

Dos Desvios das Médias dos Grupos Contemporâneos aos Modelos Animais

Luiz Alberto Fries¹

RESUMO - O objetivo deste trabalho foi apresentar a reparametrização de um modelo de desvios das médias dos grupos contemporâneos até um modelo animal não-restrito (equações dos modelos mistos de Lush - EMMML). Aplicando restrições sobre as EMMML, foi obtido o modelo animal segundo Henderson (equações dos modelos mistos de Henderson - EMMH). Para obter as EMMH, restrições do tipo $\mathbf{X}'\mathbf{A}^{-1}\boldsymbol{\lambda}\mathbf{a}^H = \mathbf{0}$ foram impostas e estas geraram a necessidade de pressuposições fortes, como ausência de seleção, para que as propriedades das EMMH pudessem ser provadas. Para diferentes situações examinadas, as restrições foram confirmadas. O somatório destas restrições pode ser erroneamente interpretado como propriedade das soluções de um modelo animal de que os animais-base tenham a soma dos seus valores genéticos igual a zero.

Palavras-chave: modelo misto sob seleção, propriedade genética, reparametrização, restrição

From Means Deviations of the Contemporary Groups up to Animal Models

ABSTRACT - The objective of this work was to present a reparametrization of a means deviation model from contemporary groups up to an unrestricted animal model (Lush's mixed model equations - LMME). Applying restrictions on the LMME, the animal model was obtained in accordance with Henderson (Henderson's mixed model equations - HMME). To obtain the HMME, restrictions of the type $\mathbf{X}'\mathbf{A}^{-1}\boldsymbol{\lambda}\mathbf{a}^H = \mathbf{0}$ were imposed and these demanded the necessity of unrealistic assumptions, as the absence of selection, so that the HMME properties could be proved. For the different examined situations, the restrictions were confirmed. The summation of these restrictions can be incorrectly understood as property of the solutions of an animal model in that the animals-base have the summation of their genetic values equal to zero.

Key Words: mixed model under selection, genetic property, reparameterization, restriction

Introdução

Durante décadas, o critério de seleção (LUSH, 1945) básico disponível foi o "desvio da média do Grupo Contemporâneo (GC)" e/ou funções destes desvios. Em gado de corte, por exemplo, pertencem ao mesmo GC de desmama os animais que nascem na mesma fazenda, ano e estação, são do mesmo sexo e manejados nos mesmos grupos e pastos até a sua desmama, realizada na mesma data. A pressuposição utilizada para o uso deste critério é a de que não existem diferenças genéticas entre as médias dos GC, o que restringe o uso pleno do método a comparações dentro de GC.

Esta situação pode ser descrita por intermédio da equação $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\delta}$, em que

\mathbf{y} é um vetor de produções medidas nos n indivíduos;

\mathbf{X} é uma matriz de incidência que indica a distribuição dos n indivíduos nos p GC;

$\boldsymbol{\beta}$ é um vetor de efeitos fixos dos p GC (note que neste modelo $\boldsymbol{\beta}$ contera todos os efeitos comuns aos animais do mesmo GC, inclusive diferentes valores genéticos médios); e

$\boldsymbol{\delta}$ é um vetor de resíduos ou desvios das médias dos GC.

Por meio da teoria dos índices de seleção, o problema é resolvido em dois passos:

(1) obtém-se uma solução para $\boldsymbol{\beta}^0$, que é igual à média aritmética de cada GC; e

(2) os desvios das médias dos GC's são calculados como $\boldsymbol{\delta}^0 = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}^0$.

Depois disso, o valor genético dos animais é predito por $\mathbf{a}^0 = \mathbf{h}^2\boldsymbol{\delta}^0$. Portanto, com o conhecimento da herdabilidade (LUSH, 1945), os desvios das médias dos GC são decompostos em:

(1) \mathbf{a}^0 , a parte genética aditiva (pelo menos no quanto ela é diferente da média do GC); e

(2) \mathbf{e}^0 , o resíduo (efeitos ambientais mais efeitos genéticos não-aditivos e de interação).

Ou seja:

$$\boldsymbol{\delta}^0 = \mathbf{1}\boldsymbol{\delta}^0 = (\mathbf{h}^2 + \mathbf{1} - \mathbf{h}^2)\boldsymbol{\delta}^0 = \mathbf{h}^2\boldsymbol{\delta}^0 + (\mathbf{1} - \mathbf{h}^2)\boldsymbol{\delta}^0 = \mathbf{a}^0 + \mathbf{e}^0$$

Assim, mesmo na falta de informações sobre a genealogia dos animais ou de capacidade operacional para utilizar esta informação, a teoria dos índices de seleção (LUSH, 1945) indica como estes desvios

¹ GenSys Consultores Associados S/C Ltda e DZ-MGA/UNESP/Jaboticabal-FAPESP. Endereço atual: Animal Genetics and Breeding Unit (AGBU) - UNE - Armidale, NSW 2351. fries@fcav.unesp.br

podem ser utilizados para prever o valor genético dos animais.

No final da década de 40, por meio de modificação nas equações normais do método dos quadrados mínimos, Henderson (HENDERSON, 1984) desenvolveu as bases para a metodologia dos modelos mistos (MMM), que passa a ser utilizada rotineiramente em gado de corte na década de 80, inclusive no Brasil. A regra mundial em gado de corte é a utilização dos modelos animais (MA), nas suas várias aplicações. SCHAEFFER (1996) apresenta e discute alguns exemplos de modelos animais.

A equação matemática de um modelo animal pode ser descrita por $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a} + \mathbf{e}$. Além dos termos já definidos, a matriz \mathbf{Z} é utilizada nesta equação como uma matriz de incidência que associa observações aos efeitos aleatórios (por exemplo, os genéticos aditivos) de cada animal. Em alguns casos, a matriz \mathbf{Z} é igual a uma identidade. Com relação à equação do modelo de desvios das médias dos GC, na equação do modelo animal substitui-se $\boldsymbol{\delta}$ por $\mathbf{Z}\mathbf{a} + \mathbf{e}$ ou, possivelmente, $\boldsymbol{\delta} = \mathbf{Z}\mathbf{a} + \mathbf{e} = \mathbf{I}\mathbf{a} + \mathbf{e}$.

O presente trabalho tem por objetivo demonstrar que as equações de um modelo animal podem ser obtidas a partir de uma reparametrização do modelo de desvios das médias dos GC. Como resultado, obtêm-se as equações de Lush. Quando restrições são impostas, obtêm-se as equações da MMMH/MA, que geram soluções condicionais. Este espaço restrito destas soluções condicionais fornecidas pela MMMH/MA é entendido como "propriedades genéticas do modelo animal".

O presente trabalho também objetiva demonstrar, por meio do uso de alguns exemplos clássicos da literatura, que estas "propriedades genéticas" são dependentes:

- (1) da distribuição dos dados nos GC e do nível de conectabilidade/conexidade destes;
- (2) se a matriz de restrições possui posto máximo (ou completo) de linhas; e
- (3) se a matriz de parentesco é completa ou não.

Materiais e Métodos

Solução do problema por intermédio do método dos quadrados mínimos (MQM)

A equação de um modelo animal, representada por $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{I}\mathbf{a} + \mathbf{e}$, pode ser utilizada para formular o problema de se estimarem simultaneamente os p efeitos ambientais dos GC ($\boldsymbol{\beta}$) e os n efeitos genéticos aditivos (\mathbf{a}). Por meio do método dos quadrados mínimos, isto seria equivalente a resolver o seguinte

problema (super-parametrizado, com $p+n$ equações), contando com n unidades de informação:

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{I} \\ \mathbf{I}'\mathbf{X} & \mathbf{I}'\mathbf{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \mathbf{a} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{I}'\mathbf{y} \end{bmatrix} \quad \{1\}$$

A alternativa oferecida pelo MQM, para fornecer a solução $[\boldsymbol{\beta}^0 \mathbf{a}^0]'$, é impor a restrição $\mathbf{X}'\mathbf{I}\mathbf{a} = \mathbf{0}$. Esta matriz de restrições significa que a soma dos valores genéticos estimados (ou a soma dos desvios das médias dos GC) dos animais pertencentes a cada GC deve ser zero para cada um e para todos os GC's e contém, portanto, p restrições, correspondendo às p linhas de \mathbf{X}' . É uma restrição não-paradoxal, alinhada com a teoria predominante em genética quantitativa e que não cria impasse de interpretação dos parâmetros ou das soluções; mas, mesmo assim, não menos importante e restritiva.

Como $\mathbf{X}'\mathbf{I}\mathbf{a} = \mathbf{X}'\mathbf{a}$ faz parte da primeira das equações em $\{1\}$, ou seja $\mathbf{X}'\mathbf{X}\boldsymbol{\beta} + \mathbf{X}'\mathbf{I}\mathbf{a} = \mathbf{X}'\mathbf{y}$ é a primeira linha de $\{1\}$, a restrição $\mathbf{X}'\mathbf{I}\mathbf{a} = \mathbf{0}$ pode ser diretamente substituída nesta (ou aplicada na primeira equação de $\{1\}$) resultando então em $\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{y}$. Sob esta restrição, quebra-se a dependência entre equações (não são mais simultâneas), embora nada impeça que elas sejam mantidas em conjunto único, e, sob $\mathbf{X}'\mathbf{I}\mathbf{a} = \mathbf{X}'\mathbf{a} = \mathbf{0}$, as equações $\{1\}$ se tornam

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{0} \\ \mathbf{I}'\mathbf{X} & \mathbf{I}'\mathbf{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}^0 \\ \mathbf{a}^0 \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{I}'\mathbf{y} \end{bmatrix} \quad \{2\}$$

A solução $\boldsymbol{\beta}^0$ (um vetor com as médias aritméticas das observações de cada GC), obtida da primeira equação de $\{2\}$ e usada na segunda equação, permite a solução $\mathbf{a}^0 = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}^0$. Note que o agora exposto é idêntico ao descrito acima, na Introdução, como o primeiro de um processo em dois passos ou estádios, ou seja, $\mathbf{a}^0 = \boldsymbol{\delta}^0 = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}^0$ continua sendo exatamente o mesmo "desvio da média do GC".

Portanto, a equação $\{1\}$ poderia ter sido escrita como

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{I} \\ \mathbf{I}'\mathbf{X} & \mathbf{I}'\mathbf{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\delta} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{I}'\mathbf{y} \end{bmatrix} \quad \{3\}$$

enquanto a equação $\{2\}$ poderia ter sido escrita como

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{0} \\ \mathbf{I}'\mathbf{X} & \mathbf{I}'\mathbf{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}^0 \\ \boldsymbol{\delta}^0 \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{I}'\mathbf{y} \end{bmatrix} \quad \{4\}$$

Combinando o MQM com a Teoria dos Índices de Seleção

A regressão dos valores genéticos aditivos \mathbf{a} sobre os valores fenotípicos \mathbf{y} é uma das muitas

definições da herdabilidade (h^2). Esta pode ser representada por:

$$h^2 = \text{Cov}(\mathbf{a}, \mathbf{y}) / \text{Var}(\mathbf{y}) = \\ \text{Var}(\mathbf{a}) / [\text{Var}(\mathbf{a}) + \text{Var}(\mathbf{e})] = \\ 1 / [1 + \text{Var}(\mathbf{e})/\text{Var}(\mathbf{a})] \text{ ou} \\ h^2 = 1/[1 + \lambda], \text{ para } \lambda = \text{Var}(\mathbf{e})/\text{Var}(\mathbf{a}).$$

Usar o conhecimento da h^2 ou de λ para as n últimas equações do sistema pode ser interpretado como estar utilizando n unidades extras de informação e, portanto, a condição está dada para que o sistema se torne determinado. Portanto, não é necessário impor as p restrições $\mathbf{X}'\mathbf{I}\boldsymbol{\delta} = \mathbf{0}$ ou $\mathbf{X}\boldsymbol{\delta}' = \mathbf{0}$ ou $\mathbf{X}'\mathbf{a} = \mathbf{0}$, quando o conhecimento de h^2 é utilizado.

Para adicionar esta informação de forma estatisticamente correta ou ortodoxa, o sistema de equações deve ser reparametrizado, considerando que o desejado é um preditor/estimador de \mathbf{a} (será usado o símbolo \mathbf{a}^L para o critério de seleção de Lush, definido abaixo).

Para o caso de valores genéticos ou DEP's de touros, LUSH (1945) faz o seguinte comentário: "O valor numérico real da solução para cada touro é em parte determinado pelo seu valor genético real, mas parcialmente também por quaisquer erros tendenciosos que as soluções contenham além dos resíduos aleatórios. Assim sendo, estas soluções irão variar mais do que os valores genéticos reais dos touros. Os touros com as soluções/estimativas mais altas são geralmente bons touros, mas provavelmente não tão bons quanto as suas soluções. Similarmente, os touros com soluções baixas são geralmente touros deficientes, mas usualmente não tão pobres quanto suas soluções. Nós erraremos menos se estimarmos o valor genético de cada touro como um ponto localizado entre a sua solução e a média da raça."

Com escalares, para errar menos e no caso de desvios das médias dos GC, deve-se usar a relação $\mathbf{a}_i^L = h^2 \boldsymbol{\delta}_i^o$ e, portanto, a reparametrização necessária é dada pelo inverso da herdabilidade, pois $\boldsymbol{\delta}_i^o = (h^2)^{-1} \mathbf{a}_i^L = [\mathbf{I} + \lambda] \mathbf{a}_i^L$.

Para os vetores $\boldsymbol{\delta}$ ($\boldsymbol{\delta} = \mathbf{I}\mathbf{a} + \mathbf{e}$) e \mathbf{a} , é necessário definir:

$\text{Var}(\mathbf{a}) = \text{Cov}(\mathbf{a}, \boldsymbol{\delta}') = \mathbf{G} = \mathbf{A}\boldsymbol{\sigma}_a^2$, em que \mathbf{A} é a matriz de coeficientes de parentesco;

$$\text{Var}(\mathbf{e}) = \mathbf{I}\boldsymbol{\sigma}_e^2; \text{Cov}(\mathbf{a}, \mathbf{e}') = \mathbf{0}; \text{ e}$$

$$\text{Var}(\boldsymbol{\delta}) = \text{Var}(\mathbf{I}\mathbf{a} + \mathbf{e}) = \mathbf{A}\boldsymbol{\sigma}_a^2 + \mathbf{I}\boldsymbol{\sigma}_e^2 = \mathbf{V}.$$

A distribuição condicional da densidade de \mathbf{a} dado o vetor $\boldsymbol{\delta}$ é dada por

$$\mathbf{a} = \text{Cov}(\mathbf{a}, \boldsymbol{\delta}') [\text{Var}(\boldsymbol{\delta})]^{-1} \boldsymbol{\delta} = \\ \mathbf{A}\boldsymbol{\sigma}_a^2 [\mathbf{A}\boldsymbol{\sigma}_a^2 + \mathbf{I}\boldsymbol{\sigma}_e^2]^{-1} \boldsymbol{\delta} = \mathbf{G}\mathbf{V}^{-1} \boldsymbol{\delta}$$

GIANOLA et al. (1989) definem este produto $\mathbf{G}\mathbf{V}^{-1}$ como \mathbf{H} e chamam \mathbf{H} de "versão ou equiva-

lente multivariado da herdabilidade" e, portanto, $\mathbf{a} = \mathbf{H}\boldsymbol{\delta}$ é apenas a versão multivariada de $\mathbf{a}_i^L = h^2 \boldsymbol{\delta}_i^o$.

O uso do MQM, da teoria dos índices de seleção (LUSH, 1945) e da matriz de parentesco de Wright permite obter o seguinte critério de seleção ou estimador/preditor dos valores genéticos dos animais:

$$\mathbf{a}^L = \mathbf{H}\boldsymbol{\delta} = \mathbf{G}\mathbf{V}^{-1}\boldsymbol{\delta} = \mathbf{G}\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \\ [\mathbf{I} + \mathbf{A}^{-1}\lambda]^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}),$$

em que $\boldsymbol{\beta}$ representa verdadeiramente os efeitos ambientais e só os efeitos ambientais dos GC.

Portanto, a relação necessária para a reparametrização, mantendo a definição de $\boldsymbol{\beta}$, fazendo $\boldsymbol{\beta}^L = \boldsymbol{\beta}$, é

$$\boldsymbol{\delta} = \mathbf{V}\mathbf{G}^{-1}\mathbf{a} = [\mathbf{A}\boldsymbol{\sigma}_a^2 + \mathbf{I}\boldsymbol{\sigma}_e^2] [\mathbf{A}\boldsymbol{\sigma}_a^2]^{-1} \mathbf{a} = [\mathbf{I} + \mathbf{A}^{-1}\lambda] \mathbf{a}$$

$$\text{Para reparametrizar } \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\delta} \end{bmatrix} \text{ de } \{3\} \text{ para } \boldsymbol{\gamma} = \begin{bmatrix} \boldsymbol{\beta}^L \\ \mathbf{a}^L \end{bmatrix}$$

por intermédio de $\boldsymbol{\gamma} = \mathbf{C} \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\delta} \end{bmatrix}$, dada a matriz

$$\mathbf{C} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} + \mathbf{A}^{-1}\lambda \end{bmatrix}^{-1}, \text{ usa-se:}$$

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{I} \\ \mathbf{I}'\mathbf{X} & \mathbf{I}'\mathbf{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\delta} \end{bmatrix} =$$

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{I} \\ \mathbf{I}'\mathbf{X} & \mathbf{I}'\mathbf{I} \end{bmatrix} \mathbf{I} \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\delta} \end{bmatrix} =$$

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{I} \\ \mathbf{I}'\mathbf{X} & \mathbf{I}'\mathbf{I} \end{bmatrix} \mathbf{C}^{-1} \mathbf{C} \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\delta} \end{bmatrix} =$$

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{I} \\ \mathbf{I}'\mathbf{X} & \mathbf{I}'\mathbf{I} \end{bmatrix} \mathbf{C}^{-1} \begin{bmatrix} \boldsymbol{\beta}^L \\ \mathbf{a}^L \end{bmatrix} =$$

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{I} \\ \mathbf{I}'\mathbf{X} & \mathbf{I}'\mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} + \mathbf{A}^{-1}\lambda \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}^L \\ \mathbf{a}^L \end{bmatrix} =$$

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{I}[\mathbf{I} + \mathbf{A}^{-1}\lambda] \\ \mathbf{I}'\mathbf{X} & \mathbf{I}'\mathbf{I}[\mathbf{I} + \mathbf{A}^{-1}\lambda] \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}^L \\ \mathbf{a}^L \end{bmatrix} =$$

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{I} + \mathbf{X}'\mathbf{A}^{-1}\lambda \\ \mathbf{I}'\mathbf{X} & \mathbf{I} + \mathbf{A}^{-1}\lambda \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}^L \\ \mathbf{a}^L \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{I}'\mathbf{y} \end{bmatrix} \{5\}$$

Nota-se que este resultado é obtido usando apenas o MQM e a teoria dos índices de seleção.

Portanto, é apropriado denominá-las de equações dos modelos mistos de Lush (EMML), ou, no caso específico de um modelo animal, de EMML/MA.

Por estarem baseadas no MQM, as EMML dependem:

- (1) de que as matrizes de incidência (no caso, \mathbf{X} e \mathbf{I}) sejam conhecidas com exatidão total; e
- (2) de que os resíduos tenham distribuição idêntica (entre níveis dos efeitos do modelo) e independente de todos os outros resíduos (representada por $e \sim \text{IID}(0, \mathbf{I}\sigma_e^2)$).

No caso do MQM ponderado, esta pressuposição é suavizada para $\text{Var}(e)=\mathbf{R}$, para uma matriz \mathbf{R} qualquer. A extensão das EMML para este caso é simples e direta. Nota-se que a forma da distribuição dos resíduos é livre para a etapa de estimação dos parâmetros de tendência central; ou seja, não é necessário pressupor/invocar normalidade ou qualquer outra distribuição para usar as EMML ou suas soluções.

As únicas outras pressuposições utilizadas para obter as EMML para um modelo descrito por uma equação geral $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a} + \mathbf{e}$ são aquelas necessárias para obter as matrizes de variância/covariância das variáveis aleatórias e que são explicitadas por:

$$\mathbf{E}[\mathbf{y}] = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a}, \mathbf{E}[\mathbf{a}] = \mathbf{a}, \text{ para } \text{Var}(\mathbf{y}) = \mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R}, \text{Var}(\mathbf{a}) = \mathbf{G}, \text{Var}(e) = \mathbf{R} \text{ e } \text{Cov}(\mathbf{a}, e') = \mathbf{0},$$

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

Uma maneira mais direta e elegante de chegar ao mesmo resultado {5} é pela substituição da mesma expressão de $\boldsymbol{\delta}$ nas duas equações do sistema {3}, obtendo-se diretamente as equações que permitem a estimação simultânea de $\boldsymbol{\beta}$ e \mathbf{a} . Substituindo $\boldsymbol{\delta}$ por $[\mathbf{I} + \mathbf{A}^{-1}\boldsymbol{\lambda}] \mathbf{a}^L$ em {3} obtêm-se igualmente:

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{I} \\ \mathbf{I}'\mathbf{X} & \mathbf{I}'\mathbf{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\delta} \end{bmatrix} =$$

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{I} \\ \mathbf{I}'\mathbf{X} & \mathbf{I}'\mathbf{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ [\mathbf{I} + \mathbf{A}^{-1}\boldsymbol{\lambda}] \mathbf{a}^L \end{bmatrix} =$$

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{I}[\mathbf{I} + \mathbf{A}^{-1}\boldsymbol{\lambda}] \\ \mathbf{I}'\mathbf{X} & \mathbf{I}'\mathbf{I}[\mathbf{I} + \mathbf{A}^{-1}\boldsymbol{\lambda}] \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \mathbf{a}^L \end{bmatrix} =$$

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{I} + \mathbf{X}'\mathbf{A}^{-1}\boldsymbol{\lambda} \\ \mathbf{I}'\mathbf{X} & \mathbf{I} + \mathbf{A}^{-1}\boldsymbol{\lambda} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \mathbf{a}^L \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{I}'\mathbf{y} \end{bmatrix}$$

Se na segunda equação de {3} $\mathbf{I}'\boldsymbol{\delta}$ for substituído diretamente por $[\mathbf{I} + \mathbf{A}^{-1}\boldsymbol{\lambda}] \mathbf{a}$, estar-se-ia "emendando" {3} e obtendo as EMMH (ver {6} abaixo) de forma heterodoxa.

Note que a MMMH é baseada no (quase) mesmo conjunto de pressuposições utilizadas na MMML, à exceção de que, para a MMMH:

$\mathbf{E}[\mathbf{y}] = \mathbf{X}\boldsymbol{\beta}$, o valor esperado de qualquer observação é dado apenas pelos efeitos fixos de GC, ou seja, todos os animais de um mesmo GC possuem o mesmo valor esperado, independentemente do valor genético do indivíduo ou do de seus pais ou da genealogia do indivíduo; e/ou

$\mathbf{E}[\mathbf{a}] = \mathbf{0}$, um vetor de zeros, ou seja, o valor esperado de todo e qualquer animal é zero; por isso, nenhuma informação prévia ou processo seletivo pode existir.

Outra maneira de obter as EMMH é por intermédio da imposição da restrição $\mathbf{X}'\mathbf{A}^{-1}\boldsymbol{\lambda} \mathbf{a}^H = \mathbf{0}$ sobre as recém obtidas EMML, eliminando-as da primeira equação de {5}, quando então se obtêm as equações simétricas, condicionais a esta restrição e chamadas de equações dos modelos mistos de Henderson (EMMH) ou EMMH para o modelo animal (EMMH/MA):

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{I} \\ \mathbf{I}'\mathbf{X} & \mathbf{I} + \mathbf{A}^{-1}\boldsymbol{\lambda} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}^H \\ \mathbf{a}^H \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{I}'\mathbf{y} \end{bmatrix} \quad \{6\}$$

Um dos objetivos deste trabalho é demonstrar que esta restrição dá origem às chamadas "propriedades" das soluções do modelo animal. Por exemplo, se a base (ou cerne ou 'kernel') de \mathbf{X} contiver uma coluna de $\mathbf{1}$'s, então a soma das estimativas dos valores genéticos dos animais-base será zero, pois, para as linhas correspondentes aos animais-base, $\mathbf{1}'\mathbf{A}^{-1}$ fornece o valor $\mathbf{1}$ (se a matriz \mathbf{A} for completa, conforme alertado por ALBUQUERQUE, 1997), enquanto para os animais restantes (com pais conhecidos) sua soma de linha ou coluna é igual a $\mathbf{0}$. Em geral, $\mathbf{1}'\mathbf{A}^{-1} = [\mathbf{1} \dots \mathbf{1} \mathbf{0} \dots \mathbf{0}]'$. Caso \mathbf{X} não contenha uma coluna de $\mathbf{1}$'s explicitamente, mesmo assim esta propriedade será constatada quando um modelo tipo "médias de subclasses ou de caselas" for utilizado. Nas EMMH/MA, se \mathbf{X} definir os GC, então a soma das soluções dos valores genéticos dos animais-base presentes em cada GC será zerada. Se um GC não contiver nenhum animal-base e todos os animais presentes tivessem informação completa de genealogia, então este GC não estaria gerando nenhuma restrição. Como regra, apenas o conjunto destas restrições (a

soma dos valores genéticos de todos os animais-base, presentes no total dos GC's) é visualizado.

Para resolver {1} ou {3}, pelo MQM, $\mathbf{X}'\boldsymbol{\delta}=\mathbf{0}$ são impostas. Em {5}, nas EMMML, nenhuma restrição sobre os efeitos genéticos ou aleatórios é imposta e $\mathbf{X}'\boldsymbol{\delta} = \mathbf{X}'\mathbf{I}\mathbf{a} + \mathbf{X}'\mathbf{A}^{-1}\boldsymbol{\lambda}$ \mathbf{a} está presente. Em {6}, nas EMMH, restrições são impostas e, quanto menor a h^2 (ou maior $\boldsymbol{\lambda}$), maior a soma de quadrados associada a esta restrição (SEARLE, 1971).

Note que em $\boldsymbol{\delta} = [\mathbf{I} + \mathbf{A}^{-1}\boldsymbol{\lambda}]\mathbf{a} = \mathbf{I}\mathbf{a} + \mathbf{A}^{-1}\boldsymbol{\lambda}\mathbf{a}$, o segundo termo, correspondendo ao chamado 'resíduo' e que é utilizado na restrição para obter {6}, contém as estimativas dos efeitos genéticos não-aditivos, das interações e do não-especificado no modelo. A decomposição de $\boldsymbol{\delta}$ se dá por intermédio das relações de variâncias. Restrições sobre o complemento dos efeitos aditivos estimados implicam em forçar o estabelecimento de relações sobre os efeitos aditivos estimados.

Outra maneira de obter as EMMML é seguir HENDERSON (1984), o qual apresenta que, caso $\boldsymbol{\beta}$ fosse conhecido, o melhor preditor (BP) de \mathbf{a} seria $\mathbf{a}^{\text{BP}} = [\mathbf{I} + \mathbf{A}^{-1}\boldsymbol{\lambda}]^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$. Definindo $\boldsymbol{\delta} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$, tem-se $\mathbf{a}^{\text{BP}} = [\mathbf{I} + \mathbf{A}^{-1}\boldsymbol{\lambda}]^{-1}\boldsymbol{\delta}$, que permite reparametrizar {3} (que poderia ser igualmente um problema MQM Generalizado ou Ponderado), por meio da relação $\boldsymbol{\delta} = [\mathbf{I} + \mathbf{A}^{-1}\boldsymbol{\lambda}]\mathbf{a}^{\text{BP}}$, fornecendo as equações:

$$\mathbf{X}'\mathbf{X}\boldsymbol{\beta} + \mathbf{X}'\mathbf{I}\boldsymbol{\delta} = \mathbf{X}'\mathbf{X}\boldsymbol{\beta} + \mathbf{X}'\mathbf{I}[\mathbf{I} + \mathbf{A}^{-1}\boldsymbol{\lambda}]\mathbf{a}^{\text{BP}} = \mathbf{X}'\mathbf{X}\boldsymbol{\beta} + \mathbf{X}'\mathbf{I}\mathbf{a}^{\text{BP}} + \mathbf{X}'\mathbf{A}^{-1}\boldsymbol{\lambda}\mathbf{a}^{\text{BP}} = \mathbf{X}'\mathbf{y} \quad \{5.1\}$$

$$\mathbf{I}'\mathbf{X}\boldsymbol{\beta} + \mathbf{I}'\mathbf{I}\boldsymbol{\delta} = \mathbf{I}'\mathbf{X}\mathbf{b} + [\mathbf{I} + \mathbf{A}^{-1}\boldsymbol{\lambda}]\mathbf{a}^{\text{BP}} = \mathbf{I}'\mathbf{y} \quad \{5.2\}$$

Os coeficientes das equações {5.1} e {5.2} são iguais aos das EMMML {5} e sua única diferença com as EMMH reside na primeira equação de {6}, na qual inexistente o termo correspondendo a $\mathbf{X}'\mathbf{A}^{-1}\boldsymbol{\lambda}\mathbf{a}^{\text{BP}}$, porque a substituição é realizada apenas na segunda equação de {3}, "emendando" esta com $\mathbf{A}^{-1}\boldsymbol{\lambda}$.

A consequência da ausência de um termo $\mathbf{X}'\mathbf{A}^{-1}\boldsymbol{\lambda}\mathbf{a}^{\text{H}}$ nas EMMH/MA {6} é que as soluções obtidas obedecem a este fato e se comportam como se a restrição $\mathbf{X}'\mathbf{A}^{-1}\boldsymbol{\lambda}\mathbf{a}^{\text{H}} = \mathbf{0}$ tivesse sido imposta. *O exemplo de KENNEDY, SCHAEFFER e SORENSEN (1988)*

No exemplo de KENNEDY et al. (1988), oito observações são feitas sobre oito animais representados por $\mathbf{a}' = [a_1 \ a_2 \ a_3 \ a_4 \ a_5 \ a_6 \ a_7 \ a_8]$. Os quatro primeiros animais formam a população base, com pais desconhecidos. Os animais 5 e 6 são filhos de 1 e 2, sendo, portanto, irmãos completos. O animal 7 é

filho de 3 e 4, enquanto o animal 8 é filho de 5 e 6 e, portanto, consanguíneo.

As EMMH/MA são:

8.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.00	3.00	1.00	0.00	0.00	-1.00	-1.00	0.00	0.00
1.00	1.00	3.00	0.00	0.00	-1.00	-1.00	0.00	0.00
1.00	0.00	0.00	2.50	0.50	0.00	0.00	-1.00	0.00
1.00	0.00	0.00	0.50	2.50	0.00	0.00	-1.00	0.00
1.00	-1.00	-1.00	0.00	0.00	3.50	0.50	0.00	-1.00
1.00	-1.00	-1.00	0.00	0.00	0.50	3.50	0.00	-1.00
1.00	0.00	0.00	-1.00	-1.00	0.00	0.00	3.00	0.00
1.00	0.00	0.00	0.00	0.00	-1.00	-1.00	0.00	3.00

$$\begin{bmatrix} \mu \\ \mathbf{a} \ 1 \\ \mathbf{a} \ 2 \\ \mathbf{a} \ 3 \\ \mathbf{a} \ 4 \\ \mathbf{a} \ 5 \\ \mathbf{a} \ 6 \\ \mathbf{a} \ 7 \\ \mathbf{a} \ 8 \end{bmatrix} = \begin{bmatrix} 72 \\ 10 \\ 9 \\ 8 \\ 7 \\ 9 \\ 10 \\ 8 \\ 11 \end{bmatrix}$$

No exemplo original, todos os animais, das três gerações, têm as performances medidas num único ambiente e estas são: [10 9 8 7 9 10 8 11]. A relação entre as variâncias residual e genética é 1.0. *O exemplo de HENDERSON, KEMPTHORNE, SEARLE e VON KROSIGK (1959)*

Neste trabalho clássico, o modelo animal (com medidas repetidas de produção de leite de uma mesma vaca em diferentes anos - note que o próprio HENDERSON (1988) aponta este trabalho como 'aparentemente a primeira aplicação de BLUP para um modelo animal') contém o efeito fixo ambiental de ano de parto, o efeito fixo de grupo genético (o trabalho sugere que estes grupos genéticos seriam os próprios touros-pais das vacas), ambos representados pela parte $\mathbf{X}\boldsymbol{\beta}$ do modelo, e o efeito aleatório da i^{ma} vaca, representado pela "capacidade mais provável de produção" (CMPP) e descrito por intermédio da parte $\mathbf{Z}\mathbf{c}$ do modelo. Tentando manter a notação do trabalho original (ver pg 195-6 deste, em que é definido que $\text{Var}(\mathbf{c}) = \mathbf{D}$), as matrizes de coeficientes das EMMML e as das EMMH, junto com o seu lado direito ('RHS') comum, são apresentadas abaixo, nesta ordem, como nova versão de {5} e {6}.

$$\begin{bmatrix} X'X & X'Z + X'Z(Z'Z)^{-1}D^{-1} \\ Z'X & Z'Z + D^{-1} \end{bmatrix} \begin{bmatrix} \beta^L \\ c^L \end{bmatrix} =$$

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + D^{-1} \end{bmatrix} \begin{bmatrix} \beta^H \\ c^H \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix} \quad \{7\}$$

As equações acima resultam de uma adaptação do original, usando a pressuposição que $\text{Var}(e) = \mathbf{I}\sigma_e^2 = \mathbf{R}$. FRIES e SCHENKEL (1993) demonstraram que o MQM ponderado, considerando os efeitos genéticos como fixos, ou as equações da MML ou o MQM generalizado de um modelo completo, fornecem as mesmas soluções verdadeiramente BLUE para β . As equações da MMMH produzem soluções de um MQM generalizado de um modelo incompleto ou parcial, que apenas contempla os efeitos β , conforme provado por (HENDERSON et al., 1959), e ignora os efeitos genéticos ou aleatórios.

Resultados e Discussão

Resultados do exemplo de KENNEDY, SCHAEFFER e SORENSEN (1988)

A estimativa da média geral pelo MQM Ordinário é 9.00 e, usando este como um valor paramétrico, a teoria dos índices de seleção fornece:

$$\mathbf{a}^{BP} = [\mathbf{0.6786 \quad 0.1786 \quad -0.5357 \quad -1.0357} \\ \mathbf{0.4405 \quad 0.7738 \quad -0.8571 \quad 1.0714}]'$$

Usando as EMMH, a solução para o efeito fixo é 8.7059 e

$$\mathbf{a}^H = [\mathbf{0.8676 \quad 0.3676 \quad -0.3676 \quad -0.8676} \\ \mathbf{0.6716 \quad 1.0049 \quad -0.6471 \quad 1.3235}]'$$

A soma de todas as soluções ($\mathbf{X}'\mathbf{a}^H = \mathbf{1}'\mathbf{a}^H$) é 2.352941. Estes resultados são também apresentados na publicação original.

Note que a soma das soluções para os quatro primeiros animais é zero e, portanto, $\mathbf{X}'\mathbf{A}^{-1}\lambda \mathbf{a}^H = \mathbf{0}$ se verifica pois $\mathbf{X}'\mathbf{A}^{-1}\lambda = [\mathbf{1 \quad 1 \quad 1 \quad 1 \quad 0 \quad 0 \quad 0 \quad 0}]$ e esta "propriedade genética das soluções" é interpretada como "os animais-base da população (sem genealogia conhecida) têm valor genético médio esperado zero".

KENNEDY, SCHAEFFER e SORENSEN (1988) chamam a atenção de que: "Um entendimento e a correta definição da população base são críticas para modelos animais. Com o uso de \mathbf{A} , animais com pais não identificados são tratados/considerados como animais da população base e pressupostos como tomados ao acaso de uma mesma população, com mérito genético 0 e variância comum σ_a^2 . A aplica-

ção de modelos animais, sem a devida consideração para a definição dos animais que formam a população base e em relação aos quais todas as estimativas de mérito genético subseqüentes estão ligadas, pode conduzir a erros. Se os animais da população base na verdade são tomados de populações com diferentes médias genéticas, como no caso de populações em cruzamento ou quando animais jovens com paternidade desconhecida entram na população continuamente, então o modelo deverá levar em consideração a estrutura destas subpopulações e permitir diferentes valores médios esperados para os animais base de cada subpopulação."

Infelizmente, estas notas acauteladoras são pouco valorizadas. Como um exemplo rudimentar disto, pode ser mencionada a falta de uso rotineiro de testes de conectabilidade/conexidade antes da realização de avaliações genéticas em populações não-experimentais. Infelizmente, apenas a parte 'fácil' do trabalho de KENNEDY, SCHAEFFER e SORENSEN (1988) é utilizada para convencer os usuários de que as soluções do modelo animal obedecem a certas dogmáticas 'propriedades genéticas'.

Para avaliar as conseqüências destas restrições, o mesmo exemplo foi utilizado, alterando apenas a possível distribuição dos animais através de GC's. Três situações (A, B e C) foram examinadas e são abaixo descritas, com seus resultados. Em algumas destas situações, foi mantido o parâmetro para a média geral, para facilitar a discussão dos resultados (obviamente então impondo limitações de estimabilidade de que a soma das soluções para GC's seja igual a zero).

Situação A

Os animais 1 e 2 pertencem ao GC 1; os animais 3 e 4, ao GC 2; e o restante (5, 6, 7 e 8), ao GC 3. Portanto, os animais-base estariam divididos em GC's distintos e apenas animais-base estariam representados nestes GC's. Todos os descendentes (filhos e netos) destes animais-base encontram-se reunidos num único GC.

A solução das EMMH para os efeitos fixos é dada por $[\mathbf{8.7333 \quad 0.4667 \quad -0.9333 \quad 0.4667}]$ e $\mathbf{a}^H = [\mathbf{0.5500 \quad 0.0500 \quad -0.0500 \quad -0.5500} \\ \mathbf{0.2833 \quad 0.6167 \quad -0.6000 \quad 0.9000}]$. É interessante notar que $\mathbf{X}'\mathbf{a}^H = [\mathbf{1.2 \quad 0.6 \quad -0.6 \quad 1.2}]$.

Estas soluções são obtidas como conseqüência de $\mathbf{X}'\mathbf{A}^{-1}\lambda \mathbf{a}^H = \mathbf{0}$. A verificação de que as restrições se cumprem é simples e podem ser feitas por inspeção. Para a presente situação, a matriz

$X'A^{-1}\lambda$ representa o conjunto de restrições de nulidade sobre a^H (as linhas da matriz $X'A^{-1}\lambda$ são separadas por //):

$$X'A^{-1}\lambda = \{ \{ 1.0 \ 1.0 \ 1.0 \ 1.0 \ 0.0 \ 0.0 \ 0.0 \ 0.0 \} // \\ [3.0 \ 3.0 \ 0.0 \ 0.0 \ -2.0 \ -2.0 \ 0.0 \ 0.0] // \\ [0.0 \ 0.0 \ 2.0 \ 2.0 \ 0.0 \ 0.0 \ -2.0 \ 0.0] // \\ [-2.0 \ -2.0 \ -1.0 \ -1.0 \ 2.0 \ 2.0 \ 2.0 \ 0.0] \}.$$

A restrição imposta por intermédio da primeira linha da matriz acima, originada do fato de o parâmetro para média geral ter sido incluído no modelo, estabelece a restrição Σ (sigma ou somatório), conhecida como "propriedade genética das soluções" e interpretada como "os animais-base da população (sem genealogia conhecida) têm valor genético esperado zero".

A restrição representada pela segunda linha determina que $3*(a_1 + a_2) = 2*(a_5 + a_6)$. Esta restrição, geneticamente, é arbitrária e não depende dos valores y observados. Por que, em qualquer situação, o valor genético médio dos filhos deveria ser igual a 1,5 vezes a média do valor genético dos pais? Note que $(a_5 + a_6) = (3/2)*(a_1 + a_2)$ se cumpre, pois $(0,9) = (1,5)(0,60)$.

Na terceira linha é estabelecido um contraponto ou uma contradição às outras restrições, pois aqui o valor genético estimado para o filho é igual a soma (note que é a soma, e não a média) dos valores genéticos estimados dos pais, ou seja, $a_7 = (a_3 + a_4)$ ou $-0.6 = -0.05 - 0.55$.

A melhor maneira encontrada para interpretar a quarta restrição é a de que esta linha determina que o duplo da soma dos valores genéticos estimados dos filhos (animais 5, 6 e 7), com o valor 0.6, seja igual à soma ponderada dos valores genéticos dos pais, considerando o número de filhos de cada pai. Esta última relação até seria discutível/aceitável, como situação geral.

As três últimas linhas/restrições estabelecem relações genéticas conceitualmente não-congruentes entre si e a soma destas resulta na primeira, que impõe a conhecida restrição Σ sobre os animais-base. Por outro lado, nenhuma das linhas envolve o animal 8, constatando-se que qualquer valor medido em y para este animal poderia ter sido observado, e isto não seria capaz de alterar em nada as relações/valores genéticos determinados pelas restrições acima.

Tendo conhecimento e considerando estas restrições, como delinear e definir, por exemplo, os pré-requisitos de um teste de progênie? Como distribuir os acasalamentos/progênes de cada touro nos GC de maneira que os resultados não sejam resultantes/dependentes desta distribuição? Como evitar que as

restrições tipo da segunda e da terceira linha acima influenciem os resultados?

Com certeza, se delineamentos simples fossem utilizados, com testes realizados dentro de unidades experimentais, as restrições teriam uma forma simples, poderiam ser evidenciadas e talvez chamassem a atenção. Contudo, a verificação de problemas deste tipo por intermédio da mera inspeção/constatação de sumários de touros ou resultados de avaliações genéticas é dificultada pela participação nestes de grande número de touros, possibilitando que os erros determinados por estas restrições fiquem dissimulados. Mesmo assim, quanto dos saltos/inversões/discrepâncias que ocorrem entre sumários poderiam ser explicados por estas restrições?

Situação B

Os animais 1, 2, 5, 6 e 8 pertencem ao GC 1 e os animais 3, 4 e 7, ao GC 2; portanto, os animais-base estariam divididos em GC's distintos e todos os descendentes (filhos e netos) de cada subconjunto de animais-base encontram-se reunidos no GC da sua base. Claramente existe uma situação de dados desconectados ou desconexos, sem nenhum laço genético entre os dois grupos.

As soluções das EMMH/MA são obtidas sem qualquer problema ou mesmo sem que alguém necessariamente se dê conta de que existem problemas de conectabilidade/conexidade.

As estimativas para os efeitos dos GC's para esta situação são $[9.6667 \ 7.6250]$ e

$$a^H = [0.2500 \ -0.2500 \ 0.2500 \ -0.2500 \ -0.0833 \ 0.2500 \ 0.1250 \ 0.5000].$$

Vale a pena comparar e notar as discrepâncias entre os valores obtidos para a^H nas duas situações.

Quando todos os animais de cada subpopulação estão dentro de um mesmo GC, a propriedade associada à primeira das restrições da situação anterior (restrição Σ) aplica-se a cada uma dessas populações isoladas.

Ou seja, $X'A^{-1}\lambda = \{ [1.0 \ 1.0 \ 0.0 \ 0.0 \ 0.0 \ 0.0 \ 0.0 \ 0.0] //$

$$[0.0 \ 0.0 \ 1.0 \ 1.0 \ 0.0 \ 0.0 \ 0.0 \ 0.0] \}.$$

Pode-se inferir que é justamente a aplicação destas restrições que permite que soluções sejam fornecidas pelas EMMH, sempre, corrigindo problemas de deficiência de posto causados por dados desconexos. Assim, é possível identificar a causa da enorme facilidade computacional (pelo menos em relação às EMMH), na solução das EMMH, de sempre fornecerem uma solução e da rápida convergência dos processos iterativos baseados nestas.

A soma das restrições acima resulta na restrição maior (Σ), válida para todos os animais-base, independentemente de suas equações serem conexas ou não.

KENNEDY et al. (1988) alertam para a necessidade de definir bem as populações e os animais-base de cada uma delas e relatam que a interpretação dos resultados é dependente desta distribuição. Esses autores reportam ainda que situações deste tipo, com problemas de estrutura e formação de subpopulações, "podem ser complementados através do uso de uma estratégia de agrupamento judiciosa/cuidadosa, como a proposta por Westell e Van Vleck (1987), e que este problema foi discutido com maiores detalhes por Henderson (1988)." Grupos genéticos podem ser bom remédio ou 'mal menor', em algumas situações. Outra opção é fazer real análise de conectabilidade, separando populações desconectadas e oferecendo relatórios separados por subpopulações, para os criadores. Porém, este assunto de conectabilidade e grupos genéticos merece mais espaço para ser discutido.

Situação C

Os animais 1, 2, 3 e 4 pertencem ao GC 1; os animais 5, 6 e 7, ao GC 2; e o animal restante, 8, está sozinho no GC 3. Portanto, todos os animais-base estariam presentes em um só GC, seus filhos estariam em um segundo GC e seus netos estariam presentes no último destes GC's. É uma situação com confundimento total entre GC's e gerações, o que é uma situação extrema, mas em linha com o que acontece na prática. Pelo menos, normalmente existe alta correlação entre o ano de nascimento do animal e o número de sua geração.

As soluções para os efeitos dos três GC's são dadas por $[8.5000 \ 8.8182 \ 10.4091]$ e

$$\mathbf{a}^H = [0.7955 \ 0.2955 \ -0.2955 \ -0.7955 \ 0.4242 \ 0.7576 \ -0.6364 \ 0.5909] .$$

Estas soluções são condicionais às seguintes restrições: $\mathbf{X}'\mathbf{A}^{-1}\boldsymbol{\lambda} = \{ [3.0 \ 3.0 \ 2.0 \ 2.0 \ -2.0 \ -2.0 \ -2.0 \ 0.0] // [-2.0 \ -2.0 \ -1.0 \ -1.0 \ 3.0 \ 3.0 \ 2.0 \ -2.0] // [0.0 \ 0.0 \ 0.0 \ 0.0 \ -1.0 \ -1.0 \ 0.0 \ 2.0] \}$.

A primeira restrição estabelece uma relação entre os somatórios dos valores genéticos dos animais da primeira e da segunda geração, para determinada (e geneticamente injustificável) ponderação (10 vs. 6). Nota-se que o número de animais-base (cuja soma de linha é 1,0) a mais no GC 1 é de quatro a mais que no GC 2. Apenas a título de especulação, note-se que se um resultado assim for repetido nos diferentes rebanhos de uma população comercial, estaria criado mecanismo espúrio para fabricar progressos genéticos

que não existem.

A segunda restrição estabelece relação entre os somatórios dos animais da segunda geração e o resto (primeira mais terceira geração). Não é possível justificar ou entender os valores dos coeficientes mas, pelo menos, para esta segunda linha, a soma dos coeficientes positivos é igual à soma dos coeficientes negativos.

A terceira linha é lógica: como o animal 8 está sozinho no GC 3, o seu valor genético só pode ser estimado como função média dos valores genéticos estimados/preditos dos seus pais.

Em qualquer situação, a soma das linhas de restrições sempre resulta em um vetor que corresponde à restrição imposta aos animais-base para que tenham soma igual a zero (restrição Σ).

Resultados do exemplo de HENDERSON, KEMPTHORNE, SEARLE e VON KROSIGK (1959)

Para este exemplo, a matriz de restrições que determina os valores das soluções para as CMPP das nove vacas, e que possui sete linhas, correspondendo aos sete parâmetros: média geral, os três anos de partos e os três grupos fixos (em uma das possibilidades apresentadas pelo autor, estes grupos poderiam ser interpretados como os três pais) das vacas, é:

$$\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{D}^{-1}\boldsymbol{\lambda} = \{ \begin{array}{cccccccccc} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & // \\ 1/3 & 1/3 & 1/2 & 1 & 0 & 0 & 0 & 0 & 0 & // \\ 1/3 & 1/3 & 1/2 & 0 & 1/2 & 1/2 & 1 & 0 & 0 & // \\ 1/3 & 1/3 & 0 & 0 & 1/2 & 1/2 & 0 & 1 & 1 & // \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & // \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & // \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & \} .$$

Como esperado, as mesmas deficiências de posto coluna de \mathbf{X} aparecem como deficiências de posto linha da matriz de restrições. Por exemplo, nota-se que a primeira linha é igual à soma das três seguintes, que também é igual à soma das três últimas. Além disso, algumas linhas são incongruentes com as outras. Para este exemplo, com uma matriz \mathbf{X}' com posto linha igual a 3, apenas as três últimas restrições (e a soma destas) se verificam. Qual restrição prepondera (se verifica) sobre quais outras é uma questão a ser resolvida.

Por exemplo, a segunda linha impõe $1/3c_1 + 1/3c_2 + 1/2c_3 + c_4 = 0$, enquanto a quinta linha de restrições força $c_1 + c_2 + c_3 + c_4 = 0$. Esta última restrição estabelece que a soma dos valores genéticos das filhas do touro 1 é igual a zero. Nota-se que esta última restrição tem sido entendida como propriedade das soluções, em modelos mistos com efeitos aleató-

rios aninhados. Nota-se também que a soma das últimas três restrições resulta na primeira, que estabelece a conhecida restrição Σ .

Para este exemplo, $[\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{D}^{-1}\mathbf{c}^{\mathbf{H}}]' = \{0 \ -17.5 \ 0.8 \ 16.7 \ 0 \ 0 \ 0\}$ e a soma das restrições 2, 3 e 4 resultam em zero. Vale a pena indicar ainda que, como consequência destas restrições, as soluções das EMMH indicam que o ano 2 foi sensivelmente pior que o ano 3, ao passo que as soluções pelo MQM Generalizado de um modelo completo (e portanto com propriedades BLUE) indicam o oposto. Isto indica que, embora as restrições sejam sobre os efeitos genéticos, as soluções para os outros efeitos (fixos ou ambientais) simultâneos do modelo também serão influenciados por estas restrições.

Este exemplo mostra que restrições reais, e não apenas limitações/constrições sobre as soluções (SEARLE, 1971), foram impostas sobre as equações, alterando os valores de funções estimáveis e formas quadráticas.

Como comentário final, pode ser dito que a simplificação das equações, por intermédio da restrição da forma $\mathbf{X}'\mathbf{A}^{-1}\boldsymbol{\lambda}\mathbf{a}^{\mathbf{H}} = \mathbf{0}$, é a origem das "propriedades das soluções" e da necessidade de fazer pressuposições, como a de ausência de seleção para que as soluções das EMMH tenham propriedades BLUE e BLUP (HENDERSON (1975) e SCHAEFFER (1996)).

Nota-se que esta restrição implica que diferentes grupos de animais-base utilizados em diferentes níveis de \mathbf{X} (diferentes locais ou anos ou GC) terão valor médio zero (ou a soma de suas soluções será zero), sendo, portanto, necessário que, na prática, seja viável e plausível pressupor que estes diferentes grupos de animais-base tenham idêntica esperança.

Nota-se a semelhança com: "A pressuposição utilizada para o uso deste critério [desvios de médias ou teoria dos índices de seleção] é a de que não existem diferenças genéticas entre as médias dos GC." Por outro lado, as EMMH que não contêm esta restrição mantêm suas propriedades mesmo na presença de seleção dos futuros pais, conforme demonstrado por FRIES e SCHENKEL (1993).

Esta restrição, ainda, lembra as várias equações propostas para serem adicionadas sobre as EMMH para contemplarem os diferentes processos seletivos (HENDERSON, 1975) e também as equações para grupos genéticos, sempre uma função aproximada da parte de \mathbf{X} , que descreve tempo e/ou origem dos animais-base. Quando isto é feito, as equações se tornam mais complexas que as EMMH, talvez até mais

que as EMMH, e o resultado produzido é sempre uma alteração/redução nos efeitos da seleção. O grau de redução dos efeitos de seleção dependerá do fato de a estratégia de grupamento ser capaz de reproduzir/captar as decisões seletivas que de fato foram tomadas.

HENDERSON (1975) demonstra que, se a seleção for do tipo 'invariante à translação', quando então $\mathbf{L}'\mathbf{X} = \mathbf{0}$ ou $\mathbf{X}'\mathbf{L} = \mathbf{0}$, as EMMH fornecerão soluções que não serão viesadas por este tipo de seleção. As decisões tomadas com relação ao valor genético dos animais, representadas por \mathbf{L} , podem ser apenas re-expressões da matriz $\mathbf{A}^{-1}\boldsymbol{\lambda}$, de dimensão $n \times n$, e estar apenas criando formas equivalentes ou alternativas às p restrições $\mathbf{X}'\mathbf{A}^{-1}\boldsymbol{\lambda}\mathbf{a}^{\mathbf{H}} = \mathbf{0}$ que de qualquer forma serão impostas.

Deveria ser melhor divulgado que a única seleção compatível com manutenção das propriedades das EMMH é a seleção entre animais dentro dos níveis de \mathbf{X} , ou seja, dentro dos GC. Portanto, "seleção dentro dos GC's" seria tradução operacional (e com interpretação clara) da obscura "seleção invariante à translação".

Pode ser dito que a teoria dos índices de seleção não proporciona comparações entre animais de diferentes GC's, a não ser que se pressuponha que as médias genéticas dos animais dos diferentes GC's sejam semelhantes e iguais a zero. Enquanto isto, a EMMH obtém estimativas/preditores por meio de pressuposição ainda mais forte (a de que o valor genético esperado de cada um dos animais é zero) e não permite que informações entre animais de diferentes GC's sejam comparadas e utilizadas para a seleção, se as propriedades do método quiserem ser mantidas. Isto significa que um método não proporciona comparações e o outro não permite que estas comparações sejam utilizadas.

Em gado de corte, a maior limitação que existia em relação à teoria dos índices de seleção era justamente a incapacidade de comparar animais de diferentes GC's, ou de diferentes rebanhos ou entre anos.

Conclusões

Para diferentes situações, verificam-se restrições do tipo $\mathbf{X}'\mathbf{A}^{-1}\boldsymbol{\lambda}\mathbf{a}^{\mathbf{H}} = \mathbf{0}$ para as soluções das EMMH. Caso a matriz de restrições seja de posto linha incompleto, então apenas parte destas restrições será verificada. Outros estudos deverão ser realizados para tentar determinar quais das restrições impostas (limitadas pelo posto de \mathbf{X}) preponderam sobre as restantes e qual a causa da maior ascendência de

algumas destas restrições. Sem ter este ponto esclarecido, fica difícil falar em funções estimáveis ou em propriedades das soluções.

Em cada uma das situações examinadas, o somatório das restrições sobre as soluções das EMMH pode ser interpretado como uma propriedade genética de que os animais-base tenham valor genético esperado igual a zero. Estas restrições são a origem das chamadas "propriedades genéticas" do modelo animal, ou seja, as 'propriedades genéticas' das soluções são consequência das relações entre as colunas de \mathbf{X} .

A dependência que foi aqui estabelecida entre o valor das soluções fornecidas pelas EMMH, em sinal e magnitude, e as restrições impostas gera inquietações mesmo quando a pressuposição de ausência de seleção for rigorosamente verdadeira.

As restrições que são impostas sobre as soluções das EMMH não dependem dos valores observados \mathbf{y} (são funções apenas de elementos da mão esquerda das equações), mas sim da definição de \mathbf{X} , da distribuição dos animais nos níveis de \mathbf{X} , do número de gerações utilizadas para definir \mathbf{A} , se esta matriz de parentesco for completa ou não, e das relações de variâncias.

Agradecimentos

Aos colegas do GenSys e UFRGS (especialmente Flávio Schenkel), da UNESP (especialmente Lucia Albuquerque) e, de forma especial, ao Dr. Larry Schaeffer, pelas saudáveis e frutíferas discussões, pela troca de idéias e pela novas compreensão destes assuntos. Contudo, a responsabilidade do trabalho é exclusivamente do autor.

Aos revisores da Revista Brasileira de Zootecnia, pelo exame detalhado, pelas sugestões valiosas e sucessivas correções.

Referências Bibliográficas

- ALBUQUERQUE, L.G. 1997. Comunicação Pessoal.
- FRIES, L.A., SCHENKEL, F.S. Estimation and prediction under a selection model. In: REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE ZOOTECNIA, 30, 1993, Rio de Janeiro. *Anais...* Viçosa: SBZ, 1993, p.1-20.
- GIANOLA, D., R.L. FERNANDO, S. I.M., FOULLEY, J.L. 1989. Likelihood estimation of quantitative genetic parameters when selection occurs: models and problems. *Genome*, 31:768-777.
- HENDERSON, C.R. 1975. Best linear unbiased estimation and prediction under a selection model. *Biometrics*, 31: 423-447.
- HENDERSON, C.R. 1984. *Applications of linear models in animal breeding*. University of Guelph. 462p.
- HENDERSON, C.R. 1988. Theoretical basis and computational methods for a number of different animal models. *J. Dairy Sci.*, 71:1-16 (Suplemento 2).
- HENDERSON, C.R., KEMPTHORNE, O., SEARLE, S.R. et al. 1959. The estimation of environmental and genetic trends from records subject to culling. *Biometrics*, 15:192-218.
- KENNEDY, B.W., SCHAEFFER, L.R., SORENSEN, D.A. 1988. Genetic properties of animal models. *J. Dairy Sci.*, 71:17-26 (Suplemento 2).
- LUSH, J.L. 1945. *Animal Breeding Plans*. Ames: Iowa State University Press.
- SCHAEFFER, L.R. Animal models: why, how and when to use them. In: SIMPÓSIO NACIONAL DE MELHORAMENTO ANIMAL, 1, 1996, Ribeirão Preto, SP. *Anais ...* Ribeirão Preto: SBMA 1996, p.21-40.
- SEARLE, S.R. 1971. *Linear models*. New York: John Wiley & Sons.

Recebido em: 23/09/97

Aceito em: 08/12/98