

USO DA MÁXIMA VEROSSIMILHANÇA RESTRITA E TRANSFORMAÇÃO CANÔNICA PARA ESTIMAÇÃO DE PARÂMETROS GENÉTICOS DE CARACTERÍSTICAS DE CRESCIMENTO EM BOVINOS

ALFREDO RIBEIRO DE FREITAS¹, ADRIANA DE CÁSSIA FAVORETTI², MAURÍCIO MELLO DE ALENCAR¹, MARIA JOSÉ PEGORIN²

RESUMO - Estimativas de parâmetros genéticos foram obtidas para o peso ao nascimento (Y_1), peso observado (Y_2) e ajustado (Y_3) à desmama, aos doze meses de idade (Y_4) e ganho de peso da desmama aos doze meses de idade (Y_5) de animais da raça Canchim. O método multivariado da Máxima Verossimi-lhança Restrita (REML), com a mesma matriz de desenho para todas as características, foi aplicada a um modelo de touros que incluiu, além do efeito médio, sexo,

ano, época de nascimento, interação ano x época e idade da vaca (efeitos fixos), e touro e erro como efeitos aleatórios, para estimar variâncias e covariâncias. As estimativas de herdabilidade (h^2) das características individuais foram de 0,33, 0,77, 0,77, 0,44 e 0,10, para Y_1 até Y_5 , respectivamente. O uso da transformação canônica foi utilizado para reduzir a análise multivariada por REML em análises univariadas. A primeira variável canônica com maior

1 - Pesquisador, EMBRAPA/CPPSE/CNPq, Caixa Postal 339, São Carlos, SP.

2 - Bacharel em Estatística - Univ. Federal de São Carlos, SP.

variabilidade genética apresentou uma h^2 de 0,83, consideravelmente maior do que as obtidas para as variáveis individuais. Como as características são correlacionadas, essa técnica, além de discernir a relação biológica entre elas, é uma alternativa valiosa para medir a importância relativa de cada uma delas, e selecionar mais eficientemente os animais pelo seu mérito genético.

Palavras-chave: características de crescimento, componentes de variâncias, gado de corte, transformação canônica.

USE OF RESTRICTED MAXIMUM LIKELIHOOD AND CANONICAL TRANSFORMATION FOR ESTIMATING GENETIC PARAMETERS FOR GROWTH TRAITS IN CATTLE

ABSTRACT - Genetic parameters estimates for birth weight (Y_1), observed (Y_2) and adjusted (Y_3) weaning weight, adjusted weight at 12 months of age (Y_4) and weight gain from weaning to 12 months of age (Y_5) were estimated using data from Canchim cattle. The multivariate restricted maximum likelihood (REML) with equal design matrices was applied to a mixed sire model for estimating variances and covariances. The sire model included, in addition to the mean, the fixed effects of sex, year, year of birth, season of birth, year x season interaction, age of dam and random effects of sires. Heritability estimates of the individual traits were

0.33, 0.77, 0.77, 0.44 and 0.10 for Y_1 to Y_5 , respectively. The canonical transformation was used to reduce the multivariate REML analysis to univariate analysis. The first canonical variable explained most genetic variation and had a heritability of 0.83, considerably higher than the original traits. This methodology was useful for assessing the relative importance of the individual traits and for choosing the subset of traits that selects animals more efficiently for genetic merit.

Keywords: canonical transformation, cattle, growth traits, variance components.

INTRODUÇÃO

O método da Máxima Verossimilhança Restrita (REML) é o preferido no melhoramento animal e disciplinas relacionadas com partição da variância fenotípica em genética e outros componentes. O REML possui várias propriedades desejáveis (HARVILLE, 1977), quando as observações têm distribuição normal, pois, sob certas condições, minimizam os vieses das estimativas de variância e covariâncias decorrentes de acasalamento seletivo, seleção através de gerações e descartes sequenciais.

Em virtude de sua natureza iterativa e da necessidade de inversões de matrizes grandes em cada iteração, a aplicação do REML é trabalhosa. Mesmo com o desenvolvimento de numerosos algoritmos especializados que exploram características específicas da estrutura dos dados ou do modelo de análise, estimar variâncias

e covariâncias de caracteres múltiplos é ainda uma tarefa difícil.

Entretanto, na classe de modelos multivariados, em que a matriz de desenho é a mesma para todas as variáveis e existe apenas um fator aleatório além do efeito residual, o procedimento da transformação canônica (TAYLOR et al. 1985, SCHAEFFER, 1986; JENSEN & MAO, 1988) reduz uma análise por REML de "q" características para "q" análises univariadas correspondentes, sem prejuízo das variáveis originais. No caso de características altamente correlacionadas, permite ainda medir a importância relativa de cada uma delas, tornando-se ainda uma alternativa útil para avaliar e selecionar animais pelo mérito genético.

O propósito deste trabalho é aplicar a técnica da transformação canônica a dados de características de crescimento em bovinos.

MATERIAL E MÉTODOS

Obtenção dos dados

Foram analisados dados do peso ao nascimento (Y_1), pesos observado (Y_2) e ajustado para 240 dias (Y_3), peso ajustado para doze meses (Y_4) e ganho de peso da desmama aos doze meses de idade (Y_5) de animais da raça Canchim,

filhos de 52 touros, nascidos e desmamados de 1982 a 1988, na Fazenda Canchim, município de São Carlos, SP, base física do Centro de Pesquisa de Pecuária do Sudeste. Foram consideradas duas épocas de nascimento (época 1: maio a julho e época 2: agosto a dezembro) e nove classes de idade da vaca ao parto, em anos, sendo ≤ 4 anos e ≥ 11 anos as extremas.

Aspectos teóricos

O modelo linear misto adotado para q características é da forma $Y = (I \cdot X)b + (I \cdot Z)u + e$, [1] em que, para i-ésima característica ($i = 1, \dots, q$), tem-se:

y = vetor de dados ($N \times 1$);

X = matriz de valores conhecidos ($N \times p$) de posto $r \leq \min(N, p)$, que associa os elementos de b ao vetor de dados y ;

$b = (b'_1, b'_2, \dots, b'_q)'$, b_i é um vetor ($p \times 1$) de efeitos fixos: média, sexo, ano de nascimento (AN), época, AN x época e idade da vaca ao parto;

Z = matriz de elementos fixos ($N \times q$), que associa os elementos de u aos elementos de y ;

$u = (u'_1, u'_2, \dots, u'_q)'$, u_i é um vetor ($t \times 1$) de efeitos aleatórios (t = número de touros);

I = matriz identidade de ordem q;

$e = (e'_1, e'_2, \dots, e'_q)'$, e_i é um vetor

As Equações de Modelos Mistos - EMM (HENDERSON, 1973), para estimar os efeitos aleatórios após a absorção dos efeitos fixos, são:

$$\begin{bmatrix} r^{11}Z'MZ + g^{11}I & r^{12}Z'MZ + g^{12}I & \dots & r^{1t}Z'MZ + g^{1q}I \\ r^{21}Z'MZ + g^{21}I & r^{22}Z'MZ + g^{22}I & \dots & r^{2t}Z'MZ + g^{2q}I \\ \dots & \dots & \dots & \dots \\ r^{q1}Z'MZ + g^{q1}I & r^{q2}Z'MZ + g^{q2}I & \dots & r^{qq}Z'MZ + g^{qq}I \end{bmatrix} \begin{bmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \dots \\ \hat{u}_q \end{bmatrix}$$

$$= \begin{bmatrix} r^{11}Z'MY_1 + r^{12}Z'MY_2 & \dots & r^{1q}Z'MY_q \\ r^{21}Z'MY_1 + r^{22}Z'MY_2 & \dots & r^{2q}Z'MY_q \\ \dots & \dots & \dots \\ r^{q1}Z'MY_1 + r^{q2}Z'MY_2 & \dots & r^{qq}Z'MY_q \end{bmatrix} \quad [2]$$

Nas EMM, M é uma matriz de projeção obtida de $(I - X(X'X)^{-1}X')$ (SEARLE et al. 1992), e tem as propriedades (SEARLE, 1971) $M = M^2$, $M = M'$, $M'M = M$; g^{ij} e r^{ij} , são os elementos das matrizes inversas de G e R , que contém as variâncias e covariâncias genéticas e residuais entre as características i e j , respectivamente, definidas por:

$$G = \begin{bmatrix} g_{11} & g_{12} & \dots & g_{1q} \\ g_{12} & g_{22} & \dots & g_{2q} \\ g_{1q} & g_{q2} & \dots & g_{qq} \end{bmatrix} \quad R = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1q} \\ r_{12} & r_{22} & \dots & r_{2q} \\ r_{1q} & r_{q2} & \dots & r_{qq} \end{bmatrix} \quad [3]$$

($N \times 1$) de resíduos aleatórios.

Transformação canônica

Seja a j -ésima observação do indivíduo representada por um vetor y_j ($q \times 1$), tem-se $\text{Var}(y_j) = P = G + R$, em que P é a matriz de variância-covariância fenotípica. Aplicando-se uma transformação linear em y_j , através de uma matriz Q apropriada, tal que $y_{cj} = Qy_j$, " c " denota escala canônica (JENSEN & MAO, 1988), então: $\text{Var}(y_{cj}) = Q(P)Q' = Q(G + R)Q' = QGQ' + QRQ' = G_c + R_c$ [4].

A matriz Q foi obtida de acordo com

SCHAFFER (1986) e foi escolhida de tal forma que G_c seja diagonal e R_c seja uma matriz identidade I_q . Assim, y_{cj} são denominadas variáveis canônicas, têm variância residual um e são não correlacionadas.

Considerando [4], tem-se covariâncias nulas e as EMM em [2] reduzem a q blocos diagonais, com o i -ésimo bloco dado por:

$$[Z'MZ + \lambda_i^{-1} * I_q] u_{ci} = Z'My_{ci}, \quad [5]$$

em que λ_i é o i -ésimo elemento diagonal de G_c e

$$y_{ci} = \sum_{m=1}^q q_{im} y_m \quad (i = 1, \dots, q)$$

Fazendo-se $C_c = (Z'MZ + \lambda_i^{-1} * I_q)^{-1}$ as estimativas dos elementos de G_c na $(k + 1)$ -ésima iteração são dadas por:

$$\hat{g}_{cii}(k+1) = (\hat{u}_{ci}(k) \hat{u}_{ci}(k) + \text{tr} C_c) / t \quad [6]$$

$$\hat{g}_{cij}(k+1) = (\hat{u}_{ci}(k) \hat{u}_{cj}(k)) / t \quad (i \neq j)$$

As estimativas dos elementos de R_c são dadas por:

$$\hat{r}_{cii}(k+1) = [\hat{e}'_{ci}(k) \hat{e}_{ci}(k) + \text{posto}(X) + q - \lambda_i^{-1} \text{tr} C_c] / n \quad [7]$$

$$\hat{r}_{cij}(k+1) = [\hat{e}'_{ci}(k) \hat{e}_{cj}(k)] / n \quad (i \neq j)$$

$$\hat{e}'_{ci}(k) \hat{e}_{cj}(k) = Y'_{ci}(k) MY_{cj}(k) - u'_{ci}(k) Z' MY_{cj}(k) - \lambda_i^{-1} u'_{ci}(k) u_{cj}(k)$$

Em cada iteração, as estimativas $\hat{G}$ e $\hat{R}$ são usadas para a obtenção de novas matrizes G_c e R_c , e as etapas de [2] até [7] continuam até que o critério de convergência seja aceito. Após a estimação de variâncias e covariâncias na escala canônica, G_c e R_c são retransformadas para a escala original conforme segue:

$$\hat{G}_{(k+1)} = Q^{-1} G_c Q^{-T}, \hat{R}_{(k+1)} = Q^{-1} R_c Q^{-T}$$

em que Q^{-T} é a transposta de Q^{-1} .

As estimativas iniciais $\hat{G}$ e $\hat{R}$ foram obtidas dos dados originais através do procedimento VARCOMP do SAS (BARR et al. página 817).

RESULTADOS E DISCUSSÃO

As estimativas de g_{ij} e r_{ij} estão nos Quadros 1 e 2, respectivamente. A convergência por REML foi assumida com trinta iterações, quando ocorreram mudanças apenas no quarto dígito.

Comparadas aos valores iniciais obtidos pelo procedimento VARCOMP do SAS, as estimativas de g_{ij} , obtidas por REML, foram maiores em quase todas as variáveis, enquanto as estimativas de r_{ij} foram praticamente similares, resultados que parecem confirmar a tendência de se obterem estimativas maiores de parâmetros genéticos por esse método. FREITAS (1993), utilizando dados semelhantes ao do presente trabalho, verificou maior discrepância nas variâncias residuais em relação às variâncias genéticas, quando comparou entre si o método IHSM e método 3, de HENDERSON.

As covariâncias genéticas e residuais das variáveis Y_1 , Y_2 e Y_3 com Y_5 foram negativas, possivelmente em razão dos efeitos de ganhos compensatórios da desmama aos doze meses. O Quadro 3 mostra as estimativas de herdabilidade (h^2) das

QUADRO 1 - Estimativas de variâncias e covariâncias genéticas (kg^2) de pesos ao nascimento (Y_1), observado (y_2) e ajustado (y_3) à desmama, aos doze meses (y_4) e ganho de peso, da desmama aos doze meses (y_5), de bovinos

	y_1	y_2	y_3	y_4	y_5
y_1	2,63106 ^a	4,83679	4,90497	3,31687	-0,01252
	2,66157 ^b	4,42257	4,58374	3,12921	
	- 0,01205				
y_2		136,80504	131,47207	116,42563	-0,11080
		152,06930	146,92242	127,57570	-0,14820
y_3			129,15749	110,27522	-0,14377
			144,11011	121,93465	-0,17058
y_4				104,57170	-0,04207
				111,58864	-0,07665
y_5					0,00080
					0,00074

^a Estimativas iniciais obtidas pelo VARCOMP, do SAS (BARR et al. 1985, página 817).

^b Estimativas obtidas, com 30 iterações, por REML.

QUADRO 2 - Estimativas de variâncias e covariâncias residuais (kg^2) de pesos, ao nascimento (Y_1), observado (Y_2) e ajustado (Y_3), à desmama, aos doze meses (Y_4), e ganho de peso, da desmama aos doze meses (Y_5), de bovinos

	Y_1	Y_2	Y_3	Y_4	Y_5
Y_1	28,226657 ^a 28,216970 ^b	32,904130 32,965783	36,908291 37,119099	28,804167 28,933677	-0,063883 -0,064786
Y_2		634,037945 633,769402	564,979814 563,393647	466,809052 464,021463	-0,777385 -0,785643
Y_3			604,123303 602,430958	502,516928 503,064464	-0,825609 -0,804843
Y_4				890,889034 898,631677	3,087181 3,143333
Y_5					0,031353 0,031603

^a Estimativas iniciais obtidas pelo VARCOMP, do SAS (BARR et al. 1985, página 817)

^b Estimativas obtidas, com 30 iterações, por REML.

características individuais. Os pesos à desmama, observado (Y_2) e ajustado (Y_3), apresentaram valores de h^2 de 0,77 e correlações fenotípicas em torno de 1,0, indicando, para esses dados, ser desnecessário o ajuste linear.

As raízes características, em ordem decrescente, foram $\hat{\lambda}_1 = 0,26187$, $\hat{\lambda}_2 = 0,09600$, $\hat{\lambda}_3 = 0,03633$, $\hat{\lambda}_4 = 0,01885$ e $\hat{\lambda}_5 = 0,00368$. A primeira variável canônica $\hat{y}_{1c}$, com uma herdabilidade de $4 * \hat{\lambda}_1 / (\hat{\lambda}_1 + 1) = 4 * (0,26187 / 1,26187) = 0,84$, representa o componente que explica a maioria da variação genética (62,84%), que é, essencialmente, a diferença entre peso à desmama e as demais características de crescimento. Essa herdabilidade é, consideravelmente, maior do que a obtida para as variáveis individuais (Quadro 3). Conseqüentemente, a segunda variável canônica, que contém a maioria da informação genética

(23,04%) após Y_{1c} , apresenta herdabilidade de 0,35, que, de certa forma, expressa o contraste do peso à desmama com as demais características de crescimento. Assim, a técnica do componente principal reduziu o espaço de cinco dimensões (Y_1 até Y_5) para duas dimensões (Y_{1c} , Y_{2c}), o que explica 86,0% da variação total, fornecendo, portanto, bom resumo dos dados.

O uso de componentes principais por meio da matriz de correlação fenotípica entre as variáveis também fornece resultados interessantes (Quadro 5).

Os dois primeiros componentes principais, $y_{1c} = 0,30037 Y_1 + 0,56192 Y_2 + 0,55084 Y_3 + 0,53114 Y_4 + 0,09213 Y_5$ e $y_{2c} = -0,17766 Y_1 - 0,09703 Y_2 - 0,20726 Y_3 + 0,25821 Y_4 + 0,92162 Y_5$, explicam 57,0% e 26,0% da variância total, respectivamente.

QUADRO 3 - Estimativas* de herdabilidade e de correlações de pesos, ao nascimento (Y_1), observado (Y_2) e ajustado (Y_3), à desmama, aos doze meses (Y_4), e ganho de peso da desmama, aos doze meses (Y_5), de bovinos

	Y_1	Y_2	Y_3	Y_4	Y_5
y_1	0,34	0,22	0,23	0,18	-0,27
y_2	0,35	0,77	0,99	0,98	-0,44
y_3	0,36	0,94	0,77	0,96	-0,52
y_4	0,31	0,80	0,76	0,44	-0,26
Y_5	0,01	0,04	-0,07	0,34	0,09

* herdabilidade na diagonal, correlações genética acima e correlações fenotípica abaixo.

Assim, pode-se reduzir o espaço de cinco dimensões (Y_1 a Y_5) para um de duas dimensões (y_{1c} e y_{2c}), o que explica 83,0% da variação total, ou seja, em estudos semelhantes de cinco variáveis, é possível, por meio dessa metodologia, considerar apenas dois componentes principais como explicativos do fenômeno.

Assim, o uso da transformação canônica e componentes principais permite reduzir o número de características para selecionar o animal. Por exemplo, investigando-se o primeiro componente, verifica-se que

as variáveis Y_2 , Y_3 e Y_4 têm pesos semelhantes (0,56192, 0,55084 0,53114). Para analisar menos variáveis sem perder a eficiência, podem-se escolher os pesos à desmama, ajustado para 240 dias (Y_3) e aos doze meses (Y_4). Como essas variáveis são correlacionadas, asseguram-se, também, maior peso ao nascimento e maior ganho diário, da desmama aos doze meses.

CONCLUSÕES

As técnicas da transformação canônica e de componentes principais

QUADRO 4 - Percentagens de variação, estimativas de componentes principais e de herdabilidades, de pesos, ao nascimento (Y_1), observado (Y_2) e ajustado (Y_3), à desmama, aos doze meses (Y_4) e ganho de peso da desmama, aos doze meses (Y_5), de bovinos

Compo- nente	% de Variação	Y_1	Y_2	Y_3	Y_4	Y_5	Herda- didade
$\hat{y}_{c1}$	62,88	-22,42779	-84,822826	38,148599	-15,209845	-13,226949	0,84
$\hat{y}_{c2}$	23,02	0,21012	0,656797	-0,331164	0,119180	0,104011	0,36
$\hat{y}_{c3}$	8,77	-0,14093	-0,657406	0,403355	-0,128051	-0,129624	0,16
$\hat{y}_{c4}$	4,53	-0,05813	0,002678	-0,075712	0,007105	-0,017145	0,08
$\hat{y}_{c5}$	0,80	0,002325	-0,034051	-0,001577	0,188440	0,044500	0,00

QUADRO 5 - Percentagens de variação e coeficientes dos componentes principais, utilizando a matriz de correlação fenotípica de pesos, ao nascimento (Y_1), observado (Y_2) e ajustado (Y_3), à desmama, aos doze meses (Y_4), e ganho de peso da desmama, aos doze meses (Y_5), de bovinos

Compo- nente	% de Variação	Y_1	Y_2	Y_3	Y_4	Y_5
$\hat{y}_{1c}$	57,00	0,30037	0,56192	0,55084	0,53114	0,09213
$\hat{y}_{2c}$	26,00	-0,17766	-0,09703	-0,20726	0,25821	0,92162
$\hat{y}_{3c}$	15,00	-0,93693	0,20388	0,20041	0,13265	-0,15125
$\hat{y}_{4c}$	1,00	-0,01381	0,35154	0,35942	-0,79547	0,33804
$\hat{y}_{5c}$	1,00	-0,01369	-0,71395	0,69589	0,02913	0,07053

foram utilizadas para estimar parâmetros genéticos de características de crescimento de animais Canchim. Essas técnicas permitiram medir a importância relativa de cada uma delas e reduzir o número de características, para selecionar os animais em razão de seu mérito genético.

REFERÊNCIAS BIBLIOGRÁFICAS

1. FREITAS, A.R. de, PEGORIN, M.J., ALENCAR, M.M. de, FAVORETTI, A.C. Estimativas de parâmetros genéticos de múltiplas características através do método Simples de Henderson. *Pesq. Agropec. Bras.*, Brasília, DF, (no prelo).
2. HENDERSON, C.R. Sire evaluation and genetics trends. In: ANIMAL BREEDING AND GENETICS SYMPOSIUM IN HONOR OF DR. JAY LUSH. Champaign, 1973. *Proceedings...* Blacksburg: American Society of Animal Science, 1973. p. 10-41.
3. JENSEN, J., MAO, I.L. Transformation algorithms in analysis of single trait and of multitrait models with equal design matrices and one random factor per trait: a review. *J. Anim. Sci.*, Champaign. v. 66, n. 11, p. 2750-2761, Nov. 1988.
4. MEYER, K. Maximum likelihood estimation of variance components for a multivariate mixed model with equal design matrices. *Biometrics*, Washington, v. 41, n. 1, p. 153-166, Mar. 1985.
5. STATISTICAL ANALYSIS SYSTEM INSTITUTE. The varcomp procedure: operating systems: All. In: _____ *User's guide: statistics*. 5.ed. Cary: 1985. cap. 41, p. 817-823.
6. SCHAEFFER, L.R. Estimation of variances and covariances within the allowable parameter space. *J. Dairy Sci.*, Champaign. v. 69, n. 1, p. 187-94, Jan. 1986.
7. SCHAEFFER, L.R., WILTON, J.W., THOMPSON, R. Simultaneous estimation of variance and covariance components from multitrait mixed model equations. *Biometrics*, Washington, v. 34, n. 2, p. 199-208, Jun. 1978.
8. SEARLE, S.R., CASELLA, G., McCULLOCH, C.E. *Variance components*. New York: John Wiley, 1992. 528p.
9. TAYLOR, J.F., BEAN, B., MARSHALL, C.E. SULLIVAN, J.J. Genetic and environmental components of semen production traits of artificial insemination Holstein bulls. *J. Dairy Sci.*, Champaign. vo. 68, n. 10, p. 2703-27, Mar. 1985.
10. THOMPSON, R. Estimation of quantitative genetic parameters. In: INTERNATIONAL CONFERENCE ON QUANTITATIVE GENETICS, Ames, 1977. *Proceedings...* Ames: Ohio State University Press, 1977. p. 639-57.