

## Avaliação de Procedimentos na Estimação de Parâmetros Genéticos em Bovinos de Corte

Alfredo Ribeiro de Freitas<sup>1</sup>

**RESUMO** - Estimativas de variâncias, covariâncias e herdabilidades ( $h^2$ ) de dados de pesos ao nascimento ( $y_1$ ), à desmama ( $y_2$ ), aos 12 ( $y_3$ ), 18 ( $y_4$ ) e 24 meses ( $y_5$ ) e circunferência escrotal aos 12 ( $y_6$ ) meses de idade foram obtidas de três amostras de dados de animais Canchim, machos e fêmeas, nascidos de 1961 a 1991, em São Carlos, SP. Foram utilizadas análises univariada, por meio do Método 3 de Henderson, Máxima Verossimilhança (ML), ML Restrita (REML), Método Iterativo Simples de Henderson (IHSM), Método da Estimação Quadrática Não-viesada de Variância Mínima (MIVQUEo) e procedimento GLM do SAS, e multivariada por IHSM, ML e REML, com e sem a inclusão da matriz de parentesco entre os animais. As estimativas de  $h^2$  variaram de 0,33 a 0,34 ( $y_1$ ), 0,30 a 0,81 ( $y_2$ ), 0,28 a 0,68 ( $y_3$ ), 0,27 a 0,27 ( $y_4$ ), 0,40 a 0,40 ( $y_5$ ) e 0,40 a 0,40 ( $y_6$ ). Detectou-se afastamento da normalidade para quase todas as características e heterogeneidade de variâncias para os efeitos fixos de sexo, ano e época de nascimento, idade da vaca ao parto, geração do animal e efeito aleatório de touros. A transformação dos dados fornecendo escores normais como resposta foi a mais eficiente em aproximar os dados a uma distribuição normal. Métodos de estimação, análise univariada ou multivariada, amostragem e tipo de transformação de dados, nesta ordem, foram os que mais influenciaram as estimativas de  $h^2$ .

Palavras-chave: bovinos de corte, componentes de variância, métodos de estimação

### Procedures Evaluation in the Estimates of Genetic Parameters of Cattle

**ABSTRACT** - Estimates of variance, covariance and heritabilities ( $h^2$ ) for body weights at birth ( $y_1$ ), weaning ( $y_2$ ), 12 ( $y_3$ ), 18 ( $y_4$ ) and 24 ( $y_5$ ) months of age, and scrotal circumference at 12 months of age ( $y_6$ ) were obtained from three data sets of purebred Canchim males and females, born from 1961 to 1969 in São Carlos, São Paulo State. The genetic parameters were obtained by means of univariate analysis by Henderson Method 3, Maximum Likelihood (ML), Restricted ML (REML), Iterative Henderson Simple Method, Minimum Variance Quadratic Unbiased Estimation (MIVQUEo) and GLM procedure of SAS, and multivariate analysis by IHSM, ML and REML, with and without the relation matrix of animals. The heritability estimates ranged from 0.33 to 0.34 ( $y_1$ ), 0.30 to 0.81 ( $y_2$ ), 0.28 to 0.68 ( $y_3$ ), 0.27 to 0.27 ( $y_4$ ), 0.40 to 0.40 ( $y_5$ ) and 0.40 to 0.40 ( $y_6$ ). No normality of data for almost all traits and heterogeneity of variance for fixed effects of gender, year and season of birth, age of dam at calving, animal generation and random effect of sire was detected. Data transformation into normal scores provided the most efficient way to bring data near to a normal distribution. Estimation methods, type of analysis (univariate or multivariate analysis), sampling and type of transformation procedure, in this order, were the factors that influenced the most the heritability estimates.

Key Words: beef cattle, estimation methods, variance components

### Introdução

Para planejar e orientar um programa de melhoramento genético de bovinos, é necessário estimar parâmetros genéticos, utilizando-se dados desbalanceados, sujeitos a grandes influências ambientais e de manejo.

A precisão desses estimadores é dependente de um conjunto de fatores, destacando-se o método de estimação, pois, embora o estado da arte seja estimar componentes de variâncias por Máxima Verossimilhança Restrita (REML) em um modelo animal e prever valores genéticos por meio do BLUP, isto nem sempre é possível e existem pelo menos outros 10 métodos de estimação em uso.

A não-normalidade dos dados, a assimetria e a curtose associadas aos dados também influenciam as inferências obtidas, sendo mais problemáticas em análises multivariadas (COCHRAN e COX, 1978; BROWNIE et al., 1990; e AZZALINI e DALLA VALLE, 1996). A heterogeneidade de variâncias, como problema correlato, quando não é devidamente estudada, pode implicar em predição viciada do valor genético, redução do progresso genético e seleção desproporcional de animais oriundos de ambientes com variâncias diferentes (RODRÍGUEZ-ALMEIDA et al., 1995a, b; MEUWISSEN et al., 1996).

Fatores como os algoritmos de estimação usando o REML, diferentes amostras de dados de um reba-

<sup>1</sup> Pesquisador da Embrapa Pecuária Sudeste, Caixa Postal 339, 13560-970 - São Carlos, SP. Bolsista do CNPq. E-mail: ribeiro@cnpq.embrapa.br

nho, tipo de transformação dos dados e de análise - univariada ou multivariada, idade do animal, sexo e raça, também, acarretam variações consideráveis nas estimativas de  $h^2$  e de correlações genéticas (RODRÍGUEZ-ALMEIDA et al., 1995a, b; MEYER e SMITH, 1996; e ROBERT e DUCROCQ, 1996).

O objetivo deste trabalho foi avaliar a magnitude da influência de tipo de análise (univariada versus multivariada), métodos e algoritmos de estimação, tipo de transformação de dados e diferentes amostragem de dados no rebanho sobre estimativas de variâncias, covariâncias e herdabilidades de características relacionadas ao desenvolvimento ponderal de animais da raça Canchim.

### Material e Métodos

#### *Obtenção dos dados*

Foram utilizadas três amostras de dados de animais da raça Canchim, nascidos na Embrapa/Centro de Pesquisa de Pecuária do Sudeste, município de São Carlos, SP: 2511 animais nascidos de 1969 a 1991 (amostra A); 1125 animais nascidos de 1982 a 1988 (amostra B); e 890 animais nascidos de 1982 a 1990 (amostra C). Foram analisados os dados de pesos ao nascimento ( $y_1$ ), à desmama ( $y_2$ ), aos 12 ( $y_3$ ), 18 ( $y_4$ ) e 24 meses ( $y_5$ ) e a circunferência escrotal aos doze meses de idade ( $y_6$ ). Foram considerados os efeitos de sexo (S), ano e época de nascimento (época 1, de maio a julho, e época 2, de agosto a dezembro), idade da vaca ao parto (IV: sendo  $\leq 4$  e  $\geq 11$  anos, as extremas), geração do animal e efeito aleatório de touros.

#### *Metodologia de análise*

##### *Estatísticas descritivas*

Foram utilizados a estatística W de Shapiro-Wilks (SAS, 1993), para verificar se os valores observados para cada característica seguiam distribuição normal, e o teste qui-quadrado de Bartlett, para testar a hipótese de homogeneidade de variâncias dentro dos níveis dos efeitos fixos e aleatórios. Os dados de cada característica foram analisados em quatro escalas: original ( $y_{obs}$ ), logarítmica ( $y_{log}$ ), padronizada  $y_{zi}=(y_i-\bar{y})/s(y_z)$  e a de escores normais de BLOM ( $y_{blom}$ ) do procedimento RANK do SAS (SAS, 1996). Essa transformação é adequada a uma variável aleatória  $y_i$  de distribuição não normal e assimétrica com média  $\bar{y}$  e variância  $\sigma^2$ , fornecendo como resposta escores normais esperados para uma distribuição normal.

#### *Estimativas dos parâmetros genéticos*

As estimativas dos parâmetros genéticos foram obtidas de análises uni e multivariada por meio dos métodos da Máxima Verossimilhança - ML (HARTLEY e RAO, 1967), ML Restrita - REML (PATTERSON e THOMPSON, 1971), Método 3 (3H) e Método Iterativo Simples de Henderson (IHSM), ambos descritos em HENDERSON (1984) e Método da Estimação Quadrática Não-viesada de Variância Mínima (MIVQUEo), "default" do procedimento VARCOMP do SAS (SAS, 1993). Para a convergência das variâncias e covariâncias por ML e REML, foram usados o algoritmo de Newton-Raphson-NR (JENRICH e SAMPSON, 1976) e o EM ("Expectation-Maximization") de DEMPSTER et al., (1977), para análises univariada e multivariada, respectivamente. Esses algoritmos são iterativos e requerem derivadas, esperanças e inversas da função de verossimilhança. Os valores iniciais das variâncias e covariâncias genéticas e residuais foram obtidos por meio do procedimento GLM do SAS, que utiliza o método dos quadrados mínimos (SAS, 1993).

Nas análises univariadas, foram utilizados os métodos 3H, IHSM, ML, REML (HENDERSON, 1984; FREITAS e VENCOVSKY 1993b) e MIVQUE(o). Para as análises multivariada com duas características e inclusão da matriz de parentesco entre os animais, foram utilizados IHSM, ML e REML (FREITAS e VENCOVSKY, 1993a); com três características, foram utilizados o REML em transformação canônica (JENSEN e MAO, 1988; FREITAS et al., 1994) e o IHSM conforme FREITAS et al. (1995).

O modelo linear misto para cada situação foi formulado considerando os efeitos fixos de sexo, ano e época de nascimento, idade da vaca ao parto, geração e época de nascimento e aleatório de touros.

### Resultados e Discussão

A amplitude dos valores das estimativas de assimetria e curtose dos pesos ao nascimento ( $y_1$ ), à desmama ( $y_2$ ), aos 12 ( $y_3$ ), 18 ( $y_4$ ) e 24 ( $y_5$ ) meses de idade e circunferência escrotal ( $y_6$ ) aos 12 meses de idade estão na Tabela 1. Rejeitou-se ( $P<0,01$ ), para quase todas as características e amostra, a hipótese de que os dados são provenientes de uma população com distribuição normal, com a assimetria variando de -0,16 ( $y_3$ ) a 0,96 ( $y_5$ ), e a curtose de -0,66 ( $y_2$ ) a 1,98 ( $y_4$ ). Apesar de a pressuposição de normalidade dos dados ser crucial para o procedi-

Tabela 1 - Amplitude dos valores dos coeficientes de assimetria e curtose obtida dos pesos ao nascimento ( $y_1$ ), à desmama ( $y_2$ ), aos 12 ( $y_3$ ), 18 ( $y_4$ ) e 24 meses de idade ( $y_5$ ) e circunferência escrotal ( $y_6$ ) aos 12 meses de idade de animais Canchim

Table 1 - Range of values of skewness and kurtosis coefficients obtained from body weights at birth ( $y_1$ ), weaning ( $y_2$ ), 12 ( $y_3$ ), 18 ( $y_4$ ) and 24 ( $y_5$ ) months of age and scrotal circumference ( $y_6$ ) at 12 months of age of Canchim herd

Caracter	Assimetria			Curtose		
Trait	Skewness			Kurtosis		
$y_1$	0,14	a	0,30*	0,11	a	0,12*
$y_2$	-0,11	a	0,30*	-0,66*	a	1,19*
$y_3$	-0,16	a	0,49*	0,09	a	1,14*
$y_4$	0,38*	a	0,73*	1,30*	a	1,98*
$y_5$	0,96*	a	0,96*	1,85*	a	1,85*
$y_6$	0,40*	a	0,40*	-0,07	a	-0,07*

\* Significativo ( $P < 0,01$ ) pela estatística W de Shapiro Wilks.

\* Significant ( $P < 0,01$ ) by W statistic of Shapiro Wilks.

mento de estimação por ML e REML, a literatura não é informativa sobre como o não atendimento desta exigência influencia as estimativas de parâmetros genéticos no melhoramento animal. Na análise de variância, entretanto, há evidências de que os valores da curtose e, em menor extensão, da assimetria dos dados (SCHEFFÉ, 1959) são os indicadores mais importantes de como os desvios de normalidade dos dados influenciam as inferências obtidas. Segundo COCHRAN e COX (1978), a não-normalidade dos dados pode influir negativamente na estimação dos efeitos fixos, no uso dos testes t e F e na heterogeneidade da variância do erro. Entretanto, como a estimação dos componentes de variâncias e covariâncias geralmente é feita por REML, cuja função de verossimilhança é construída com base em contrastes de resíduos, após o ajuste dos efeitos fixos do modelo misto (DAVIDIAN e GILTINAN, 1996), quaisquer deficiências desses ajustes irão interferir nas estimativas por REML.

Observando-se ainda a Tabela 1, pode-se inferir que a qualidade dos dados decresce (maiores coeficientes de assimetria e de curtose) com a idade. Segundo BROWNIE et al. (1990) e DAVIDIAN e GILTINAN (1996), em estudos de crescimento, é comum observar variabilidade crescente na resposta dos indivíduos em função do tempo, denominada de “fenômeno da inflação de variância”. Para dados de desenvolvimento ponderal de bovinos, o aumento da variância entre indivíduos, com a idade, pode ser explicado também pela maior dificuldade de se pesar o animal adulto e pelo acúmulo de influências

ambientais. Esses resultados sugerem maior rigor dos métodos de análise de dados de desenvolvimento ponderal de bovinos à idade adulta.

Com exceção de algumas características e amostra, houve heterogeneidade de variâncias ( $P < 0,05$ ), pelo teste Bartlett, para os efeitos estudados (Tabela 2). Observou-se heterogeneidade de variâncias crescente na resposta dos animais com a idade, fato associado ao fenômeno da “inflação de variância” e também quando se aumenta o número de níveis de um efeito (Tabela 1). Esse último pode ser explicado pelo maior desbalanceamento no número de repetições e nos efeitos ambientais associados a este.

Os problemas decorrentes da heterogeneidade de variâncias sobre as propriedades dos testes t e F da ANOVA são enormes e têm sido discutidos em vários trabalhos. Uma abordagem interessante sobre o assunto pode ser vista em BROWNIE et al. (1990). No melhoramento animal, quando a heterogeneidade de variância dos efeitos do modelo não é devidamente estudada, pode ocorrer predição viciada do valor genético e, conseqüentemente, das DEPs dos animais (RODRIGUEZ-ALMEIDA et al., 1995a, 1995b). Segundo MEUWISSEN et al. (1996), apenas com a correção da heterogeneidade de variâncias dentro de rebanhos em dados de produção de leite o viés dos valores genéticos reduziu em cerca de 19%. Redução do progresso genético e ainda seleção desproporcional de animais oriundos de ambientes com variâncias diferentes foram encontradas por vários autores (BROTHERSTONE e HILL, 1986; DE VEER e VAN VLECK, 1987; BOLDMAN e FREEMAN, 1990; GIANOLA et al., 1992; WEIGEL e GIANOLA, 1992; COOK e WEISBERG, 1994; BAXTER, 1995; e RAMOS et al., 1996). Este fato comprova a hipótese de homogeneidade de variância por meio dos efeitos fixos, assumida na REML, em que o método mais utilizado para estimação de variâncias e covariâncias, nem sempre é atendida. Assumir esta hipótese na estrutura de dispersão dos dados, quando a mesma não é atendida, implica em prejuízos na resposta à seleção dos animais (GIANOLA et al., 1992). A situação é mais complicada, quando se trata de análise multivariada (WEIGEL e GIANOLA, 1992), pois, não raro, heterogeneidade de variâncias também significa heterogeneidade da matriz de covariâncias, e o impacto deste fato sobre o progresso genético esperado e realizado é mais importante que no caso univariado. Dos procedimentos utilizados para minimizar os efeitos da heterogeneidade de variâncias nas estimativas de parâmetros genéticos,

Tabela 2 - Razão entre a maior e a menor variância de cada efeito, dos pesos ao nascimento ( $y_1$ ), à desmama ( $y_2$ ), aos 12 meses ( $y_3$ ), 18 ( $y_4$ ) e ( $y_5$ ) 24 meses de idade e circunferência escrotal ( $y_6$ ) aos 12 meses de idade de animais Canchim  
 Table 2 - Ratio between the highest and lowest variance within each effect from body weights at birth ( $y_1$ ), weaning ( $y_2$ ), 12 ( $y_3$ ), 18 ( $y_4$ ) and 24 months of age ( $y_5$ ) and scrotal circumference ( $y_6$ ) at 12 months of age of Canchim herd

Efeito <i>Effect</i>	$y_1$	$y_2$	$y_3$	$y_4$	$y_5$	$y_6$
Sexo <i>Sex</i>	1,3 a 1,3	1,0-1,4	1,2 a 1,6	1,4 a 2,0	2,3*a 2,3*	-
Geração <i>Generation</i>	1,0 a 1,2	1,4-1,6	1,3 a 1,3	1,4 a 1,7*	1,5 a 1,8*	
Idade da vaca ao parto <i>Age of dam at calving</i>	1,4 a 1,6*	1,1-2,0*	1,5 a 1,6*	1,4 a 1,9*	1,5* a 2,1*	4,2** a 4,2**
Ano <i>Year</i>	2,2* a 20,0**	8,8** a 4,1**	12,5**-18,1*	17,2** a 115,4**	15,4** a 150,0**	2,4* a 2,4**
Touro <i>Sire</i>	49,0** a 337,0**	345,3** a 1806,1**	453,9**-596,4**	31,6** a 188,1**	1484,5** a 750,0**	26,2** a 26,2**

\* ( $P < 0,05$ ) e \*\* ( $P < 0,01$ ), pelo teste de qui-quadrado de Bartlett para homogeneidade de variâncias.  
 ( $P < .05$ ) and \*\* ( $P < .01$ ), by chi-square test of Bartlett for variances homogeneity.

pode-se destacar o de VISSCHER (1992), que estuda esse efeito dentro de correlações intraclasse, e o de MEUWISSEN et al. (1996), que estimam valores genéticos e corrigem heterogeneidade de variância fenotípica dentro de rebanho-ano.

Para a estimação dos componentes de variâncias e covariâncias, usando-se os métodos ML e REML e a amostra de dados B (Tabela 3), foram necessárias oito iterações para  $y_1$  e  $y_2$ , nas análises univariadas por meio do algoritmo de NR e 40 iterações nas análises multivariada, por meio do EM.

A velocidade de convergência utilizando-se o NR foi similar à encontrado na literatura para ML e REML em análises univariadas e tamanhos de arquivos semelhantes (HEMMERLE e HARTLEY, 1973; TAYLOR e EVERETT, 1985). Embora esse algoritmo seja sensível a valores iniciais pobres, permite localizar o valor máximo da função de verossimilhança com poucas iterações, quando o mesmo está próximo ao valor verdadeiro (HARVILLE, 1977).

O aumento no número de iterações nas análises multivariada justifica-se pelo fato de que a verossimilhança pelo EM incrementa monotonicamente em progressão geométrica, exigindo, porém, número grande de iterações (JENSEN e MAO, 1988; MEYER e SMITH, 1996; ROBERT e DUCROCQ, 1996; e GEOFFREY e KRISHNAN, 1997).

Para estimar os componentes de variâncias e covariâncias, usando-se o método IHSM e a amostra de dados B (Tabela 3), foram necessárias 6 e 13 iterações, respectivamente, para  $y_1$  e  $y_2$ , nas análises univariadas; com três características (Tabela 4),

convergir com 20 iterações. Devido à escassez de informações na literatura sobre o IHSM, não se pode compará-lo com o ML e o REML. Entretanto, segundo OUWELTJES et al., (1988), o mesmo parece convergir mais rapidamente entre os métodos que utilizam pseudoesperanças.

Segundo MEYER (1989, 1991), MEYER e SMITH (1996), o NR e o EM convergem mais rapidamente que os procedimentos livres de derivadas (DF) que usam aproximação numérica das derivadas e exploram a técnica da matriz esparsa. Segundo MISZTAL (1994), os métodos DF são fáceis de se usar, adequados para manejar grandes volumes de dados, como no caso do modelo animal. Entretanto, esses métodos são dependentes do número de características e dos valores iniciais das variâncias e covariâncias, além de apresentarem convergência lenta e manuseio difícil em análises multivariada.

MISZTAL (1994) mostrou que a técnica da matriz esparsa pode reduzir o tempo e requerimento de memória para o REML com o EM, tornando-o mais vantajoso que o DF, pois sempre converge para máximo global. Entretanto, quando se usa o REML em um modelo animal, o DF pode ser vantajoso, pois elimina o trabalho complexo das derivadas. GEOFFREY e KRISHNAN (1997) fazem duas críticas ao EM: a lentidão na convergência e o não-fornecimento da matriz de covariâncias no caso univariado.

Independentemente do algoritmo usado, mesmo que ML e REML proporcionem estimativas de parâmetros genéticos dentro do espaço de parâmetros, deve-se verificar se estas representam solução máxi-

Tabela 3 - Estimativas de covariâncias genética ( $g_{ij}$ ) e residual ( $r_{ij}$ ) e de herdabilidades ( $h^2_i$ ) associadas aos pesos ao nascimento ( $y_1$ ), à desmama ( $y_2$ ), aos 12 ( $y_3$ ), 18 ( $y_4$ ) e 24 meses de idade ( $y_5$ ) e circunferência escrotal ( $y_6$ ) aos 12 meses de idade de animais Canchim. (A: 2511 animais nascidos de 1961 a 1969; B: 1125 animais nascidos de 1982 a 1988; e C: 890 bezerros nascidos de 1982 a 1990)

Table 3 - Estimates of genetic ( $g_{ij}$ ) and residual ( $r_{ij}$ ) covariances and heritabilities ( $h^2_i$ ) from body weights at birth ( $y_{21}$ ), weaning ( $y_2$ ), 12 ( $y_3$ ) and 18 ( $y_4$ ) months of age of Canchim cattle (A: 2.511 animals born from 1961 to 1969; B: 1.125 animals born from 1982 to 1988 and C: 890 animals born from 1982 to 1990)

Método Method	$g_{22}(r_{22})$	$g_{23}(r_{23})$	$g_{33}(r_{33})$	$h^2_2$	$h^2_3$	Amostra Sample
Análise univariada Univariate analysis						
3H	84,84 (572,27)	47,15 (478,28)	123,78 (990,98)	0,52	0,44	A
MIVQUE(o)	59,02 (595,63)	26,17 (497,28)	63,71 (1045,34)	0,36	0,23	A
ML	88,15 (525,92)	43,66 (496,56)	111,75 (997,62)	0,57	0,40	A
REML	97,33 (575,44)	59,60 (484,15)	129,94 (1012,91)	0,58	0,45	A
3H	118,17 (594,80)	100,40 (493,77)	95,33 (886,35)	0,66	0,39	B
IHSM	124,77 (597,09)	104,98 (493,72)	85,49 (903,98)	0,69	0,34	B
MIVQUE(o)	86,25 (622,74)	69,20 (521,14)	67,66 (910,62)	0,48	0,28	B
ML	132,10 (585,28)	122,81 (485,11)	91,23 (877,64)	0,74	0,38	B
REML	144,50 (595,99)	134,66 (521,14)	106,85 (892,32)	0,78	0,43	B
3H	164,06 (792,34)	188,60 (606,70)	239,04 (1173,19)	0,68	0,68	C
REML	$g_{11}(r_{11})$ 2,43 (27,36)	$g_{14}(r_{14})$ -1,67 (41,08)	$g_{44}(r_{44})$ 96,47 (1349,11)	$h^2_1$ 0,33	$h^2_4$ 0,27	A
Análise multivariada Multivariate analysis						
	$g_{22}(r_{22})$	$g_{23}(r_{23})$	$g_{33}(r_{33})$	$h^2_2$	$h^2_3$	
IHSM	148,26 (584,21)	121,75 (486,26)	118,50 (876,54)	0,81	0,48	B
ML	128,15 (578,44)	110,07 (481,30)	94,54 (871,97)	0,72	0,39	B
REML	131,18 (627,80)	110,55 (515,15)	93,19 (914,18)	0,69	0,37	B
REML <sub>c</sub>	144,11 (602,43)	121,93 (503,06)	111,59 (898,63)	0,77	0,44	B

REML<sub>c</sub> = REML e transformação canônica.

REML<sub>c</sub> = REML and canonical transformation.

Tabela 4 - Estimativas de covariâncias genética ( $g_{ij}$ ) e residual ( $r_{ij}$ ) e de herdabilidades ( $h^2_i$ ) obtidas dos pesos à desmama ( $y_2$ ), aos 12 ( $y_3$ ) meses e circunferência escrotal ( $y_6$ ) aos 12 meses de idade de animais Canchim obtidos de 890 animais nascidos de 1982 a 1990 (C)

Table 4 - Estimates of genetic ( $g_{ij}$ ) and residual ( $r_{ij}$ ) covariances from body weights at weaning ( $y_2$ ), and at 12 months ( $y_3$ ) and scrotal circumference ( $y_6$ ) at 12 months of age of Canchim cattle obtained from 890 animals born from 1982 to 1990 (C)

$g_{22}/r_{22}$	$g_{23}/r_{23}$	$g_{33}/r_{33}$	$g_{26}/r_{26}$	$g_{36}/r_{36}$	$G_{66}/r_{66}$	$h^2_2$	$h^2_3$	$h^2_6$
167,83	192,07	278,82	9,99	14,28	0,94	0,64	0,63	0,40
882,14	727,23	1485,48	23,52	58,33	8,50			

$g_{ij}/r_{ij}$  Covariância genética na primeira linha e residual na segunda linha.

$g_{ij}/r_{ij}$  Genetic covariances in the first row and residual in the second row.

ma local ou global (MISZTAL, 1994; BOLDMAN et al., 1995). No caso de dados desbalanceados, tem-se certa garantia de solução máxima global, quando os critérios abaixo forem utilizados, proporcionando a mesma solução: usar diferentes valores iniciais; utilizar valores abaixo e acima dos obtidos na primeira convergência; e usar dois ou três dígitos significativos para a  $h^2$  (TAYLOR e EVERETT, 1985). Além desses, BOLDMAN et al. (1995) recomendam ainda: não utilizar covariâncias entre característica cuja correlação seja próxima de 1; dividir o valor da

característica pelo respectivo desvio-padrão para dados que diferem grandemente em escala; no caso de variâncias próximas de zero, multiplicar os dados por um fator a ser escolhido de modo a obter convergência e, em seguida, retransformar as estimativas para as escalas originais.

As estimativas de variâncias e covariâncias genéticas ( $g_{ij}$ ), residuais ( $r_{ij}$ ) entre as características  $i$  e  $j$ , e de herdabilidades ( $h^2_i$ ), associadas aos pesos ao nascimento ( $y_1$ ), desmama ( $y_2$ ), 12 ( $y_3$ ), 18 ( $y_4$ ) e 24 meses ( $y_5$ ) e circunferência escrotal aos 12 ( $y_6$ )

meses de idade estão nas Tabela 3 e 4.

Na análise univariada (Tabela 3), observa-se variabilidade nas estimativas de variâncias e covariâncias e de  $h^2$  entre método de estimação e amostra de dados. Nas amostras A e B, valores crescentes de  $g_{ij}$  são obtidos por MIVQUE(o), ML e REML, nessa ordem, refletindo diferenças entre os métodos para estimar  $g_{ij}$ . Comparado ao REML, que possui propriedades mais desejáveis para estimar parâmetros genéticos, as estimativas obtidas por 3H são subestimadas. Este método somente é adequado quando se considera uma população base, pois não corrige para vícios devidos à seleção, e somente produz estimativas não-viciadas de variâncias entre touros, quando há independência entre si (HENDERSON, 1953). Quanto aos métodos MIVQUE(o) e IHSM, por usarem pseudoesperanças nas suas construções, fornecem soluções apenas aproximadas das estimativas de variâncias e covariâncias. Uma vez que são fáceis de se usar, mesmo em situações de análise multivariadas e grandes conjunto de dados (HENDERSON, 1984; FREITAS et al., 1995), são recomendados métodos mais elaborados, como o ML e REML, para estimar valores iniciais de variâncias e covariâncias.

O ML, por não considerar os efeitos fixos associados ao modelo, é recomendável somente quando o número de observações for muito grande em relação ao número de elementos do vetor  $\beta$  (efeitos fixos). Sua distribuição teórica baseia-se em resultados assintóticos (HARVILLE, 1977). Em adição a estes métodos, o REML elimina vícios decorrentes da seleção e as estimativas de variâncias e covariâncias são obtidas pela maximização da função de verossimilhança construída com base nos resíduos, após ajustar os dados para os efeitos fixos (HARVILLE, 1977).

As estimativas de  $h^2$  (Tabela 3) variaram de 0,30 a 0,78 para pesos à desmama e de 0,28 a 0,68 para pesos aos doze meses, com comportamento similar ao das variâncias e covariâncias genéticas. Isto pode ser explicado pelo fato de que, no cálculo de  $h^2$ , os valores de  $g_{ij}$  são dominantes em relação aos de  $r_{ij}$ , e esses tiveram magnitude irregular nos diferentes métodos de estimação. Outros fatores podem interferir nas estimativas de  $h^2$  para determinada característica, raça, rebanho ou cruzamento: a seleção, o grau de parentesco entre os reprodutores, o tamanho amostral e ainda o modelo adotado, quando se usa um método específico. Quando se usam vários métodos para estimar covariâncias, a estrutura de distribuição

dos dados merece atenção especial. As duas análises realizadas por REML na amostra A e por 3H no C tem o propósito de ilustrar variações nos estimadores de parâmetros genéticos, quando se varia a amostra de dados.

Na análise multivariada (Tabela 3), ML e REML, por serem métodos correlatos, proporcionaram  $g_{ij}$  concordantes, ao contrário dos resultados obtidos das análises univariada, confirmando as observações de POLLAK (1985) e MEYER (1991), ou seja, na análise multivariada e ainda com a inclusão da matriz de parentesco entre os animais, são estimadas variâncias e covariâncias genéticas mais apropriadas. Quanto ao IHSM, apesar de os resultados satisfatórios terem sido obtidos (KRIESE et al., 1991a; KRIESE et al., 1991b; E FREITAS et al., 1995), o mesmo pareceu superestimar os parâmetros genéticos.

As estimativas de herdabilidade (Tabela 3) obtidas por ML e REML foram similares entre si e inferiores às do IHSM, porém foram superiores às obtidas com o método 3 de Henderson (3H).

Além dos fatores acima, o efeito da seleção praticada dentro do rebanho Canchim pode ter interferido nesses estimadores. Na presença desse efeito, há tendência de obter valores menores para  $g_{ij}$  (HENDERSON, 1984; CARABAÑO et al., 1989). Entretanto, o fato de ML e REML controlar e/ou minimizar vícios decorrentes da seleção sobre as estimativas (HENDERSON, 1984; OUTWELTJES et al., 1988) explica, em parte, os valores maiores de  $g_{ij}$  e de  $h^2$  obtidos por esses métodos.

Refinamentos para facilitar a manipulação dos dados também podem influenciar as estimativas de  $h^2$ . Utilizando-se análise univariada por REML na amostra B, FREITAS et al. (1994) estimaram  $h^2$  de 0,33; 0,77; 0,44; e 0,10, para pesos ao nascimento, pesos observados à desmama, pesos ajustados linearmente para 205 dias à desmama e ganho de peso da desmama aos 12 meses, respectivamente. Numa análise multivariada por REML e transformação canônica (REMLc), utilizando-se os mesmos dados, apenas a primeira variável canônica explicou 86,0% da variação detectada, proporcionando  $h^2$  conjunta de 0,83, consideravelmente maior que as obtidas individualmente, favorecendo, portanto, a seleção dos animais.

Uma vez que nessas análises foi assumida a hipótese de homogeneidade de variâncias dentro dos efeitos, a qual foi rejeitada, admite-se, em consequência desse fato, prejuízos nas estimativas de parâmetros genéticos. Segundo GIANOLA et al. (1992) e WEIGEL e GIANOLA (1992), esses efeitos

são maiores em análises multivariadas, pois não raro, heterogeneidade de variâncias geralmente implica em heterogeneidade de covariâncias, e o impacto deste fato sobre o progresso genético esperado é mais importante que no caso univariado.

Os valores dos coeficientes de assimetria, curtose e  $h^2$  obtidos com os dados da amostra A, analisados em quatro escalas, estão na Tabela 5. Embora tenha havido similaridade nas estimativas de herdabilidade e respectivos erros-padrão para todas as características, a transformação pelo método dos escores normais comportou-se como a mais adequada, pois os coeficientes de assimetria e de curtose foram praticamente nulos. Segundo BAXTER (1995), transformações que usam escores ou postos, além de serem úteis, quando os dados têm assimetria alta, como no presente estudo, eliminam as diferenças de grupos, quando aplicadas a dados bimodais ou multimodais, sendo mais robustas à presença de “outliers” que as outras. O mesmo autor ressalta, no entanto, que o

uso dessas transformações, uma vez que não utiliza toda a informação dos dados originais, proporciona estimativas não-suficientes. A transformação logarítmica, apesar de aproximar os dados a uma normal, estabilizar as variâncias e induzir um modelo linear para a resposta transformada (BAXTER, 1995; AZZALINI e VALLE, 1996; e DAVIDIAN e GILTINAN, 1996), quando o desvio-padrão dos dados, na escala original, varia linearmente com a média (BOLDMAN e FREEMAN, 1990), mostrou-se ineficaz para essa amostra de dados. FREITAS e GRAÇA (1996), analisando a amostra de dados A, observaram relação não-linear entre a média e o desvio-padrão para os efeitos de sexo, ordem de parto, ano de nascimento e touro, em que os coeficientes de variação variam de 2 a 33%, com oscilação crescente nesta ordem, o que pode ser justificado pelo fato de que, à medida que se aumenta o número de níveis de um efeito, há maior desbalanceamento no número de repetições e nos efeitos ambientais associados a este.

Tabela 5 - Estimativas da assimetria (Ass), curtose (Cur) e de  $h^2$  obtidas para pesos ao nascimento ( $y_1$ ), à desmama ( $y_2$ ), aos 12 ( $y_3$ ), 18 ( $y_4$ ) e 24 meses de idade ( $y_5$ ) e circunferência escrotal ( $y_6$ ) aos 12 meses de idade, obtidas de 2511 animais Canchim, nascidos de 1969 a 1991 (A) com os dados analisados em quatro escalas: observados ( $y_{obs}$ ), logarítmica ( $y_{log}$ ), padronizada ( $y_z$ ) e transformação de BLOM ( $y_{blom}$ )

Table 5 - Estimates of skewness, kurtosis and  $h^2$  obtained from body weights at birth ( $y_1$ ), weaning ( $y_2$ ), 12 ( $y_3$ ), 18 ( $y_4$ ) and 24 ( $y_5$ ) months of age obtained from 2511 Canchim cattle, born from 1969 to 1991 (A), with the data analysed in four scales: observed ( $y_{obs}$ ), logarithmic ( $y_{log}$ ), standard ( $y_z$ ) and BLOM ( $y_{blom}$ )

Carácter Trait	$y_{obs}$		$y_{log}$		$y_z$		$y_{blom}$	
	Ass Skewness	Cur Kurtosis	Ass Skewness	Cur Kurtosis	Ass Skewness	Cur Kurtosis	Ass Skewness	Cur Kurtosis
$y_1$	0,1459	0,1100	-0,4860	1,0055	0,1359	0,1100	0,0121	-0,0387
$y_2$	-0,1171	-0,6613	-0,7676	0,5549	-0,1171	-0,6613	-0,0013	-0,1247
$y_3$	0,1072	0,0936	-0,6002	0,6819	0,1072	0,0936	-0,0005	-0,0332
$y_4$	0,7333	1,3055	0,0114	0,7885	0,7333	1,3055	0,0010	-0,0342
$y_5$	0,9612	1,8580	0,2468	0,3620	0,9612	1,8580	0,0007	-0,0404
	Herdabilidade: $h^2_{\pm s}$ ( $h^2$ ) Heritability: $h^2_{\pm s}$ ( $h^2$ )							
$y_1$	0,3894 $\pm$ 0,1120		0,3980 $\pm$ 0,1148		0,3896 $\pm$ 0,1120		0,3923 $\pm$ 0,1129	
$y_2$	0,5415 $\pm$ 0,1612		0,5500 $\pm$ 0,1620		0,5360 $\pm$ 0,1612		0,5568 $\pm$ 0,1602	
$y_3$	0,5344 $\pm$ 0,1577		0,4870 $\pm$ 0,1485		0,5344 $\pm$ 0,1577		0,5182 $\pm$ 0,1543	
$y_4$	0,1630 $\pm$ 0,0702		0,1475 $\pm$ 0,0688		0,1630 $\pm$ 0,0700		0,1537 $\pm$ 0,0695	
$y_5$	0,2657 $\pm$ 0,1086		0,2990 $\pm$ 0,1158		0,2647 $\pm$ 0,1086		0,3043 $\pm$ 0,1174	

## Conclusões

Dados do desenvolvimento ponderal e da circunferência escrotal de animais Canchim não seguem distribuição normal, com a qualidade dos dados decrescendo com a idade adulta do animal.

Houve heterogeneidade de variâncias para os efeitos fixos de sexo, ano e época de nascimento, idade da vaca ao parto, geração do animal e efeito aleatório de touros, o que inviabiliza o uso de métodos de análise de variância tradicionais para estimar parâmetros genéticos.

A transformação que fornece escores normais como resposta foi eficiente para ajustar os dados a uma distribuição normal; contudo, as transformações usadas não influenciaram as estimativas de herdabilidade.

O método IHSM mostrou-se de fácil uso e adequado para manejar grandes conjunto de dados mesmo em situações de análise multivariadas, podendo ser recomendado para estimar valores iniciais de variâncias e covariâncias de métodos mais elaborados, como o ML e REML.

Na análise multivariada, ML e REML proporcionaram variâncias e covariâncias genéticas mais concordantes que as obtidas das análises univariada.

Métodos de estimação, análise univariada ou multivariada, diferentes amostras e tipo de transformação de dados, nesta ordem, foram os que mais influenciaram as estimativas de  $h^2$ .

## Referências Bibliográficas

AZZALINI, A., VALLE, A.D. 1996. The multivariate skew-normal distribution. *Biometrika*, 83(4):715-726.

BAXTER, M.J. 1995. Standardization and transformation in principal component analysis, with applications to archaeometry. *J. R. Statist. Soc. C*, 44(4):513-527.

BOLDMAN, K.G., FREEMAN, A.E. 1990. Adjustment for heterogeneity of variance by herd production level in dairy cow and sire evaluation. *J. Dairy Sci.*, 73(2):503-512.

BOLDMAN, K.G., KRIESE, L.A., VAN VLECK, L.D. et al. 1995. *A manual for use of MTDFREML*. A set of programs to obtain estimates of variances and covariances. Washington, DC: USDA-ARS. 120p.

BROTHERSTONE, S., HILL, W.G. 1986. Heterogeneity of variance amongst herds for milk production. *Animal Production*, 42(3):297-303.

BROWNIE, C., BOOS, D.D., OLIVER, J.H. 1990. Modifying the  $t$  and ANOVA  $F$  tests when treatment is expected to increase variability relative to controls. *Biometrics*, 46(1):259-266.

CARABAÑO, M.J., VAN VLECK, L.D., WIGGANS, G.R. et al. 1989. Estimation of genetic parameters for milk and fat yields of dairy cattle in Spain and the United States. *J. Dairy Sci.*, 22(11):3013-3022.

CHAPUIS, H., TIXIER-BOICAHRD, M., DELABROSSE, Y.

et al. 1996. Multivariate restricted maximum likelihood estimation of genetic parameters for production traits in three selected turkey strains. *Genet. Sel. Evol.*, 28(2):197-215.

COCHRAN, W.G., COX, D.F. 1978. *Deseno experimentales*. Mexico: Trillas. 661p.

COOK, R.D., WEISBERG, S. 1994. Transforming a response variable for linearity. *Biometrika*, 81(4):731-737.

DAVIDIAN, M., GILTINAN, D.M. 1996. *Nonlinear models for repeated measurement data*. 2.ed. London: Chapman Hall. 359p.

DEMPSTER, A.P., LAIRD, N.M., RUBIN, D.B. 1977. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Statist. Soc. B*, 39:1-38.

DE VEER, J.C., VAN VLECK, L. D. 1987. Genetic parameters for first lactation milk yields at three levels of herd production. *J. Dairy Sci.*, 70(7):1434-41.

FREITAS, A.R. de, FAVORETTI, A.C., ALENCAR, M.M. et al. 1994. Uso da máxima verossimilhança restrita e transformação canônica para estimação de parâmetros genéticos de características de crescimento em bovinos. *R. Soc. Bras. Zootec.*, 23(3):394-401.

FREITAS, A.R., GRAÇA, F.X.J. 1996. A qualidade dos dados de pesos de bovinos e o melhoramento genético. *Multiciência*, 1(1):93-102.

FREITAS, A.R., PEGORIN, M.J., ALENCAR, M.M. et al. 1995. Estimativas de parâmetros genéticos de múltiplas características através do método simples de Henderson. *Pesq. Agropec. Bras.*, 30(5):711-722.

FREITAS, A.R., VENCOVSKY, R. 1993a. Comparação de métodos para estimação de componentes de variâncias e parâmetros afins de múltiplas características em bovinos. *Pesq. Agropec. Bras.*, 28(4):453-463.

FREITAS, A.R., VENCOVSKY, R. 1993b. Métodos de estimação de variâncias e parâmetros afins de características de crescimento em bovinos. *Pesq. Agropec. Bras.*, 28(7):855-861.

GEOFFREY, J.M., KRISHNAN, K. 1997. *The EM algorithm and extensions*. New York: Wiley. 274p.

GIANOLA, D., FOULLEY, J.L., FERNANDO, R.L. et al. 1992. Estimation of heterogeneous variances using empirical bayes methods: Theoretical considerations. *J. Dairy Sci.*, 75(10):2805-2823.

HARTLEY, H.O., RAO, J.N.K. 1967. Maximum likelihood estimation for the mixed analysis of variance model. *Biometrika*, 54(1/2): 93-108.

HARVILLE, D.A. 1977. Maximum likelihood approaches to variance component estimation and to related problems. *J. American Statist. Association*, 72(358):320-339.

HEMMERLE, W.J., HARTLEY, H.O. 1973. Computing maximum likelihood estimates for the mixed model AOV model using the W transformations. *Technometrics*, 15(4):819-831.

HENDERSON, C.R. 1984. *Applications of linear models in animal breeding*. Ontario: University of Guelph. 426p.

HENDERSON, C.R. 1953. Estimation of variance and covariance components. *Biometrics*, 4(2):226-252.

JENRICH, R.I., SAMPSON, P.F. 1976. Newton-Raphson and related algorithms for maximum likelihood variance component estimation. *Tecnometrics*, 18(1):11-17.

JENSEN, J., MAO, I.L. 1988. 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.*, 66(11):2750-2761.

KRIESE, L.A., BERTRAND, J.K., BENYSHEK, L.L. 1991a. Age adjustment factors, heritabilities and genetic correlations for scrotal circumference and related growth traits in hereford and brangus bulls. *J. Anim. Sci.*, 69(2):478-489.



- KRIESE, L.A., BERTRAND, J.K., BENYSHEK, L.L. 1991b. Genetic and environmental growth trait parameter estimates for Brahman and Brahman-derivative cattle. *J. Anim. Sci.*, 69(6):2362-2370.
- MEUWISSEN, T.H.E., JONG, G., ENGEL, B. 1996. Joint estimation of breeding values and heterogeneous variances of large data sets. *J. Dairy Sci.*, 79(2):310-316.
- MEYER, K. 1991. Estimating variances and covariances for multivariate animal models by restricted maximum likelihood. *Genet. Sel. Evol.*, 23(1):67-83.
- MEYER, K. 1989. Restricted maximum likelihood to estimate variance components for animal models with several random effects using a derivative-free algorithm. *Genet. Sel. Evol.*, 21(3):247-398.
- MEYER, K., SMITH, S.P. 1996. Restricted maximum likelihood estimation for animal models using derivatives of the likelihood. *Genet. Sel. Evol.*, 28(1):23-49.
- MISZTAL, I. 1994. Comparison of computing properties of derivative and derivative-free algorithms in variance-component estimation by REML. *J. Anim. Breed. Gen.*, 111(5-6):346-355.
- Ouweltjes, W., Schaeffer, L.R., Kennedy, B.W. 1988. Sensitivity of methods of variance component estimation to culling type of selection. *J. Dairy Sci.*, 71(3):723-729.
- PATTERSON, H.D., THOMPSON, R. 1971. Recovery of interblock information when block sizes are unequal. *Biometrika*, 58:545-554.
- POLLAK, E.J. 1985. Genetic evaluation of beef cattle from performance. In: SIMPÓSIO INTERNACIONAL DE PRODUÇÃO ANIMAL, 1, 1983, Ribeirão Preto. *Anais...* Ribeirão Preto: Sociedade Brasileira de Genética, p.73-82.
- RAMOS, A.A., VALÊNCIA, E.F.T., WEICHSLER, F.S. et al. Heterogeneidade de variâncias das características de produção de bovinos da raça holandesa no trópico II. Estratificação por nível de produção de vacas. In: REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE ZOOTECNIA, 33, 1996, Fortaleza. *Anais...* Fortaleza: SBZ, 1996, p.61-73.
- ROBERT, C., DUCROCQ, V. 1996. Computation of all eigenvalues of matrices used in restricted maximum likelihood estimation of variance components using sparse matrix techniques. *Genet. Sel. Evol.*, 28(1):51-65.
- RODRÍGUEZ-ALMEIDA, F.A., VAN VLECK, L.D., CUNDIFF, L.V. 1995a. Effect of accounting for different phenotypic variances by sire breed, and sex on selection of sires based on expected progeny differences for 200- and 365-day weights. *J. Anim. Sci.*, 73(9):2589-2599.
- RODRÍGUEZ-ALMEIDA, F.A., VAN VLECK, L.D., CUNDIFF, L.V. et al. 1995b. Heterogeneity of variance by sire breed, sex, and dam breed in 200- and 365-day weights of beef cattle from a top cross experiment. *J. Anim. Sci.*, 73(9):2579-2588.
- SAS INSTITUTE. 1993. In: *SAS/STAT User's guide: statistics*. versão 6, v.2. 4.ed. Cary.
- SAS INSTITUTE. 1996. In: *Procedures*. versão 6, v.1. 3.ed. Cary.
- SCHEFFÉ, A. 1959. *The analysis of variances*. [S.l.]: John Wiley. 477p.
- TAYLOR, J.F., EVERET, R.W. 1985. Estimation of variance components by the expectation-maximization algorithm for restricted maximum likelihood in a repeatability model for semen production. *J. Dairy Sci.*, 68(11):48-53.
- VISSCHER, P.M. 1992. Power of likelihood ratio tests for heterogeneity of intraclass correlation and variance in balanced half-sib design. *J. Dairy Sci.*, 75(5):1320-1330.
- WEIGEL, K.A., GIANOLA, D. 1992. Estimation of heterogeneous within-herd variance components using empirical bayes methods: a simulation study. *J. Dairy Sci.*, 75(10):2824-2833.

**Recebido em:** 19/11/98

**Aceito em:** 13/08/99