvio-padrão residual.

<sup>a</sup> G/REML = Gaussian/Restricted Maximum Likelihood; G/Bayes = Gaussian/Inference; R/Bayes = Robust (Contaminated Normal)/Bayesian Inference.

<sup>b</sup> VF = phenotypic value without contamination; VF1 = phenotypic value with contamination given for  $\tau_1$ ; VF2 = phenotypic value with contamination given for  $\tau_2$ .

<sup>c</sup> MSE = Mean square error of the regression of true genetic values on predicted genetic values.

<sup>d</sup>  $\sigma_a^2$  = additive genetic variance;  $\sigma_e^2$  = residual variance of the not contaminated population;  $\sigma_e^2/\tau$  = residual variance of the contaminated population;  $h^2$  = heritability of the not contaminated population;  $h^{2*}$  = heritability of the contaminated population;  $\tau$  and  $\phi$  = parameters of contaminated normal distribution.

<sup>e</sup>  $\sigma_a$  = additive genetic standard deviation;  $\sigma_e$  = residual standard deviation.



Tabela 6 - Resultados considerados na comparação dos modelos lineares mistos Gaussianos e Robustos para o 9<sup>o</sup> ( $\phi = 0,4$ ;  $\tau_1 = 0,171875$  e  $\tau_2 = 0,140625$ , respectivamente, para VF1 e VF2) e o 10<sup>o</sup> ( $\phi = 0,4$ ;  $\tau_1 = 0,15625$  e  $\tau_2 = 0,1484375$ , respectivamente, para VF1 e VF2) conjunto de parâmetros, com variâncias genética aditiva e residual médias iguais a 36,0 e 64,0, respectivamente  
 Table 6 - Results used for comparison of Gaussian and Robust mixed linear models, for the 9<sup>th</sup> ( $\phi = 0,4$ ;  $\tau_1 = 0,171875$  and  $\tau_2 = 0,140625$ , respectively for VF1 and VF2) and 10<sup>th</sup> sets of parameters ( $\phi = 0,4$ ;  $\tau_1 = 0,15625$  and  $\tau_2 = 0,1484375$ , respectively for VF1 and VF2), using averages for additive genetic and residual variances equal to 36,0 and 64,0, respectively

Modelo/Método <sup>a</sup> Model/Method <sup>a</sup>	Valor fenotípico <sup>b</sup> Phenotypic value <sup>b</sup>	R <sup>2</sup>		Estimativa <sup>d</sup> Estimate						Desvio-padrão <sup>e</sup> Standard deviation				
		Mínimo Minimum	Máximo Maximum	Médio Mean	QMR <sup>c</sup> MSE <sup>c</sup>	$\sigma^2_a$	$\sigma^2_e$	$\sigma^2_e/\tau$	h <sup>2</sup>	h <sup>2*</sup>	$\tau$	$\phi$	$\sigma_a$	$\sigma_e$
9 <sup>o</sup> conjunto 9 <sup>th</sup> group														
G/REML	VF	0,46	0,52	0,50	16,8667	38,76	—	—	0,37	—	—	—	—	—
	VF1	0,29	0,34	0,32	23,1577	43,93	—	—	0,19	—	—	—	—	—
	VF2	0,26	0,32	0,29	24,0165	44,83	—	—	0,17	—	—	—	—	—
G/Bayes	VF	0,46	0,52	0,50	16,9313	39,69	—	—	0,37	—	—	—	7,97	5,71
	VF1	0,29	0,34	0,32	23,2450	45,92	—	—	0,20	—	—	—	14,32	13,42
	VF2	0,26	0,32	0,29	24,1028	47,02	—	—	0,18	—	—	—	15,67	15,38
R/Bayes	VF	0,47	0,52	0,50	16,8506	40,43	476,07	0,38	0,38	0,08	0,1383	0,0039	7,71	5,83
	VF1	0,32	0,40	0,36	21,9448	48,69	358,82	0,43	0,43	0,12	0,1785	0,4367	15,99	16,94
	VF2	0,32	0,38	0,35	22,1852	50,04	437,92	0,45	0,45	0,10	0,1387	0,4291	14,19	15,77
10 <sup>o</sup> conjunto 10 <sup>th</sup> group														
G/REML	VF	0,47	0,54	0,51	17,5028	37,96	—	—	0,38	—	—	—	—	—
	VF1	0,28	0,35	0,32	24,3629	50,64	—	—	0,21	—	—	—	—	—
	VF2	0,27	0,35	0,31	24,5856	51,18	—	—	0,21	—	—	—	—	—
G/Bayes	VF	0,48	0,53	0,51	17,5589	39,18	—	—	0,38	—	—	—	7,58	5,34
	VF1	0,28	0,35	0,32	24,4553	53,13	—	—	0,22	—	—	—	15,77	13,65
	VF2	0,27	0,34	0,31	24,6800	53,75	—	—	0,22	—	—	—	16,24	14,17
R/Bayes	VF	0,48	0,54	0,51	17,5145	39,55	438,10	0,39	0,39	0,08	0,1391	0,0145	7,35	6,32
	VF1	0,34	0,39	0,37	22,5331	48,00	427,98	0,43	0,43	0,10	0,1494	0,3706	12,77	14,11
	VF2	0,34	0,39	0,37	22,5761	46,55	436,92	0,43	0,43	0,10	0,1430	0,3753	12,72	14,30

a G/REML = Gaussiano/Máxima Verossimilhança Restrita; G/Bayes = Gaussiano/Inferência Bayesiana; R/Bayes = Robusto (Normal Contaminada)/Inferência Bayesiana.  
 b VF = valor fenotípico sem contaminantes; VF1 = valor fenotípico com contaminação dada por  $\tau_1$ ; VF2 = valor fenotípico com contaminação dada por  $\tau_2$ .  
 c QMR = quadrado médio do resíduo da regressão dos valores genéticos verdadeiros sobre os valores genéticos preditos.  
 d  $\sigma^2_a$  = variância genética aditiva;  $\sigma^2_e$  = variância residual da população não-contaminante;  $\sigma^2_e/\tau$  = variância residual da população contaminante; h<sup>2</sup> = herdabilidade da população não-contaminante; h<sup>2\*</sup> = herdabilidade da população contaminante;  $\tau$  e  $\phi$  = parâmetros da distribuição normal contaminada.  
 e  $\sigma_a$  = desvio-padrão genético aditivo;  $\sigma_e$  = desvio-padrão residual.  
 a G/REML = Gaussian/Restricted Maximum Likelihood; G/Bayes = Gaussian/Bayesian Inference; R/Bayes = Robust (Contaminated Normal)/Bayesian Inference.  
 b VF = phenotypic value without contamination; VF1 = phenotypic value with contamination given for  $\tau_1$ ; VF2 = phenotypic value with contamination given for  $\tau_2$ .  
 c MSE = Mean square error of the regression of true genetic values on predicted genetic values.  
 d  $\sigma^2_a$  = additive genetic variance;  $\sigma^2_e$  = residual variance of the not contaminated population;  $\sigma^2_e/\tau$  = residual variance of the contaminated population; h<sup>2</sup> = heritability of the not contaminated population; h<sup>2\*</sup> = heritability of the contaminated population;  $\tau$  and  $\phi$  = parameters of contaminated normal distribution.  
 e  $\sigma_a$  = additive genetic standard deviation;  $\sigma_e$  = residual standard deviation.

obtidos pela distribuição *a posteriori*, tanto em modelo Gaussiano como Robusto, foram semelhantes entre si.

Para o modelo Robusto, os desvios aditivos para o VF1 e VF2 foram menores que aqueles do modelo Gaussiano. Para a maior parte dos conjuntos de parâmetros (Tabelas de 2 a 6), os desvios residuais, pelo modelo Robusto, para o VF1 foram maiores que para a característica com menor valor de  $\tau$  (VF2), indicando que maiores diferenças entre as subpopulações são mais fáceis de detectar.

As estimativas médias de  $h^2$  para VF em todos os conjuntos de dados foram próximas dos valores reais tanto para as obtidas nas análises pelos modelos Gaussianos, REML ou inferência Bayesiana quanto na análise com modelo Robusto, evidenciando que o modelo Robusto, mesmo quando utilizado em conjuntos de dados não contaminados, é adequado para a estimação dos parâmetros. As estimativas de  $h^2$  obtidas por REML e pela distribuição *a posteriori* em modelo Gaussiano para VF1 e VF2 devem ser consideradas como a média das herdabilidades das duas subpopulações ponderadas por  $\phi$ . Utilizando-se o modelo Robusto, é possível obter estimativas de herdabilidade separadas para cada subpopulação, embora a interpretação desses valores quando se considera o conjunto das subpopulações seja difícil. As estimativas obtidas foram próximas dos valores reais para cada uma das subpopulações. Nas análises com modelos Gaussianos não é possível separar as populações e, assim, encontra-se um valor médio para as duas subpopulações. Comparando as estimativas de  $h^2$  obtidas por REML e Inferência Bayesiana, sob modelo Gaussiano e por Inferência Bayesiana, sob modelo Robusto (Tabelas de 2 a 6), verifica-se, como para as estimativas das variâncias, que as estimativas de  $h^2$  são semelhantes entre as características na estimação por REML e Inferência Bayesiana, em modelo Gaussiano e diferem da maioria dos conjuntos de parâmetros obtidos pelo modelo Robusto. Resultados semelhantes também foram encontrados por Strandén & Gianola (1999) ao compararem o modelo Gaussiano, via Amostrador de Gibbs, ao modelo Robusto utilizando a distribuição *t*.

Os valores de  $R^2$  mínimos, máximos e médios resultantes da regressão dos valores genéticos verdadeiros sobre os preditos foram semelhantes quando os componentes de variância foram estimados por REML e Inferência Bayesiana, ambos sob o modelo Gaussiano, para os três valores fenotípicos avaliados (VF, VF1 e VF2), e apresentaram valores decrescentes nessa mesma ordem (Tabelas 2 a 6), o que evidencia a não-adequação do modelo Gaussiano para as características que possuem indivíduos contaminantes (presença de *outliers*). Os  $R^2$  diminuem menos proporcio-

nalmente de VF para a VF1 que de VF1 para VF2. Portanto, a pior qualidade de ajuste do modelo Gaussiano se acentua ainda mais à medida que aumenta o grau de contaminantes, o que seria esperado.

Uma vez que no processo de simulação foram confeccionados 50 conjuntos de dados com estrutura de população diferente, escolheu-se apenas um desses conjuntos de dados: a 5ª repetição do 2º conjunto de parâmetros, no qual ocorreu uma sensível melhora no ajuste promovido pelo modelo Robusto. Para a característica com VF1 e VF2, que apresentam indivíduos contaminantes, houve melhora de 11 e 16%, respectivamente, para a estimação pelo modelo Robusto, atestando a melhor qualidade de ajuste desse modelo e propiciando valores genéticos mais acurados em situações de dados com heterogeneidade de variância residual.

Para o modelo Robusto, os  $R^2$  foram superiores aos encontrados para o modelo Gaussiano para a VF1 e VF2 em todos conjuntos de parâmetros (Tabelas 2 a 6), o que comprova a melhor qualidade de ajuste do modelo Robusto em relação ao Gaussiano. Para a VF, a qual não possui contaminantes, o  $R^2$  manteve-se praticamente o mesmo, indicando que o modelo Robusto pode ser usado para análise de dados que não apresentem contaminação. A melhora do ajuste do modelo Robusto foi maior para a VF2, a qual apresentava menor valor de  $\tau$ . Parece que, quanto maior a diferença entre as variâncias das subpopulações, mais eficiente se torna o ajuste do modelo Robusto em relação ao modelo Gaussiano. O melhor ajuste ocorreu para o 1º conjunto de parâmetros (Tabela 2), de 0,26 para 0,38, o que representa melhora de 46%, diminuindo o efeito da heterogeneidade da variância residual. Em estudo com dados de produção de leite, comparando modelo Gaussiano com Robusto, no qual utilizaram a distribuição de *t*, Strandén & Gianola (1999) também observaram redução da estimativa da variância residual obtida pelo modelo Robusto. Outro indicativo estatístico de qualidade de ajuste, no caso das regressões, é o quadrado médio do resíduo (Tabelas 2 a 6), que, embora seja proporcional ao  $R^2$ , comprova quantitativamente a variância do erro de predição dos valores genéticos. Portanto, a predição dos valores genéticos dos animais para as características contendo contaminantes utilizando uma distribuição normal contaminada como alternativa robusta produziu menores erros de predição.

## Conclusões

As estimativas das variâncias genética aditiva e residual obtidas utilizando o modelo Gaussiano por REML e

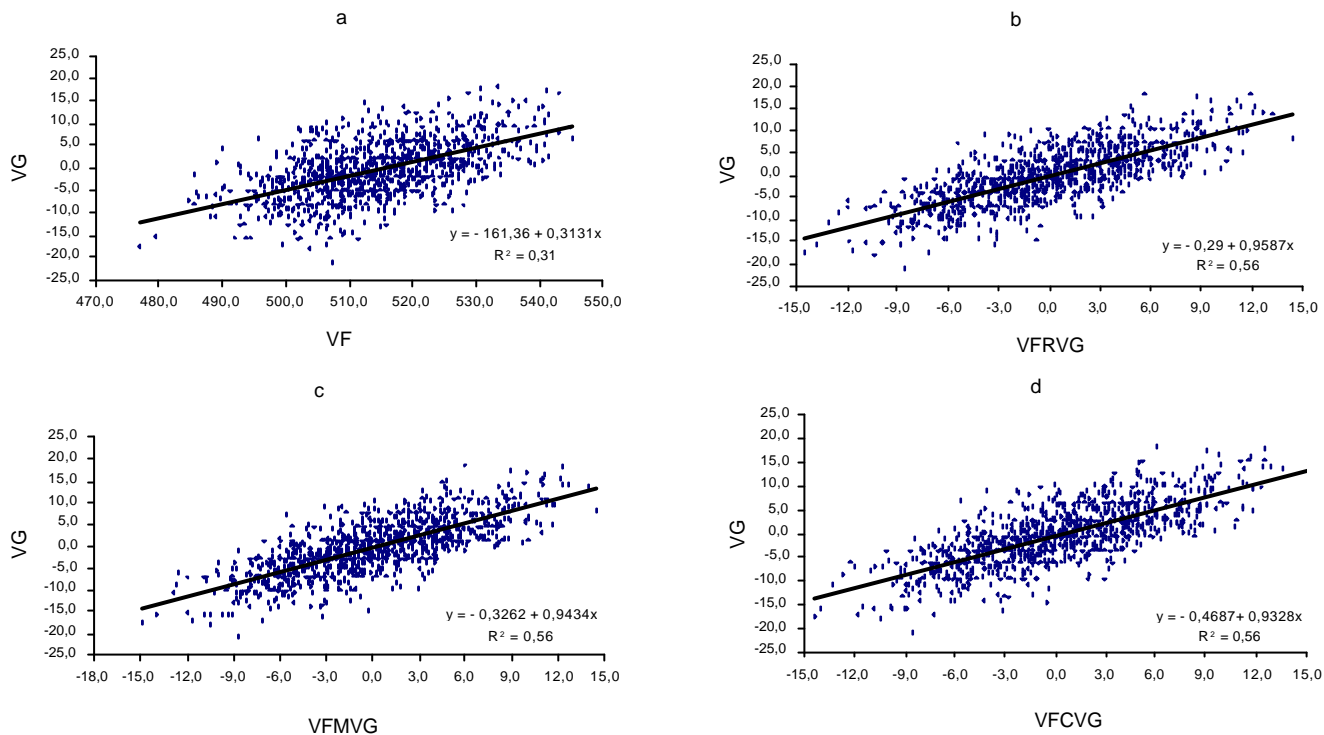


Figura 1 - Regressão dos valores genéticos verdadeiros (VG) sobre o valor fenotípico VF (sem contaminação), dada pelo 2º conjunto de parâmetros ( $\phi = 0,4$ ) na 5ª repetição para: a) valor fenotípico (VF); e para valores genéticos preditos: b) Máxima Verossimilhança Restrita (VFRVG), c) Inferência Bayesiana (VFMVG) e d) Robusto (VFCVG).

Figure 1 - Regression of true genetic values (VG) on phenotypic value VF (without contamination), for the 2<sup>nd</sup> set of parameters ( $\phi = 0.4$ ) in the 5<sup>th</sup> repetition for: a) phenotypic value (VF); and for predicted genetic values: b) Restricted Maximum Likelihood (VFRVG), c) Bayesian Inference (VFMVG) and d) Robust (VFCVG).

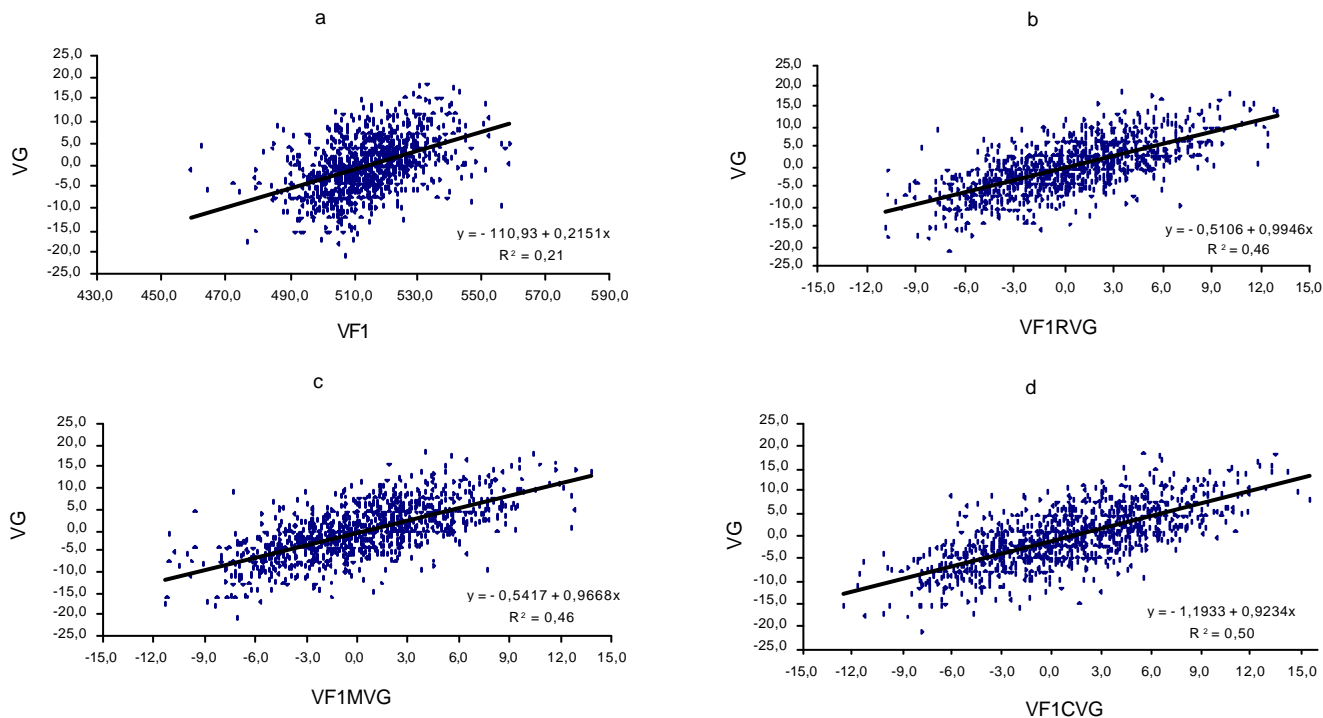


Figura 2 - Regressão dos valores genéticos verdadeiros (VG) sobre o valor fenotípico VF1 (com contaminação), dada pelo 2º conjunto de parâmetros ( $\phi = 0,4$  e  $\tau_1 = 0,1875$ ) na 5ª repetição para: a) valor fenotípico (VF1); e para valores genéticos preditos: b) Máxima Verossimilhança Restrita (VF1RVG), c) Inferência Bayesiana (VF1MVG) e d) Robusto (VF1CVG).

Figure 2 - Regression of true genetic values (VG) on phenotypic value VF1 (with contamination), for the 2<sup>nd</sup> set of parameters ( $\phi = 0.4$  and  $\tau_1 = 0.1875$ ) in the 5<sup>th</sup> repetition for: a) phenotypic value (VF1); and for predicted genetic values: b) Restricted Maximum Likelihood (VF1RVG), c) Bayesian Inference (VF1MVG) and d) Robust (VF1CVG).

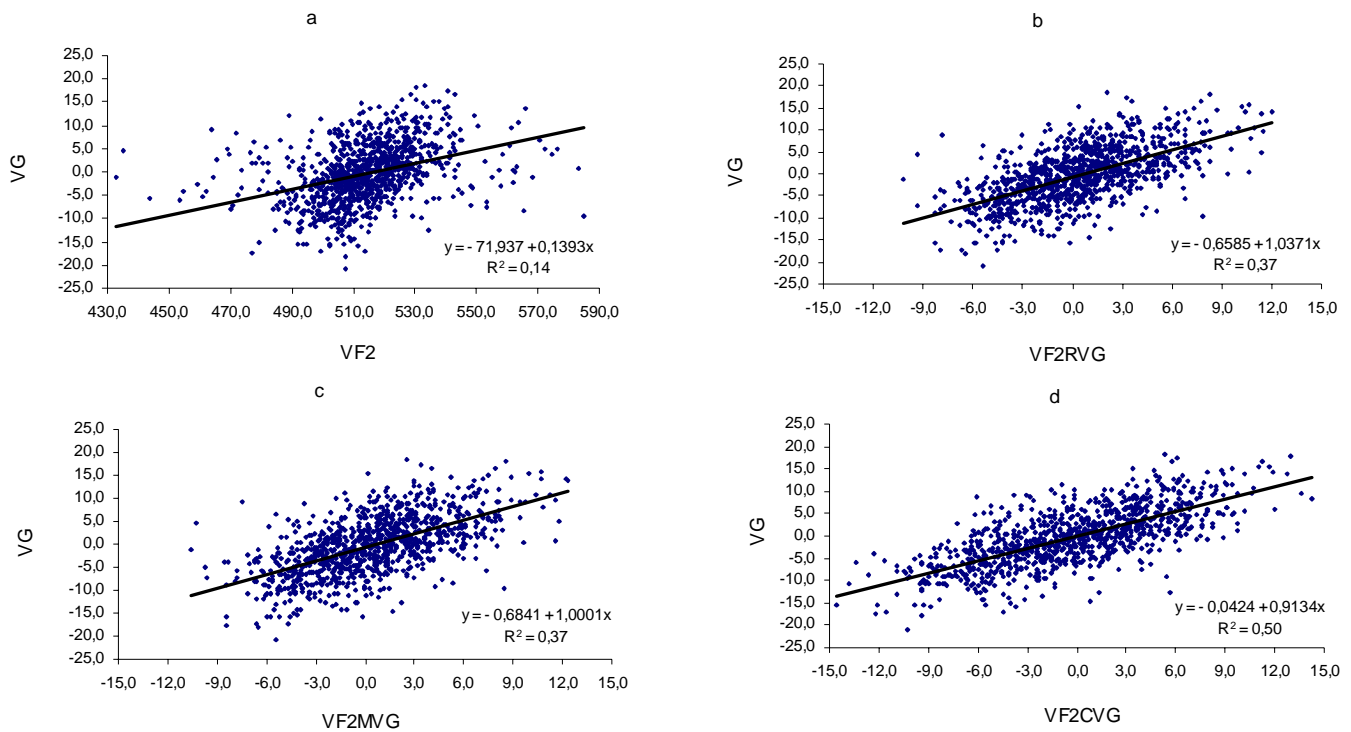


Figura 3 - Regressão dos valores genéticos verdadeiros (VG) sobre o valor fenotípico VF2 (com contaminação), dada pelo 2º conjunto de parâmetros ( $\phi = 0,4$  e  $\tau_2 = 0,125$ ) na 5ª repetição para: a) valor fenotípico (VF2); e para valores genéticos preditos: b) Máxima Verossimilhança Restrita (VF2RVG), c) Inferência Bayesiana (VF2MVG) e d) Robusto (VF1CVG).

Figure 3 - Regression of true genetic values (VG) on phenotypic VF2 (with contamination), for the 2<sup>nd</sup> set of parameters ( $\phi = 0.4$  and  $\tau_2 = 0.125$ ) in the 5<sup>th</sup> repetition for: a) phenotypic value (VF2); and for predicted genetic values: b) Restricted Maximum Likelihood (VF2RVG), c) Bayesian Inference (VF2MVG) and d) Robust (VF2CVG).

Inferência Bayesiana foram similares. As estimativas da variância residual para as características contendo contaminantes foram bastante aumentadas, demonstrando que o modelo Gaussiano foi inadequado para estimar a variância residual de dados que apresentam contaminação. Para a característica em que os dados seguem distribuição normal, os modelos Gaussiano e Robusto produzem resultados semelhantes. A utilização do modelo Robusto em situações em que existe heterogeneidade de variância residual permite identificar esta contaminação e obter melhores resultados na predição do valor genético.

### 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: U.S. Department of Agriculture/Agriculture Research Service, 1995. 120p.
- CARNEIRO JR., J.M.; ASSIS, G.M.L.; EUCLYDES, R.F. et al. Influência da informação *a priori* na avaliação genética animal utilizando dados simulados. **Revista Brasileira de Zootecnia**, v.34, n.6, p.1905-1913, 2005.
- GELMAN, A.; CARLIN, J.B.; STERN, H.S. et al. **Bayesian data analysis**. Boca Raton: Chapman and Hall/CRC Press, 1995. 523p.
- ROGERS, W.H.; TUKEY, J.W. Understanding some long-tailed distributions. **Statistica Neerlandica**, v.26, p.211-226, 1972.

- ROSA, G.J.M.; PADOVANI, C.R.; GIANOLA, D. Robust linear mixed models with normal/independent distributions and Bayesian MCMC implementation. **Biometrical Journal**, v.45, n.5, p.573-590, 2003.
- STATISTICAL ANALYSIS SYSTEM – SAS - **User's guide**: statistics. Cary: 1999. (CD-ROM).
- STRANDÉN, I.; GIANOLA, D. Mixed effects linear models with *t*-distributions for quantitative genetic analysis: a Bayesian approach. **Genetic Selection Evolution**, v.31, p.25-42, 1999.
- Van TASSEL, C.P.; CASELLA, G.; POLLAK, E.J. Effects of selection on estimates of variance components using Gibbs sampling and restricted maximum likelihood. **Journal of Dairy Science**, v.78, n.3, p.678-692, 1995.
- Van TASSELL, C.P.; Van VLECK, L.D. **A manual for use of MTGSAM**: a set of Fortran programs to apply Gibbs sampling to animal models for variance components estimation. Lincoln: U.S. Department of Agriculture, Agricultural Research Service, 1995. 85p.
- Van TASSEL, C.P.; Van VLECK, L.D. Multiple-trait sampler for animal models: flexible programs for Bayesian and likelihood-based (co)variance component inference. **Journal of Animal Science**, v.74, p.2586-2597, 1996.
- WRIGHT, D.R.; STERN, H.S.; BERGER, P.J. Comparing traditional and Bayesian Analyses of selection experiments in animal breeding. **Journal of Agricultural, Biological, and Environmental Statistics**, v.5, n.2, p.240-256, 2000.