



## Inferência robusta e heterocedástica em componentes de variância, parâmetros e valores genéticos multirraciais<sup>1</sup>

Mauricio Morgado de Oliveira<sup>2</sup>, Fernando Flores Cardoso<sup>2,3</sup>, José Carlos da Silveira Osório<sup>2</sup>

<sup>1</sup> Pesquisa financiada pelo CNPq.

<sup>2</sup> Departamento de Zootecnia/FAEM, UFPEL, Pelotas, RS. Bolsista do CNPq.

<sup>3</sup> Embrapa Pecuária Sul, Bagé, RS.

**RESUMO** - Os objetivos neste estudo foram avaliar um modelo animal multirracial (MAMR) heterocedástico para estimar a heterogeneidade de variância residual devida à raça, à heterozigosidade, ao sexo e ao grupo de contemporâneos (GC) e a robustez a dados extremos. Foram avaliadas diferenças nas estimativas de componentes de variância, herdabilidades e valores genéticos. Foram utilizados 4.016 registros de ganho de peso pós-desmame (GPD) do rebanho de formação da raça Brangus-Ibagé da Embrapa Pecuária Sul. Os dados foram analisados por um MAMR com quatro especificações diferentes da distribuição dos resíduos, definidas pela natureza da variância residual homocedástica (HO) ou heterocedástica (HE) e pela distribuição marginal dos resíduos Gaussiana (G) ou *t* de Student (T). O melhor ajuste aos dados foi pelo MAMR-T-HE e foi observada heterocedasticidade devida ao sexo e à heterozigosidade dos animais e entre grupos de contemporâneos. Houve diferenças substanciais nas variâncias genéticas de grupos raciais estimadas utilizando-se diferentes modelos e nos heterocedásticos, MAMR-G-HE e MAMR-T-HE. As maiores diferenças nas estimativas de herdabilidade ocorreram entre grupos raciais diferentes. Também foram observados reordenamentos na classificação de animais superiores quanto ao mérito genético quando variâncias residuais heterogêneas e robustez a dados extremos foram consideradas. Avaliações genéticas multirraciais utilizando o modelo *t* de Student para descrever resíduos e assumindo variâncias residuais heterogêneas são mais apropriadas para realizar interferência no mérito genético de animais produtos de cruzamento entre as raças Angus e Nelore.

Palavras-chave: Angus, avaliação genética, bovinos de corte, cruzamento, inferência bayesiana, Nelore

## Robust and heteroskedastic inference in multibreed variance components, genetic parameters and breeding values

**ABSTRACT** - The objectives of this study were to evaluate a multibreed heteroskedastic animal model (MAMR) to estimate residual variance heterogeneity due to breed, heterozygosity, sex and contemporary groups (CG) and robustness to outliers. Differences on variance components, heritability and genetic values estimates were evaluated. It was used 4,016 records of post weaning weight gain (PWG) from Brangus foundation herd of Embrapa Pecuária Sul. Data were analyzed by MAMR with four different residual distribution specifications, which were defined by the nature of residual variance, homoskedastic (HO) or Heteroskedastic (HE) and by the Gaussian (G) or Student *t* marginal residual distribution. The best data fit was achieved by using MAMR-T-HE and it was observed heteroskedasticity due to sex and animal heterozygosity and among contemporary groups. There were substantial differences in genetic variances of breed groups estimated by using different models and in the heteroskedastic, MAMR-G-HE and MAMR-T-HE. The greatest differences on heritability estimates occurred among different breed groups. It was also noted a reordering in the ranking of superior animals regarded to genetic merit when heterogenous residual variances and robustness to outliers were considered. Multibreed genetic evaluations by using the Student *t* distribution model to describe residuals and assuming heterogenous residual variances are more appropriate to perform inference on genetic merit of animals from crossbreeding progenies of Angus and Nelore purebred animals.

Key Words: Angus, Bayesian inference, beef cattle, crossbreeding, genetic evaluation, Nelore

### Introdução

Os ambientes de criação onde o desempenho de bovinos de corte é registrado para as avaliações genéticas têm

geralmente ampla diversidade climática e de sistemas de produção. A qualidade dos dados pode ser comprometida por ocorrência de erros de registro, de tratamento preferencial e de potencial efeito de lesões ou doenças (Cardoso et al., 2005).

A utilização de modelos hierárquicos bayesianos é uma alternativa para as avaliações genéticas de populações com registros de dados caracterizados por diferentes tipos de heterogeneidade em múltiplos níveis (Sorensen & Gianola, 2002). No melhoramento animal esses modelos foram utilizados para ajustar variâncias residuais heterogêneas (Foulley et al., 1992; Gianola et al., 1992; Sancristobal et al., 1993) e avaliar robustez a dados extremos em virtude de tratamento preferencial (Stranden & Gianola, 1998; 1999). Heterocedasticidade e robustez geralmente são avaliadas isoladamente, entretanto, não há dificuldade em considerá-las concomitantemente (Kizilkaya & Tempelman, 2003; Sorensen & Waagepetersen, 2003; Kizilkaya & Tempelman, 2005), o que é especialmente relevante uma vez que esses fatores podem estar correlacionados (Cardoso et al., 2005).

Na presença de heterocedasticidade residual, os valores genéticos de animais criados em ambientes com maior variância seriam superestimados (Carvalho et al., 2002), visto que os dados ajustados para os efeitos de ambiente têm menor contribuição genética do que a ponderação que lhes estaria sendo aplicada (Martins, 2002).

Os objetivos neste estudo foram aplicar um modelo animal multirracial considerando heterogeneidade de variância residual devido à raça, heterozigose, sexo e grupo de contemporâneos, e avaliar possíveis diferenças nas estimativas de componentes de variância, herdabilidade e valores genéticos de bovinos produtos do cruzamento entre as raças Angus e Nelore quando variâncias residuais heterogêneas são consideradas.

## Material e Métodos

Registros de ganho de peso pós-desmame (GPD) padronizados para um período de teste de 345 dias do rebanho de formação da raça Brangus-Ibagé (3/8 Nelore - 5/8 Angus) da EMBRAPA Pecuária Sul foram analisados. A formação desse rebanho foi iniciada em 1940, mas dados de peso corporal foram registrados somente a partir de 1956, quando nasceu a primeira geração de animais 3/8 Nelore.

Foram controlados 6.499 nascimentos e realizados 110.250 registros de peso corporal ao nascer, ao desmame e mensalmente até os dois anos de idade, entre 1956 e 1979. Foram excluídos animais sem registro de peso ao desmame e, para obter o GPD dentre as múltiplas pesagens disponíveis, foram formados grupos de contemporâneos (GC) considerando ano e estação de nascimento, sexo e data da pesagem. Foram mantidas aquelas avaliações em que a idade média do GC estava mais próxima de 550 dias, e GC em que a idade média

na data da pesagem era menor que 490 dias ou maior que 610 dias ou GC com menos de quatro animais foram eliminados. Foram ainda eliminados registro de GPD com 3,5 desvios-padrão acima da média do GC ao qual o animal pertencia, restando um subconjunto de 4.060 registros.

A conectabilidade entre grupos contemporâneos foi avaliada em função do número total de laços genéticos (mínimo 10), utilizando o programa AMC (Roso & Schenkel, 2006). Foram eliminados registros de 44 animais pertencentes a nove grupos contemporâneos. Portanto, foram utilizados para as análises 4.016 registros, havendo 4.819 animais no arquivo de pedigree. Os animais avaliados apresentaram composições raciais que variavam de Angus puro a 1/4 Angus × 3/4 Nelore. Todos os animais foram criados em pastajeo sob condições extensivas.

Os dados foram analisados utilizando um modelo animal multirracial (MAMR):

$$y_{ijkl} = \mu + gc_i + sexo_j + \beta_{IDV} (idv_k) + \beta_{IDV^2} (idv_k^2) + \beta_A (f_l) + \beta_D (\delta_l) + \beta_{AA} (2[1-f_l]f_l) + u_l + e_{ijl}, \quad [1]$$

em que:  $y_{ijkl}$  = registro de GPD do  $l$ -ésimo animal, pertencente ao  $i$ -ésimo grupo de contemporâneos,  $j$ -ésimo sexo e filho da  $k$ -ésima vaca;  $\mu$  = média geral da população;  $gc_i$  = efeito do  $i$ -ésimo GC ( $i = 1, 2, \dots, 268$ ), sendo  $gc_i \sim N(0, \sigma_{gc}^2)$ ;  $sexo_j$  = efeito do  $j$ -ésimo sexo ( $j=1, j=2$  e  $j=3$ , para macho, macho castrado e fêmea, respectivamente);  $\beta_{IDV}$  e  $\beta_{IDV^2}$  = respectivamente, coeficientes de regressão linear e quadrático da idade da vaca, mãe do animal  $l$ , em meses ( $23 \leq IDV \leq 192$ );  $\beta_A$  = coeficiente do efeito aditivo ( $A$ ) da raça Angus,  $\beta_D$  do efeito de dominância Nelore-Angus ( $D$ ) e  $\beta_{AA}$  do efeito epistático Nelore-Angus ( $AA$ );  $f_l$  = proporção esperada de alelos oriundos da raça Angus no animal  $l$ ;  $\delta_l$  = coeficiente de heterozigosidade, que representa a probabilidade que um dos alelos de um locus aleatoriamente escolhido seja derivado da raça Nelore e o outro alelo seja derivado da raça Angus;  $2[1-f_l]f_l$  = coeficiente epistático para o animal  $l$  baseado na definição de perdas por recombinação de Kinghorn (1980);  $u_l$  = efeito genético aditivo do animal  $l$ , assumindo que  $u = \{u_l\}_{l=1}^{4819} \sim N(0, G)$ , em que  $G$  é a matriz de (co)variância genética multirracial (Lo et al., 1993), que é uma função do parentesco aditivo entre os animais e das variâncias genéticas aditivas específicas por raça, para Angus  $\sigma_{g(1)}^2$  e Nelore  $\sigma_{g(2)}^2$ ;  $e_{ijl} \sim N(0, \sigma_{e_{ijl}}^2)$  representa os resíduos, para a qual a pressuposição sobre sua variância ( $\sigma_{e_{ijl}}^2$ ) e sua distribuição caracteriza os distintos modelos avaliados. Assumiu-se que a variância de segregação entre as duas raças era igual a zero,  $\sigma_{s(12)}^2 = 0$ , nesta população (Oliveira et al., 2010).

Quatro especificações diferentes da distribuição e variância dos resíduos foram consideradas em um arranjo fatorial  $2 \times 2$ , no qual o primeiro fator foi definido pela variâncias residual, homocedástica ou heterocedástica, e o segundo fator pela distribuição marginal dos resíduos, Gaussiana (G) ou  $t$  de Student (T). Os quatro modelos resultantes são apresentados em detalhes a seguir.

**MAMR-T-HE:** os erros foram considerados heterocedásticos e distribuídos de acordo com uma distribuição  $t$  de Student. Assumiu-se que a variância residual específica para o registro do animal  $l$  era determinada por uma função multiplicativa da variância residual global ( $\sigma_e^2$ ) e fatores escalares multiplicativos específicos para GC ( $\tau_{GC_i}$ ), sexo ( $\tau_{S_j}$ ), proporção da raça Angus ( $\gamma_A$ ) e heterozigiosidade ( $\gamma_D$ ):  $\sigma_{e_{ijl}}^2 = \sigma_e^2 \tau_{GC_i} \tau_{S_j} (\gamma_A)^{f_i} (\gamma_D)^{\delta_i} w_l^{-1}$ . A contribuição do grupo de contemporâneos sobre a variância residual ( $\tau_{GC_i}$ ) foi considerada uma variável aleatória com distribuição *a priori*  $\tau_{GC_i} \sim \text{Gama}^{-1}(\eta, \eta-1)$ , sendo  $\eta > 1$ , em que  $\eta$  é o parâmetro desconhecido que define a heterocedasticidade da variância residual em função do grupo contemporâneo a que pertence o animal  $l$ , uma vez que  $E(\tau_{GC_i}) = 1$ , e  $\text{Var}(\tau_{GC_i}) = \frac{1}{\eta-2}$  a influência do efeito aleatório de GC na heterocedasticidade residual desaparece quando  $\eta \rightarrow \infty$  (Cardoso et al., 2005; Kizilkaya & Tempelman, 2005). A variável  $w_l$  representa a ponderação específica para o registro do animal  $l$  e sua distribuição *a priori* foi definida como  $w_l \sim \text{Gama}\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$ , sendo  $w_l > 0$  e  $\nu > 0$ ;  $\nu$  representa um hiperparâmetro de robustez desconhecido. Com essa especificação, a distribuição marginal dos resíduos *a priori* é  $t$  de Student com  $\nu$  graus de liberdade (Lange & Sinsheimer, 1993; Rosa et al., 2003).

**MAMR-G-HE:** neste modelo o erro também foi considerado heterocedástico, entretanto, a distribuição marginal dos resíduos foi considerada Gaussiana. Assumiu-se  $w_l = 1$  para todo  $l$ , isto é, todos os registros tiveram a mesma ponderação. Por exemplo, a variância residual para um animal macho inteiro  $F_1$  do grupo de contemporâneos 2 (isto é,  $i=2, j=1, f_i=0,5, \delta_i=0,5$  e  $w_l=1$ ) foi estimada como:  $\sigma_{e_{ijl}}^2 = \sigma_e^2 \tau_{GC_2} \tau_{S_1} (\gamma_A)^{0,5} (\gamma_D)^{0,5} (1)^{-1}$ .

**MAMR-T-HO:** neste modelo os erros foram considerados homocedásticos e distribuídos de acordo com uma distribuição  $t$  de Student. Neste modelo, impõe-se que  $\tau_{GC_i} = \tau_{S_j} = (\gamma_A)^{f_i} = (\gamma_D)^{\delta_i} = 1$  e, portanto, a variância residual para o registro do animal  $l$  é homocedástica e definida por  $\sigma_{e_{ijl}}^2 = \sigma_e^2 w_l^{-1}$ . Assim como no primeiro modelo, a variável de

ponderação específica para o registro do animal  $l$  foi considerada aleatória e distribuída de acordo com  $w_l \sim \text{Gama}\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$ .

**MAMR-G-HO:** neste modelo assumiram-se erros Gaussianos homocedásticos, sendo  $\tau_{GC_i} = \tau_{S_j} = (\gamma_A)^{f_i} = (\gamma_D)^{\delta_i} = 1$  e  $w_l = 1$ . Assim,  $\sigma_{e_{ijl}}^2 = \sigma_e^2$  é a variância residual homocedástica para todo  $l$ .

Distribuições *a priori* uniformes finitas foram adotadas para os efeitos ambientais  $\mu$ ,  $sexo_j$  e todos os coeficientes  $\beta^*$ 's, bem como para os efeitos  $\tau_{S_j}$ ,  $\gamma_A$  e  $\gamma_D$  sobre a variância residual. Para os efeitos aleatórios de  $gc_j$  e  $u$  foram presumidas distribuições *a priori* normais com médias e variâncias conforme definidas acima na descrição do modelo. Especificações conjugadas foram adotadas como distribuições *a priori* para os componentes de variância. As seguintes distribuições quadrado invertidas escalonadas foram empregadas:  $\sigma_{gc}^2 | \nu_{gc}, s_{gc}^2 \sim s_{gc}^2 \chi_{gc}^{-2}$  e  $\sigma_{g(r)}^2 | \nu_{g(r)}, s_{g(r)}^2 \sim s_{g(r)}^2 \chi_{g(r)}^{-2}$ , para a variância devida ao efeito de grupo contemporâneo e variância aditiva, respectivamente, em que  $r=1$  para Angus e  $r=2$  para Nelore. Os hiperparâmetros representados por  $\nu_{gc}$ , e  $\nu_{g(r)}$  indicam o grau de confiança nos valores *a priori* dos componentes de variâncias  $\sigma_{gc}^2$  e  $\sigma_{g(r)}^2$ , dados respectivamente pelos hiperparâmetros  $s_{gc}^2$  e  $s_{g(r)}^2$ . Os valores adotados para os hiperparâmetros foram  $\nu_{g(1)} = \nu_{g(2)} = 8$ ;  $\nu_{gc} = 1$ ;  $s_{g(1)}^2 = s_{g(2)}^2 = 150 \text{ kg}^2$  e  $s_{gc}^2 = 1000 \text{ kg}^2$ . Como distribuição *a priori* da variância residual assumiu-se  $\sigma_e^2 | \nu_e, s_e^2 \sim s_e^2 \chi_e^{-2}$ , com  $\nu_e = 1$  e  $s_e^2 = 350 \text{ kg}^2$  e foi especificada uma distribuição gama *a priori* para o parâmetro de heterocedasticidade  $\eta$  com hiperparâmetros  $\alpha_\eta = 0,03$  e  $\beta_\eta = 0,01$ , resultando numa média *a priori* de  $E_\eta = 3$ , com variância  $\text{var}(\eta) = 300$ . Similarmente, foram especificadas distribuições gama *a priori* para o parâmetro de robustez  $\nu$ , com hiperparâmetros  $\alpha_\nu = 0,04$  e  $\beta_\nu = 0,01$  para os modelos em que o erro era distribuído como  $t$  de Student (MAMR-T-HO e MAMR-T-HE), tal que a média *a priori* foi  $E_{(\nu)} = 4$ .

A variância genética de um grupo racial  $g$  ( $\sigma_g^2$ ) foi obtida pela média das variâncias raciais ponderadas pela composição racial do grupo,  $\sigma_g^2 = f_1^g \sigma_{g(1)}^2 + f_2^g \sigma_{g(2)}^2$ , em que  $f_r^i$  é a fração da raça  $r$  observada no grupo racial  $g$ .

A variância residual marginal específica para um grupo racial  $g$  ( $\sigma_{e(g)}^2$ ) foi estimada de acordo com as pressuposições de cada modelo, seguindo Cardoso et al. (2005). Para MAMR-G-HO tem-se  $\sigma_{e(g)}^2 = \sigma_e^2$ ; para MAMR-T-HO,  $\sigma_{e(g)}^2 = \frac{\nu}{\nu-2} \sigma_e^2$ ; para MAMR-G-HE,  $\sigma_{e(g)}^2 = \sigma_e^2 \tau_{S_1}^{0,146} \tau_{S_2}^{0,241} (\gamma_A)^{f_i} (\gamma_D)^{\delta_i}$ , sendo

$E(\tau_{GCi}) = 1$  em todo grupo contemporâneo  $i$  e que 0,146 e 0,241 são, respectivamente, a proporção amostral de animais do sexo 1 (macho inteiro) e sexo 2 (macho castrado); para MAMR-T-HE  $\sigma_{e(g)}^2 = \frac{v}{v-2} \sigma_e^2 \tau_{s_1}^{0,146} \tau_{s_2}^{0,241} (\gamma_A)^{f_i} (\gamma_B)^{\delta_i}$ . A herdabilidade para o grupo  $g$  pode ser obtida por:

$$h_g^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2}$$

Os valores da variância residual de referência  $\sigma_e^2$  são afetados pela parametrização do modelo estrutural e não são diretamente comparáveis entre modelos com distintas distribuições marginais dos resíduos (Cardoso et al., 2005), como os empregados no presente estudo. Portanto, calculou-se a variância residual marginal global  $\sigma_E^2$ , que representa uma estimativa do valor médio para dispersão residual adotada para todas as observações. Para os dois modelos de erro homocedástico (MAMR-G-HO e MAMR-T-HO),  $\sigma_E^2$  é idêntica para todos os grupos raciais e obtida por  $\sigma_E^2 = \sigma_e^2$  e  $\sigma_E^2 = \frac{v}{v-2} \sigma_e^2$ , respectivamente. Para os modelos heterocedásticos (MAMR-G-HE e MAMR-T-HE),  $\sigma_E^2$  foi ponderada pela proporção observada de sexo do terneiro (0,146 e 0,241, respectivamente, para macho inteiro e macho castrado), pela proporção de média de genes da raça Angus esperada na população ( $f^- = 0,631$ ) e pelo coeficiente médio de heterozigiosidade ( $\bar{\delta} = 0,473$ ), sendo obtida por  $\sigma_E^2 = \sigma_e^2 \tau_{s_1}^{0,146} \tau_{s_2}^{0,241} (\gamma_A)^{0,6131} (\gamma_B)^{0,473}$  e  $\sigma_E^2 = \frac{v}{v-2} \sigma_e^2 \tau_{s_1}^{0,146} \tau_{s_2}^{0,241} (\gamma_A)^{0,6131} (\gamma_B)^{0,473}$ , respectivamente, para MAMR-G-HE e MAMR-T-HE.

A inferência foi baseada no método Monte Carlo via Cadeias de Markov (MCMC) de 110.000 ciclos, após 10.000 ciclos do período de descarte e as amostras foram retiradas a cada dez ciclos. Utilizou-se o programa INTERGEN (Cardoso, 2008). Médias, modas, percentis (0,025 e 0,975) e os desvios-padrão das densidades marginais *a posteriori* dos parâmetros foram calculados. Estimativas de correlações de Spearman entre as médias *a posteriori* dos efeitos genéticos aditivos de ganho de peso pós-desmame, obtidas utilizando todos os animais e somente para touros (reprodutores) em cada classe ou somente aqueles classificados nos 10% superiores de cada classe.

Para verificar o ajuste dos modelos foram utilizados o critério de informação de desvio (DIC) (Spiegelhalter et al., 2002) e ordenadas preditivas condicionais (Gelfand, 1996).

O valor de critério de informação de desvio é composto por uma medida de ajuste global, a média *a posteriori* do desvio e uma penalização para a complexidade do modelo. O desvio do modelo  $M_r$ , segundo Spiegelhalter et al. (2002), pode ser estimado por  $\bar{D}_r = \frac{1}{C} \sum_{c=1}^C -2 \log p(\mathbf{y} | \theta^{(c)}, M_r)$ , em

que  $\theta^{(c)}$  inclui o valor amostrado para todos os parâmetros do modelo  $M_r$  no ciclo  $c$  de  $C$  ciclos totais. A complexidade do modelo  $M_r$  é determinada pelo número efetivo de parâmetros  $p_{D(r)} = \bar{D}_r - D_r(\bar{\theta})$ , em que  $D_r(\bar{\theta}) = -2 \log p(\mathbf{y} | \bar{\theta}, M_r)$ , sendo  $\bar{\theta}$  a média *a posteriori* de  $\theta$ , isto é,  $p_{D(r)}$  representa a diferença entre a média *a posteriori* do desvio e o desvio calculado com a média *a posteriori* dos parâmetros do modelo  $M_r$ . O critério de informação de desvio para o modelo  $M_r$  é determinado por:

$$DIC_r = \bar{D}_r + p_{D(r)}$$

Menores valores do critério de informação de desvio são indicativos de melhor ajuste do modelo.

Para observação  $y_{ijkl}$ , a ordenada preditiva condicional é  $p(y_{ijkl} | \mathbf{y}_{(-ijkl)}, M_r)$ , destinada ser a densidade de validação cruzada, que sugere quais valores de  $y_{ijkl}$  são prováveis quando o modelo  $M_r$  é ajustado a todas as outras observações  $y_{(-ijkl)}$ , exceto  $y_{ijkl}$ . Para um dado modelo  $M_r$ , as ordenadas preditivas condicionais foram obtidos pela média harmônica de  $C$  ciclos de MCMC, pela seguinte aproximação (Gelfand, 1996):

$$\hat{p}(y_{ijkl} | \mathbf{y}_{(-ijkl)}, M_r) \approx \frac{1}{\frac{1}{C} \sum_{c=1}^C p^{-1}(y_{ijkl} | \theta^{(c)}, M_r)}$$

Finalmente, um estimador do desvio do modelo  $r$  baseado nas ordenadas preditivas condicionais de todas as observações foi obtido por  $-2 \sum \log \hat{p}(y_{ijkl} | \mathbf{y}_{(-ijkl)}, M_r)$ , isto é, multiplicando-se por -2 a soma dos logaritmos de todos os registros. De mesma forma que para DIC, menores valores do desvio baseado em ordenadas preditivas condicionais indicam melhor ajuste.

## Resultados e Discussão

Os critérios adotados para comparar o ajuste dos modelos, o critério de informação de desvio e o desvio baseado em ordenada preditiva condicional favoreceram os modelos multirraciais heterocedásticos em detrimento aos modelos multirraciais homocedástico (Tabela 1). Com relação à avaliação da distribuição marginal dos resíduos (Gaussiana ou  $t$  de Student), os modelos com resíduos  $t$  de Student foram melhores. Dessa forma, o MAMR-T-HE foi o modelo mais adequado por ambos os critérios. Isto implica que esse modelo não somente ajusta melhor os dados conforme indicado pelo seu menor critério de informação de desvio, mas também seu menor desvio baseado em ordenadas preditivas condicionais aponta para uma melhor habilidade preditiva de dados que não fazem parte das observações. Esta última característica é

particularmente relevante ao se avaliarem modelos utilizados em avaliações genéticas, uma vez que o principal objetivo é prever fenótipos não observados, ou seja, o desempenho das futuras gerações. Esses resultados concordam com os obtidos por Cardoso et al. (2005), que aplicaram os mesmos modelos deste estudo na análise do ganho pós desmama de bovinos, produto de cruzamentos entre as raças Nelore e Hereford e concluíram que o modelo  $t$  de Student heterocedástico era o que melhor ajustava os dados.

Nos modelos robustos  $t$  de Student, os graus de liberdade ou parâmetros de robustez ( $\nu$ ) indicam o grau de curtose da distribuição dos resíduos e quanto menores os seus valores mais leptocúrtica e de caudas mais longas são essas distribuições, ou seja, com mais valores que se afastam da média em vários múltiplos do desvio-padrão. Quando considerados desconhecidos, como no presente estudo, os graus de liberdade podem convergir para um valor infinito se a distribuição dos resíduos é normal sem dados extremos (Lange & Sinsheimer, 1993). A média *a posteriori* de  $\nu$  para o MAMR-T-HO e o MAMR-T-HE estão entre quatro e oito (Tabela 2), o que indica a presença de dados extremos com respeito à pressuposição de distribuição normal dos resíduos (Stranden & Gianola, 1998). Isso indica que modelos com distribuição residual  $t$  devem propiciar melhor ajuste aos dados.

A média  $\pm$  desvio-padrão *a posteriori* de  $\nu$  neste estudo foi bastante similar aos resultados obtidos por Cardoso et al. (2005) para o modelo  $t$  de Student homocedástico de

$4,79 \pm 0,21$  e para o modelo  $t$  de Student heterocedástico de  $7,33 \pm 0,48$ , indicando estrutura residual similar nestes dois conjuntos de dados de ganho pós-desmama. Os maiores valores de  $\nu$  associados aos modelos heterocedásticos em relação aos homocedásticos em ambos os estudos era esperado, pois alguns resíduos que aparecem como extremos sob especificações de erro homocedástico e contribuem para menores estimativas de  $\nu$ , ou seja, caudas mais longas, podem pertencer a grupo contemporâneo com maiores variâncias residuais no modelo heterocedástico. Neste caso, essas variâncias aumentadas permitem maior amplitude nos valores dos resíduos e, conseqüentemente, que resíduos moderadamente extremos sejam contemplados dentro da pressuposição de normalidade pela maior dispersão da distribuição (Cardoso et al., 2005).

A heterocedasticidade residual entre os grupos de contemporâneos foi ajustada no modelo pela especificação dos correspondentes efeitos escalares aleatórios  $\tau_{GCi}$  e pelo parâmetro de heterocedasticidade  $\eta$ . Os valores observados de  $\eta$  apresentam evidência de heterocedasticidade entre os grupos de contemporâneos (Tabela 2), o que pode ser verificado também pelo desvio-

padrão dos  $\tau_{GCi}$ ,  $DP(\tau_{GCi}) = \sqrt{\frac{1}{\eta-2}}$ , que foram de 0,71 e 0,63

para MAMR-G-HE e MAMR-T-HE, respectivamente. Esses resultados são muito similares àqueles de Kizilkaya & Tempelman (2005), que estimaram a média *a posteriori* de desvio-padrão ( $\tau_{GCi}$ ) para pesos ao nascer e de escores da

Tabela 1 - Critérios bayesianos para comparação do ajuste de diferentes modelos aos dados de ganho pós-desmama

Modelo <sup>1</sup>	Média do desvio	Número efetivo de parâmetros	Critério de informação de desvio <sup>2</sup>	Desvio das ordenadas preditivas condicionais <sup>2</sup>
MAMR-G-HO	35.094	1.261	36.356	36.649
MAMR-T-HO	34.246	1.582	35.828	36.398
MAMR-G-HE	34.700	1.042	35.742	36.393
MAMR-T-HE	34.264	1.293	35.558	36.272

<sup>1</sup> MAMR-G-HO = Modelo Animal Multirracial Gaussiano homocedástico; MAMR-T-HO = Modelo Animal Multirracial  $t$  de Student homocedástico; MAMR-G-HE = Modelo Animal Multirracial Gaussiano heterocedástico; MAMR-T-HE = Modelo Animal Multirracial  $t$  de Student heterocedástico.

<sup>2</sup> Valores menores indicam melhor ajuste.

Tabela 2 - Médias *a posteriori* para parâmetros de robustez e heterocedasticidade, e para fatores escalares obtidos por diferentes modelos

Parâmetro	Modelo Animal Multirracial $t$ de Student homocedástico	Modelo Animal Multirracial Gaussiano heterocedástico	Modelo Animal Multirracial $t$ de Student heterocedástico
$\nu$	$4,76 \pm 0,51$	na	$8,16 \pm 1,67$
$\eta$	na	$3,97 \pm 0,74$	$4,50 \pm 0,93$
$\gamma_A$	na	$0,73 \pm 0,24$	$0,70 \pm 0,24$
$\gamma_D$	na	$0,35 \pm 0,09$	$0,36 \pm 0,10$
$\tau_{S_1}$	na	$1,38 \pm 0,19$	$1,39 \pm 0,20$
$\tau_{S_2}$	na	$1,01 \pm 0,13$	$1,03 \pm 0,13$
$\tau_{S_3}$	na	$1,00 \pm 0,00$	$1,00 \pm 0,00$

$\nu$  = graus de liberdade ou parâmetro de robustez;  $\eta$  = parâmetro de heterocedasticidade de grupos de contemporâneos;  $\gamma_A$  = fator escalar para proporção da raça Angus;  $\gamma_D$  = fator escalar para heterozigiosidade;  $\tau_{S_1}$  = fator escalar para macho inteiro;  $\tau_{S_2}$  = fator escalar para macho castrado;  $\tau_{S_3}$  = fator escalar para fêmea – restringido ao valor um; na = não se aplica.

facilidade de parto na raça italiana Piedmontês de  $0,74 \pm 0,14$  e  $0,60 \pm 0,09$  para esses respectivos modelos. Resultados similares, embora com maior magnitude, também foram observados por Cardoso et al. (2005), respectivamente de  $0,84 \pm 0,07$  e  $0,72 \pm 0,06$ . Nesses três estudos as médias *a posteriori* de desvio-padrão ( $\tau_{GCi}$ ) sob o modelo de erros gaussianos heterocedásticos foram sempre ligeiramente maiores do que aquelas estimadas pelo modelo de erros *t* de Student heterocedásticos. Esses resultados podem ser atribuídos à especificação de erros *t* de Student que atenua os efeitos residuais extremos, enquanto que a distribuição gaussiana pode resultar em estimativas da variância específica de grupo de contemporâneos inflacionadas e consequentemente maiores estimativas de desvio-padrão ( $\tau_{GCi}$ ) (Cardoso et al., 2005).

A variabilidade residual diminui com o aumento da proporção da raça Angus (Tabela 2), já que a média *a posteriori* de  $g_A$  é menor que 1. Entretanto, grandes desvios-padrão foram observados. Os efeitos da heterozigidade ( $g_D$ ) na heterocedasticidade residual foram mais evidentes. Determinou-se que  $\Pr(\gamma_2 > 1 | y) = 0,0001$  para o modelo de erro *t* de Student heterocedástico, indicando que enquanto a heterozigidade aumenta (e também a heterose), a variabilidade residual diminui. Este resultado é consistente com a teoria que a heterozigidade pode minimizar a variação ambiental (Lerner, 1954), como relatado por outros autores (Cardoso et al., 2005).

A média *a posteriori* do efeito de sexo na variabilidade residual foi muito similar nos dois modelos de erro heterocedástico (Tabela 2). Os machos inteiros ( $\tau_{S_1}$ ) apresentaram variabilidade residual quase 40% maior a

que os machos castrados ( $\tau_{S_2}$ ) e fêmeas ( $\tau_{S_3}$ ). Esses resultados estão de acordo com os obtidos por outros pesquisadores, que concluíram que sexo é uma fonte significativa de heterocedasticidade residual para características de crescimento em gado de corte (Garrick et al., 1989; Rodriguez-Almeida et al., 1995). As diferenças entre sexos observadas neste estudo não podem ser atribuídas exclusivamente a um efeito de escala, pois os animais foram criados em condições ambientais relativamente pobres e os ganhos médios diários observados foram de  $0,35 \pm 0,12$  kg para machos inteiros, de  $0,31 \pm 0,11$  kg para machos castrados e  $0,21 \pm 0,12$  kg para fêmeas. A média de ganho dos machos castrados foi mais próxima dos machos inteiros, enquanto sua variabilidade residual foi mais similar à das fêmeas (Tabela 2), contrariando a expectativa que seria um aumento da variância proporcional ao aumento da média da característica se houvesse efeito de escala.

As estimativas das médias (Tabela 3) e distribuições *a posteriori* (Figura 1) das variâncias genéticas de ganho de peso diário obtidas para as raças Nelore e Angus diferiram substancialmente entre os diferentes modelos. A raça Angus apresentou variância genética em média duas vezes maior que a obtida para a raça Nelore no MAMR-G-HO (Tabela 3), entretanto, para os modelos MAMR-T-HO, MAMR-G-HE e MAMR-T-HE, houve ampla sobreposição dos valores no 95% IPP (Figura 1).

O mesmo padrão de variação das estimativas das variâncias genéticas entre modelos considerando ou não robustez e heterocedasticidade também foi observado por Cardoso et al. (2005) em uma população produto do

Tabela 3 - Componentes de variâncias genéticas aditivas e residuais da população Brangus-Ibagé obtidas por diferentes modelos<sup>1</sup>

Componente de variância	Média <i>a posteriori</i>	Desvio-padrão <i>a posteriori</i>	Moda <i>a posteriori</i>	Intervalo com 95% dos valores mais prováveis <i>a posteriori</i>		Tamanho efetivo da amostra <sup>2</sup>
<b>MAMR-G-HO</b>						
Angus	189,76	27,06	192,28	140,19	a 244,49	819
Nelore	83,58	23,33	75,92	45,88	a 136,50	2115
Residual	365,87	17,70	365,71	331,60	a 400,82	1226
<b>MAMR-T-HO</b>						
Angus	130,07	23,02	124,06	88,79	a 176,91	688
Nelore	93,74	27,76	81,88	46,99	a 156,87	972
Residual	411,48	22,89	409,69	368,66	a 458,96	1940
<b>MAMR-G-HE</b>						
Angus	127,97	26,82	125,66	77,65	a 183,80	488
Nelore	131,82	46,89	106,67	62,04	a 242,47	399
Residual	388,39	28,32	383,61	336,74	a 448,26	1601
<b>MAMR-T-HE</b>						
Angus	113,67	24,26	113,73	70,54	a 166,59	489
Nelore	126,78	40,80	112,62	61,40	a 217,79	385
Residual	402,97	28,26	402,19	351,16	a 462,34	2228

<sup>1</sup> MAMR-G-HO = Modelo Animal Multirracial Gaussiano homocedástico; MAMR-T-HO = Modelo Animal Multirracial *t* de Student homocedástico; MAMR-G-HE = Modelo Animal Multirracial Gaussiano heterocedástico; MAMR-T-HE = Modelo Animal Multirracial *t* de Student heterocedástico.

<sup>2</sup> Conforme proposto por Sorensen et al. (1995).

cruzamento entre as raças Nelore e Hereford. Esses autores observaram que nos modelos homocedásticos Gaussiano e *t* de Student a raça Hereford apresentou maior variabilidade genética, enquanto o inverso foi observado nos

modelos heterocedásticos Gaussiano e *t* de Student, com a raça Nelore tendo maior variância genética.

As estimativas de variância residual marginal obtidas com diferentes modelos considerados apresentaram

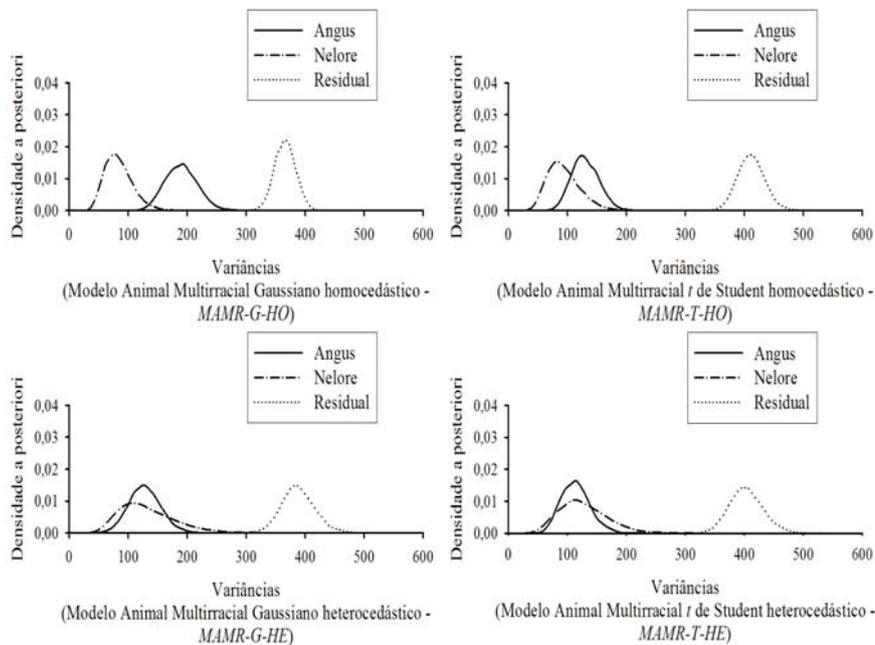


Figura 1 - Densidades *a posteriori* da variância genética aditiva para as raças Angus e Nelore e da variância residual marginal para ganho pós-desmame obtidas por diferentes modelos.

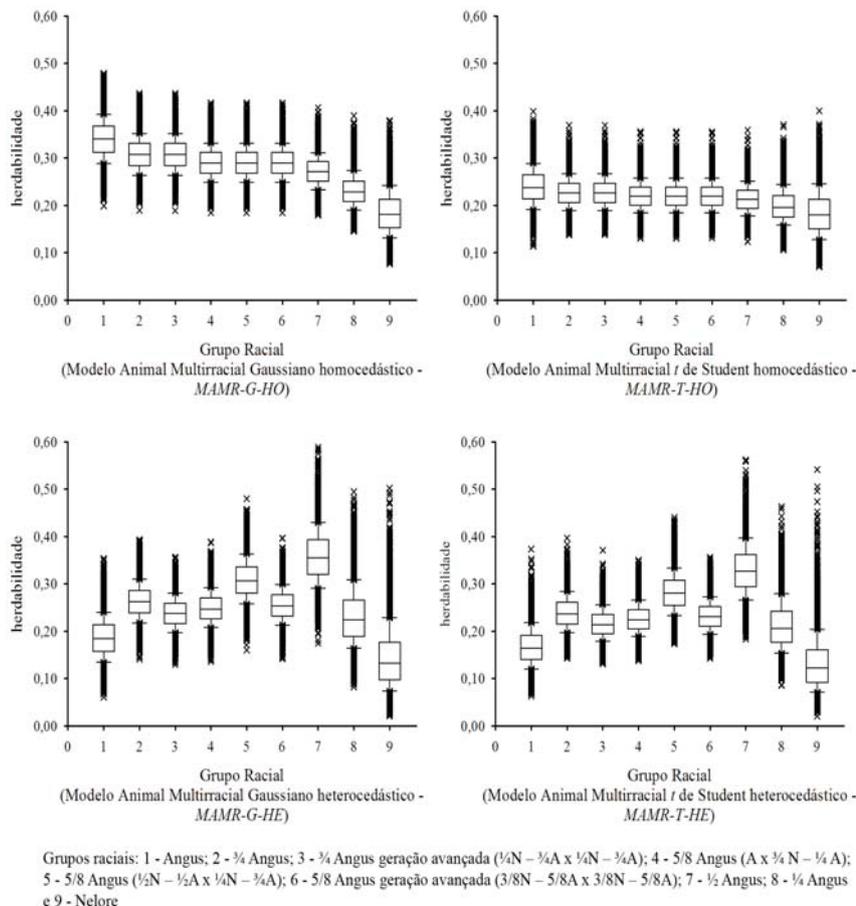


Figura 2 - Distribuição *a posteriori* da herdabilidade para ganho pós-desmame de acordo com os grupos raciais e o modelo utilizado.

sobreposição de todos os 95% IPP e grandes diferenças entre as estimativas de variância residual dos diferentes modelos não foram observadas.

Nos modelos com homocedasticidade residual, MAMR-G-HO e MAMR-T-HO, as diferenças em herdabilidade são exclusivamente devidas à variância genética uma vez que a variância residual marginal é a mesma para todos os grupos e, portanto, as herdabilidades observadas foram crescentes com a proporção de Angus e com maiores diferenças para MAMR-G-HO (Figura 2). Nos modelos heterocedásticos, MAMR-G-HE e MAMR-T-HE, observam-se maiores diferenças de herdabilidade entre os distintos grupos raciais (Figura 2), pois, além de diferenças genéticas, estas incorporam também diferenças na variância residual devido à raça e à heterozigosidade. Assim esses modelos representam alternativas metodológicas para maximizar a eficiência das avaliações genéticas quando fenótipos de animais cruzados são utilizados, especialmente, pelo adequado ajuste da dispersão dos valores genéticos dentro de cada composição racial presente na população.

As correlações de Spearman entre os valores genéticos obtidos pelos modelos MAMR-G-HO, MAMR-T-HO, MAMR-G-HE e MAMR-T-HE foram estimadas considerando todos os animais avaliados, incluindo touros, vacas e produtos, e, separadamente, para touros que aparecem como pais de outros animais com registro de ganho de peso diário, para os animais classificados como 10% superiores de todos os animais e para os touros com progênie 10% superiores (Tabela 4). As estimativas de correlações entre todos os modelos foram altas entre os valores genéticos quando considerados todos os animais e ou todos os

touros com progênie (maior ou igual a 0,970). Entretanto, quando foram incluídos nessas estimativas somente indivíduos classificados como 10% superiores, que justamente são os de maior interesse na seleção, a concordância de ordenação entre os modelos decresceu substancialmente, variando entre 0,496 e 0,948, quando incluídos animais de todas as categorias e entre -0,03 e 0,721 quando considerados somente touros com progênie.

As correlações de ordem entre MAMR-G-HE e MAMR-T-HE foram relativamente altas para todas as situações estudadas, indicando concordância entre os ordenamentos desses modelos (Tabela 4). Por outro lado, entre MAMR-G-HO e MAMR-T-HO, quando considerados animais e touros com progênie classificados nos 10% superiores, a correlação de ordem entre valores genéticos para os animais candidatos à seleção diminuiu consideravelmente (Tabela 4). Esses últimos resultados têm maiores implicações para as avaliações genéticas e as decisões de seleção que as correlações resultantes quando são envolvidos todos os animais. As correlações para os animais superiores podem ser esperadas como mais baixas quando registros discrepantes pela distribuição gaussiana devem conduzir no MAMR-G-HO às predições mais extremas do mérito genético do animal, em relação ao MAMR-T-HO, onde as caudas longas atenuam dados extremos (Cardoso et al., 2007). Correlações relativamente baixas também foram obtidas entre MAMR-G-HO e MAMR-G-HE para os 10% melhores animais e touros com progênie, MAMR-G-HE provavelmente permitiria uma seleção mais equilibrada dos animais através dos ambientes por considerar a heterocedasticidade, visto que, quando as variâncias heterogêneas são ignoradas na modelagem estatística, uma proporção maior dos animais dos ambientes mais variáveis tende a ser classificada no extremo superior da distribuição de valores genéticos (Hill, 1984; Gianola, 1986).

## Conclusões

Avaliações genéticas multirraciais assumindo variâncias residuais heterogêneas, e não normalidade de resíduos, podem ser mais apropriadas para realizar inferência quanto ao mérito genético de animais produtos de cruzamento entre as raças Angus e Nelore. A flexibilidade de modelos que consideram resíduos heterocedásticos permite estimar herdabilidades distintas para diferentes grupos raciais. Heterozigosidade e efeitos ambientais podem influenciar a variabilidade residual e conseqüentemente a predição de valores genéticos para ganho de peso pós desmama de animais cruzados. Assim, esses fatores devem ser considerados na estimação de componentes de variância.

Tabela 4 - Correlação de ordem de Spearman entre valores genéticos individuais obtidos por diferentes modelos<sup>1</sup> para todos os animais avaliados e para touros-pais, considerando todos os indivíduos (acima da diagonal) e somente os 10% superiores (abaixo da diagonal)

	MAMR-G-HO	MAMR-T-HO	MAMR-G-HE	MAMR-T-HE
Touros pais <sup>2</sup>				
MAMR-G-HO		0,986	0,974	0,974
MAMR-T-HO	0,576		0,970	0,980
MAMR-G-HE	0,430	-0,030		0,993
MAMR-T-HE	0,200	0,261	0,721	
Todos os animais				
MAMR-G-HO		0,983	0,973	0,969
MAMR-T-HO	0,714		0,977	0,987
MAMR-G-HE	0,594	0,752		0,995
MAMR-T-HE	0,496	0,819	0,948	

<sup>1</sup> MAMR-G-HO = Modelo Animal Multirracial Gaussiano homocedástico; MAMR-T-HO = Modelo Animal Multirracial *t* de Student homocedástico; MAMR-G-HE = Modelo Animal Multirracial Gaussiano heterocedástico; MAMR-T-HE = Modelo Animal Multirracial *t* de Student heterocedástico.

<sup>2</sup> Reprodutores machos pais de outros animais com registro.

## Agradecimentos

À Embrapa Pecuária Sul e ao Dr. Nelson Manzoni de Oliveira, pelo fornecimento dos dados utilizados neste trabalho. Ao CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico, pela concessão da bolsa de estudos.

## Referências

- CARDOSO, F.F. **Manual de utilização do Programa INTERGEN - Versão 1.0 em estudos de genética quantitativa animal**. Embrapa Pecuária Sul. Bagé, 2008. p.74. (Documentos 74).
- CARDOSO, F.F.; ROSA, G.J.M.; TEMPELMAN, R.J. Multiple-breed genetic inference using heavy-tailed structural models for heterogeneous residual variances. **Journal of Animal Science**, v.83, n.8, p.1766-1779, 2005.
- CARDOSO, F.F.; ROSA, G.J.M.; TEMPELMAN, R.J. Accounting for outliers and heteroskedasticity in multibreed genetic evaluations of postweaning gain of Nelore-Hereford cattle. **Journal of Animal Science**, v.85, n.4, p.909-918, 2007.
- CARVALHEIRO, R.; FRIES, L.A.; SCHENKEL, F.S. et al. Efeitos da Heterogeneidade de Variância Residual entre Grupos de Contemporâneos na Avaliação Genética de Bovinos de Corte. **Revista Brasileira de Zootecnia**, v.31, p.1680-1688, 2002.
- FOULLEY, J.L.; CRISTOBAL, M.S.; GIANOLA, D. et al. Marginal likelihood and Bayesian approaches to the analysis of heterogeneous residual variances in mixed linear Gaussian models. **Computational Statistics & Data Analysis**, v.13, n.3, p.291-305, 1992.
- GARRICK, D.J.; POLLAK, E.J.; QUAAS, R.L. et al. Variance heterogeneity in direct and maternal weight traits by sex and percent purebred for Simmental-sired calves. **Journal of Animal Science**, v.67, n.10, p.2515-2528, 1989.
- GELFAND, A.E. Model determination using sampling-based methods. In: GILKS, W.R.; RICHARDSON, S. et al. (Eds.). **Markov Chain Monte Carlo in practice**. London: Chapman & Hall, 1996. Model determination using sampling-based methods, 1996. p.145-161.
- GIANOLA, D. On selection criteria and estimation of parameters when the variance is heterogeneous. **Theoretical and Applied Genetics**, v.72, n.5, p.671-677, 1986.
- GIANOLA, D.; FOULLEY, J.L.; FERNANDO, R.L. et al. Estimation of heterogeneous variances using empirical bayes methods - Theoretical considerations. **Journal of Dairy Science**, v.75, n.10, p.2805-2823, 1992.
- HILL, W.G. On selection among groups with heterogeneous variance. **Animal Production**, v.39, p.473-477, 1984.
- KINGHORN, B. The expression of recombination loss in quantitative traits. **Zeitschrift Fur Tierzucht Und Zuchtungsbiologie - Journal of Animal Breeding and Genetics**, v.97, n.2, p.138-143, 1980.
- KIZILKAYA, K.; TEMPELMAN, R.J. **An assessment of heteroskedastic t-error linear mixed models for the analysis of field data collected from diverse environments**. International Biometric Society - ENAR Spring Meeting. Tampa, FL: 2003. 62p.
- KIZILKAYA, K.; TEMPELMAN, R.J. A general approach to mixed effects modeling of residual variances in generalized linear mixed models. **Genetics Selection Evolution**, v.37, n.1, p.31-56, 2005.
- LANGE, K.; SINSHEIMER, J.S. Normal/independent distributions and their applications in robust regression. **Journal of the American Statistical Association**, v.2, n.2, p.175-198, 1993.
- LERNER, I.M. **Genetic homeostasis**. Edinburgh: Oliver and Boyd, 1954. 134p.
- LO, L.L.; FERNANDO, R.L.; GROSSMAN, M. Covariance between relatives in multibreed populations: additive model. **Theoretical And Applied Genetics**, v.87, n.4, p.423-430, 1993.
- MARTINS, E.N. Avaliação genética e heterogeneidade de variância. In: REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE ZOOTECNIA, 2002, Recife. **Anais...** Recife, 2002. p.205-214.
- OLIVEIRA, M.M.; CARDOSO, F.F.; OSÓRIO, J.C.S. Componentes de variância e parâmetros genéticos em uma população multirracial Nelore-Angus sob enfoque Bayesiano. **Revista Brasileira de Zootecnia**, v.39, p.2426-2433, 2010 .
- RODRIGUEZ-ALMEIDA, F.A.; VANVLECK, L.D.; CUNDIFF, L.V. et al. Heterogeneity of variance by sire breed, sex, and dam breed in 200-day and 365-day weights of beef-cattle from a top cross experiment. **Journal of Animal Science**, v.73, n.9, p.2579-2588, 1995.
- 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.
- ROSO, V.M.; SCHENKEL, F.S. AMC - A computer program to assess the degree of connectedness among contemporary groups. In: WORLD CONGRESS ON GENETICS APPLIED TO LIVESTOCK PRODUCTION, 8., 2006, Belo Horizonte, MG, Brasil. **Proceedings...** Belo Horizonte, 2006. p.27-26.
- SANCRISTOBAL, M.; FOULLEY, J.L.; MANFREDI, E. Inference about multiplicative heteroskedastic components of variance in a Mixed Linear Gaussian Model with an application to beef-cattle breeding. **Genetics Selection Evolution**, v.25, n.1, p.3-30, 1993.
- SORENSEN, D.A.; ANDERSEN, S.; GIANOLA, D. et al. Bayesian-inference in threshold models using Gibbs sampling. **Genetics Selection Evolution**, v.27, n.3, p.229-249, 1995.
- SORENSEN, D.A.; GIANOLA, D. **Likelihood, Bayesian, and MCMC methods in quantitative genetics**. New York: Springer-Verlag, 2002. 740p.
- SORENSEN, D.A.; WAAGEPETERSEN, R. Normal linear models with genetically structured residual heterogeneity: A case study. **Genetics Research**, v.82, p.207-222, 2003.
- SPIEGELHALTER, D.J.; BEST, N.G.; CARLIN, B.P. et al. Bayesian measures of model complexity and fit. **Journal of the Royal Statistical Society Series B-Statistical Methodology**, v.64, p.583-616, 2002.
- STRANDEN, I.; GIANOLA, D. Attenuating effects of preferential treatment with Student-t mixed linear models: a simulation study. **Genetics Selection Evolution**, v.30, n.6, p.565-583, 1998.
- STRANDEN, I.; GIANOLA, D. Mixed effects linear models with t-distributions for quantitative genetic analysis: a Bayesian approach. **Genetics Selection Evolution**, v.31, n.1, p.25-42, 1999.