

AVALIAÇÃO GENÉTICA: DOS DADOS ÀS DEP'S

Elias Nunes Martins

INTRODUÇÃO

A avaliação genética visa a identificação dos indivíduos geneticamente superiores de tal sorte que, usados na reprodução, leguem aos seus descendentes sua superioridade, alterando dessa forma a média da população. Sob esta definição o indivíduo é visto como um veículo de genes que devem se expressar na geração subsequente.

O recurso disponível para a execução da avaliação genética é a observação das características economicamente importantes. O texto que segue tem por objetivo apresentar, com o auxílio de um exemplo simulado, os procedimentos envolvidos no processamento dos dados até a obtenção das predições das diferenças esperadas nas progênes (DEP's).

DESCRIÇÃO DA OBSERVAÇÃO

O primeiro passo para a utilização racional das observações das características economicamente importantes é o estabelecimento de um modelo que explique, convenientemente, a estrutura dos dados observados.

Tem sido admitido que o modelo adequado seja linear, de maneira que a observação é definida como o somatório de vários efeitos classificados como ambientais e genéticos. Contudo, nem todo efeito ambiental é identificável, por não ter origem conhecida e, assim, um modelo possível é:

$$y_{ij} = f_i + g_{ij} + E_{ij},$$

em que

y_{ij} é a observação referente ao indivíduo j criado no ambiente i ;

f_i é o conjunto de efeitos identificáveis do ambiente i ;

g_{ij} é o conjunto de efeitos genéticos no indivíduo j criado no ambiente i ;

E_{ij} é o conjunto de efeitos ambientais não identificáveis no indivíduo j criado no ambiente i .

Naturalmente, entre os efeitos genéticos o particular interesse reside naqueles aditivos porque são eles os determinantes dos valores genéticos dos indivíduos. Então o modelo pode ser reescrito como:

$$y_{ij} = f_i + a_{ij} + d_{ij} + E_{ij},$$

em que

a_{ij} é o valor genético do indivíduo j criado no ambiente i ;

d_{ij} é o conjunto de efeitos genéticos não-aditivos no indivíduo j criado no ambiente i .

De acordo com esse modelo a_{ij} , d_{ij} e E_{ij} não são estimáveis visto que estão confundidos por estarem associados exclusivamente à observação y_{ij} . Por outro lado, f_i é estimável se d_{ij} , d_{ij} e E_{ij} têm esperança nula. Esta pressuposição é coerente com a premissa de que não ocorre mudança na estrutura genética da população se os indivíduos que se acasalam são uma amostra aleatória, desconsiderando-se a existência de mutações, migrações e erros de amostragem.

UMA FORMA DE ELIMINAR OS EFEITOS AMBIENTAIS IDENTIFICÁVEIS

Usando esse conceito pode-se redefinir o modelo inicialmente proposto da seguinte forma:

$$y_{ij} = f_i + \varepsilon_{ij},$$

em que $\varepsilon_{ij} = a_{ij} + d_{ij} + E_{ij}$, com média dada por

$$E(\varepsilon_{ij}) = 0 \text{ e variância dada por } E(\varepsilon_{ij}^2) = \sigma_\varepsilon^2.$$

Um conveniente estimador para f_i pode ser obtido de forma que $E(\varepsilon_{ij}^2)$ seja mínima, como segue.

Na forma matricial o modelo é escrito como:

$$y = X\beta + \varepsilon,$$

em que

y é o vetor de observações;

X é a matriz de incidência dos efeitos fixos de ambiente;

β é o vetor de efeitos fixos de ambiente a serem estimados;

ε é o vetor de resíduos.

Minimizar $E(\varepsilon_{ij}^2)$ equivale a minimizar $E(\varepsilon'\varepsilon)$, então,

$$\begin{aligned} E(\varepsilon'\varepsilon) &= E[(y - X\beta)'(y - X\beta)] \\ &= E(y'y - 2\beta'X'y + \beta'X'X\beta) \end{aligned}$$

Pode ser observado que $E(\varepsilon'\varepsilon)$ é uma função quadrática de β , o vetor de efeitos fixos de ambiente a serem estimados. Assim, a derivada dessa função, em relação a β , quando nula permite a estimação desejada:

$$\begin{aligned}\delta[E(\varepsilon'\varepsilon)]/\delta\beta' &= \delta[E(y'y - 2\beta'X'y + \beta'X'X\beta)]/\delta\beta' \\ &= -2X'y + 2X'X\beta \\ -X'y + X'X\beta &= 0 \\ X'X\beta &= X'y.\end{aligned}$$

Considerando que sejam feitas adequadas restrições se $X'X$ não tem posto completo, então:

$$\hat{\beta} = (X'X)^{-1} X'y.$$

Uma vez estimados os efeitos fixos de ambiente as observações podem ser corrigidas por $y - X\hat{\beta}$.

Para ilustrar a estimação de efeitos fixos de ambiente e o ajustamento dos dados considere-se o conjunto de informações simuladas de ganho de peso da desmama ao sobreano em gado Nelore apresentado na Tabela 12.1.

Além do ganho de peso, são fornecidas também informações sobre a genealogia, sexo do animal, ano de nascimento e rebanho em que o animal foi criado. Essas informações permitem formular o seguinte modelo para os ganhos de peso observados.

$$y_{ij} = \mu + gc_i + \varepsilon_{ij},$$

em que

y_{ij} é a observação de ganho de peso referente ao indivíduo j pertencente ao grupo contemporâneo i ;

gc_i é o efeito de ambiente no grupo contemporâneo i ;

ε_{ij} é o resíduo no ganho de peso do indivíduo j pertencente ao grupo contemporâneo i .

Os grupos contemporâneos foram formados por animais do mesmo sexo, nascidos num mesmo ano e no mesmo rebanho, como mostrado na Tabela 12.2. Deve-se ter em mente que o conjunto de dados aqui apresentado tem apenas a finalidade de ilustrar procedimentos de avaliação genética. Assim, é certo que aparecem situações por vezes muito distantes do que acontece nas aplicações práticas. Assim, a formação de grupos contemporâneos com número muito reduzido de animais como ocorre no exemplo não devem ser utilizados em situações reais.

O processo de estimação dos efeitos de ambiente nos grupos contemporâneos, por meio do Método de Quadrados Mínimos produz as estimativas apresentadas na Tabela 12.3.

Uma vez obtidas as estimativas dos efeitos de ambiente de cada grupo contemporâneo, o ajustamento dos dados é feito subtraindo-se de cada observação a estimativa de efeito de ambiente correspondente ao grupo contemporâneo onde ela foi produzida, como pode ser visto na Tabela 12.4.

TABELA 12.1. Dados Simulados de Ganho de Peso da Desmama ao Sobreano

ANIMAL	PAI	MÃE	SEXO	ANO	REBANHO	GANHO
1	-	-	1	82	1	-
2	-	-	1	82	2	-
3	-	-	1	82	3	-
4	-	-	2	80	1	-
5	-	-	2	80	2	-
6	-	-	2	81	2	-
7	-	-	2	81	3	-
8	-	-	2	80	1	-
9	-	-	2	81	2	-
10	-	-	2	81	3	-
11	-	-	2	80	1	-
12	-	-	2	80	2	-
13	-	-	2	81	2	-
14	-	-	2	81	3	-
15	-	-	2	81	3	-
16	1	4	2	84	1	53,64
17	1	5	1	84	2	72,40
18	1	6	2	84	2	67,65
19	1	7	1	84	3	74,65
20	2	8	2	84	1	77,02
21	2	9	1	84	2	95,65
22	2	10	1	84	3	90,40
23	3	11	2	84	1	73,65
24	3	12	2	84	2	72,15
25	3	13	1	84	2	79,90
26	3	14	1	84	3	101,65
27	3	15	1	84	3	99,40
28	17	24	1	87	2	84,77
29	17	41	1	87	3	81,77

TABELA 12.1. Dados Simulados de Ganho de Peso da Desmama ao Sobreano (cont.)

ANIMAL	PAI	MÃE	SEXO	ANO	REBANHO	GANHO
30	22	23	1	87	1	88,52
31	22	16	2	87	1	68,40
32	22	42	2	87	2	56,77
33	22	43	2	87	2	88,27
34	26	18	1	87	2	92,65
35	26	20	1	87	1	95,84
36	26	44	1	87	3	123,40
37	26	45	2	87	3	90,90
38	26	46	2	87	3	81,90
39	27	47	1	87	1	46,90
40	27	48	2	87	1	63,90

TABELA 12.2. Formação dos Grupos Contemporâneos (GC) para os Animais com Dados Observados de Ganho de Peso da Desmama ao Sobreano

ANIMAL	SEXO	ANO	REBANHO	GANHO	GC
17	1	84	2	72,40	1
21	1	84	2	95,65	1
25	1	84	2	79,90	1
26	1	84	3	101,65	2
22	1	84	3	90,40	2
19	1	84	3	74,65	2
27	1	84	3	99,40	2
39	1	87	1	46,90	3
35	1	87	1	95,84	3
30	1	87	1	88,52	3
34	1	87	2	92,65	4
28	1	87	2	84,77	4
36	1	87	3	123,40	5
29	1	87	3	81,77	5

TABELA 12.2. Formação dos Grupos Contemporâneos (GC) para os Animais com Dados Observados de Ganho de Peso da Desmama ao Sobreano (cont.)

ANIMAL	SEXO	ANO	REBANHO	GANHO	GC
20	2	84	1	77,02	6
16	2	84	1	53,64	6
23	2	84	1	73,65	6
18	2	84	2	67,65	7
24	2	84	2	72,15	7
40	2	87	1	63,90	8
31	2	87	1	68,40	8
32	2	87	2	56,77	9
33	2	87	2	88,27	9
37	2	87	3	90,90	10
38	2	87	3	81,90	10

Tabela 12.3. Estimativas de efeitos de ambiente nos grupos contemporâneos, obtidas por meio do Método de Quadrados Mínimos

	GRUPOS CONTEMPORÂNEOS									
	1	2	3	4	5	6	7	8	9	10
Efeito	82,65	91,52	77,08	88,71	102,58	68,02	69,90	66,15	72,52	86,40

TABELA 12.4. Ajustamento dos Dados Observados de Ganho de Peso para os Efeitos de Ambiente dos Grupos Contemporâneos (GC)

ANIMAL	GC	GANHO	AJUSTAMENTO	GANHO AJUSTADO
17	1	72,40	72,40 - 82,65	-10,25
21	1	95,65	95,65 - 82,65	13,00
25	1	79,90	79,90 - 82,65	-2,75
26	2	101,65	101,65 - 91,52	10,13
22	2	90,40	90,40 - 91,52	-1,12
19	2	74,65	74,65 - 91,52	-16,87
27	2	99,40	99,40 - 91,52	7,88
39	3	46,90	46,90 - 77,08	-30,18

TABELA 12.4. Ajustamento dos Dados Observados de Ganho de Peso para os Efeitos de Ambiente dos Grupos Contemporâneos (GC) (cont.)

ANIMAL	GC	GANHO	AJUSTAMENTO	GANHO AJUSTADO
35	3	95,84	95,84 - 77,08	18,76
30	3	88,52	88,52 - 77,08	11,44
34	4	92,65	92,65 - 88,71	3,94
28	4	84,77	84,77 - 88,71	-3,94
36	5	123,40	123,40 - 102,58	20,82
29	5	81,77	81,77 - 102,58	-20,81
20	6	77,02	77,02 - 68,02	9,00
16	6	53,64	53,64 - 68,02	-14,62
23	6	73,65	73,65 - 68,02	5,63
18	7	67,65	67,65 - 69,90	-2,25
24	7	72,15	72,15 - 69,90	2,25
40	8	63,90	63,90 - 66,15	-2,25
31	8	68,40	68,40 - 66,15	2,25
32	9	56,77	56,77 - 72,52	-15,75
33	9	88,27	88,27 - 72,52	15,75
37	10	90,90	90,90 - 86,40	4,50
38	10	81,90	81,90 - 86,40	-4,75

Dessa forma efeitos identificáveis de ambiente foram eliminados dos dados e agora os ganhos de peso ajustados são descritos pelo modelo

$$\varepsilon_i = a_i + d_i + E_i,$$

em que:

a_i é o valor genético do indivíduo i ;

d_i é o conjunto de efeitos genéticos não aditivos no indivíduo i ;

E_i é o conjunto de efeitos ambientais não identificáveis no indivíduo i .

UMA FORMA DE OBTER A PREDIÇÃO DOS VALORES GENÉTICOS

Nosso interesse é, em particular, obter uma predição para o valor genético a_i e então para facilitar o entendimento podemos reescrever o modelo como

$$\varepsilon_i = a_i + e_i,$$

em que $e_i = d_i + E_i$.

Agora para se estabelecer um preditor para os valores genéticos é conveniente escrever expressões para as médias e variâncias de ε_i e a_i .

As médias para ε_i e a_i , dadas por $E(\varepsilon_i)$ e $E(a_i)$, são por definição iguais a zero, enquanto que as variâncias são dadas respectivamente por $E(\varepsilon_i^2) = \sigma_\varepsilon^2$, que é a variância fenotípica, e $E(a_i^2) = \sigma_a^2$, que é a variância genética aditiva. Adicionalmente, pode-se definir a covariância entre ε_i e a_i por $E(\varepsilon_i a_i) = \sigma_a^2$, que é a própria variância genética aditiva, visto que a_i está contido em ε_i e é independente de e_i .

Um possível preditor para a_i pode ser $\hat{a}_i = b\hat{\varepsilon}_i$, de tal sorte que o valor obtido pelo preditor \hat{a}_i seja o mais próximo possível do valor genético verdadeiro a_i . Em outras palavras significa encontrar o valor de b que torna a variância do erro de predição, dada por $Var(\hat{a}_i - a_i) = E[(\hat{a}_i - a_i)^2]$, o menor possível. Isso pode ser feito como segue:

$$\begin{aligned} E[(\hat{a}_i - a_i)^2] &= E[(b\hat{\varepsilon}_i - a_i)^2] \\ &= E(b^2\hat{\varepsilon}_i^2 - 2b\varepsilon_i a_i + a_i^2) \\ &= E(b^2\hat{\varepsilon}_i^2) - E(2b\varepsilon_i a_i) + E(a_i^2) \end{aligned}$$

Como b é uma constante

$$= b^2 E(\hat{\varepsilon}_i^2) - 2b E(\varepsilon_i a_i) + E(a_i^2)$$

Usando as definições feitas anteriormente para as variâncias e covariâncias de ε_i e a_i , temos:

$$E[(\hat{a}_i - a_i)^2] = b^2 \sigma_\varepsilon^2 - 2b \sigma_a^2 + \sigma_a^2$$

Por meio dessa expressão verifica-se que a variância do erro de predição é uma função quadrática de b , e assim a derivada dessa função, em relação a b , quando nula permite obter o valor de b que torna mínima a variância do erro de predição.

$$\begin{aligned} \delta \left\{ E[(\hat{a}_i - a_i)^2] \right\} / \delta b &= 2b \sigma_\varepsilon^2 - 2 \sigma_a^2 \\ 2b \sigma_\varepsilon^2 - 2 \sigma_a^2 &= 0 \end{aligned}$$

$$b = \frac{\sigma_a^2}{\sigma_\varepsilon^2} = h^2 = \text{coeficiente de herdabilidade.}$$

Esse resultado significa que se alguém pretende predizer o valor genético a partir dos desempenhos individuais dos animais, a predição mais próxima do valor genético verdadeiro é obtida corrigindo-se as observações para os efeitos fixos de ambiente e em seguida multiplicando-se o resultado pelo coeficiente de herdabilidade.

Supondo-se um coeficiente de herdabilidade de 0,23, a título de ilustração, pode-se aplicar essa conclusão ao conjunto de dados fornecido na Tabela 12.1, como se mostra na Tabela 12.5.

Observa-se que a classificação dos animais com base nos dados ajustados e com base na predição do valor genético é a mesma. Isto é óbvio visto que a predição do valor genético é obtida multiplicando-se os dados ajustados por uma constante, que é o coeficiente de herdabilidade.

TABELA 12.5. Predições para os Valores Genéticos com Base nos Dados Individuais de Ganho de Peso da Desmama ao Sobreano Ajustados para os Efeitos de Ambiente dos Grupos Contemporâneos (GC)

ANIMAL	GANHO	GANHO AJUSTADO	VALOR GENÉTICO PREDITO
36	123,40	20,82	4,79
35	95,84	18,76	4,31
33	88,27	15,75	3,62
21	95,65	13,00	2,99
30	88,52	11,44	2,63
26	101,65	10,13	2,33
20	77,02	9,00	2,07
27	99,40	7,88	1,81
23	73,65	5,63	1,29
37	90,90	4,50	1,03
34	92,65	3,94	0,91
31	68,40	2,25	0,52
24	72,15	2,25	0,52
22	90,40	-1,12	-0,26
18	67,65	-2,25	-0,52
40	63,90	-2,25	-0,52
25	79,90	-2,75	-0,63
28	84,77	-3,94	-0,91
38	81,90	-4,75	-1,03
17	72,40	-10,25	-2,36
16	53,64	-14,62	-3,36
32	56,77	-15,75	-3,62
19	74,65	-16,87	-3,88
29	81,77	-20,81	-4,79
39	46,90	-30,18	-6,94

A PREDIÇÃO DO VALOR GENÉTICO CONSIDERANDO OS PARENTESCOS

Anteriormente foi comentado que o valor genético não era estimável por estar confundido com os efeitos genéticos não aditivos e ambientais, visto que esses três efeitos estão associados exclusivamente à observação y_{ij} . No entanto, se for dada atenção à estrutura genealógica (Figura 12.1) verifica-se que os valores genéticos apresentam associações com outras observações, além daquela feita no próprio animal.

Estas associações podem ser descritas pelas relações de parentesco. Assim, se estas associações forem consideradas, as observações fornecidas por parentes também podem ser utilizadas como base para a predição dos valores de terminado indivíduo.

Para que se tenha uma melhor noção de como isso pode ser feito, tome-se como exemplo o animal 17 que tem os animais 16, 18 e 19 como meio-irmãos, os animais 28 e 29 como filhos e os animais 31 e 34 como sobrinhos. Os coeficientes de parentesco do animal 17 com seus meio-irmãos é 0,25; com seus filhos é 0,5 e com seus sobrinhos é 0,125. Assim, um preditor para o valor genético do animal 17, que leve em consideração o seu desempenho e os de seus parentes seria:

$$\hat{a}_{17} = b_1 \hat{e}_{17} + b_2 \hat{e}_{16} + b_3 \hat{e}_{18} + b_4 \hat{e}_{19} + b_5 \hat{e}_{28} + b_6 \hat{e}_{29} + b_7 \hat{e}_{31} + b_8 \hat{e}_{34}.$$

A questão agora se resume em definir os valores de b_1 até b_8 , que são os coeficientes de regressão do valor genético do animal 17 em função das observações com ele associadas. Aqui novamente é desejado um preditor que produza predições o mais próximo possível dos valores genéticos verdadeiros, o que equivale a minimizar a variância do erro de predição dada por $Var(\hat{a}_i - a_i) = E[(\hat{a}_i - a_i)^2]$.

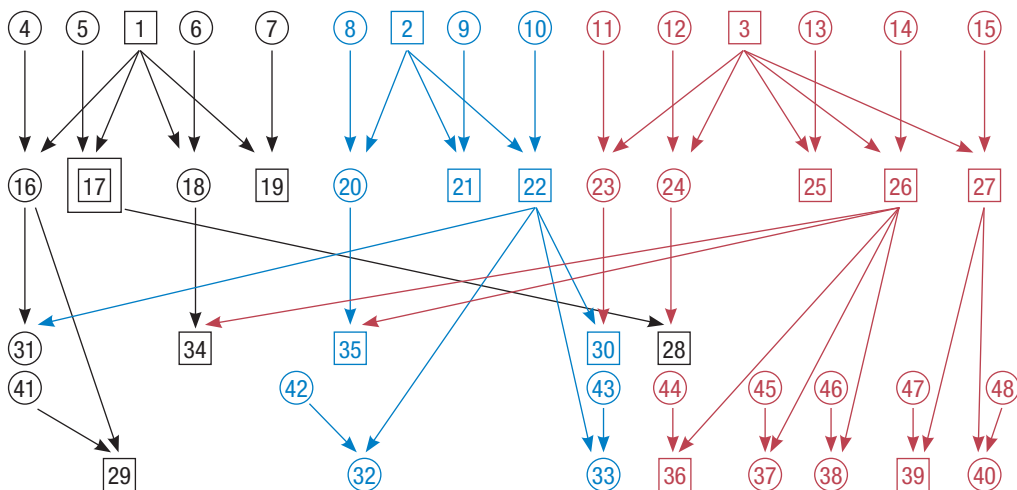


FIGURA 12.1. Genealogia dos animais.

Na forma matricial o preditor \hat{a}_{17} pode ser escrito como

$$\hat{a}_{17} = [b_1 \quad b_2 \quad b_3 \quad b_4 \quad b_5 \quad b_6 \quad b_7 \quad b_8] \begin{bmatrix} \hat{\epsilon}_{17} \\ \hat{\epsilon}_{16} \\ \hat{\epsilon}_{18} \\ \hat{\epsilon}_{19} \\ \hat{\epsilon}_{28} \\ \hat{\epsilon}_{29} \\ \hat{\epsilon}_{31} \\ \hat{\epsilon}_{34} \end{bmatrix}$$

$$\hat{a}_{17} = b' \epsilon$$

e então

$$\begin{aligned} E[(\hat{a}_1 - a_i)^2] &= E[(b' \hat{\epsilon} - a_i)^2] \\ &= E(b' \hat{\epsilon} \hat{\epsilon}' b - 2b' \hat{\epsilon} a_i + a_i^2) \\ &= E(b' \hat{\epsilon} \hat{\epsilon}' b) - E(2b' \hat{\epsilon} a_i) + E(a_i^2) \\ &= b' E(\hat{\epsilon} \hat{\epsilon}') b - 2b' E(\hat{\epsilon} a_i) + E(a_i^2) \end{aligned}$$

O termo $E(\hat{\epsilon} \hat{\epsilon}')$ é a matriz V de variância e covariância fenotípica referente ao vetor de informações e o termo $E(\hat{\epsilon} a_i)$ é o vetor g de covariância entre o valor genético a ser predito e o vetor de informações. Então,

$$E[(\hat{a}_i - a_i)^2] = b' V b - 2b' g + \sigma_a^2$$

Verifica-se aqui também que a variância do erro de predição é uma função quadrática de b , e assim a derivada dessa função, em relação a b , quando nula permite obter o valor de b que torna mínima a variância do erro de predição.

$$\begin{aligned} \delta \left\{ E[(\hat{a}_i - a_i)^2] \right\} / \delta b' &= 2b' V - 2g' \\ 2b' V - 2g' &= 0 \end{aligned}$$

Para ilustrar esse resultado pode-se aplicá-lo para obter-se a predição do valor genético para o animal 17. Nesse caso a matriz V e o vetor g serão

$$V = \begin{bmatrix} \sigma_e^2 & \frac{1}{4}\sigma_a^2 & \frac{1}{4}\sigma_a^2 & \frac{1}{4}\sigma_a^2 & \frac{1}{2}\sigma_a^2 & \frac{1}{2}\sigma_a^2 & \frac{1}{8}\sigma_a^2 & \frac{1}{8}\sigma_a^2 \\ \frac{1}{4}\sigma_a^2 & \sigma_e^2 & \frac{1}{4}\sigma_a^2 & \frac{1}{4}\sigma_a^2 & \frac{1}{8}\sigma_a^2 & \frac{1}{8}\sigma_a^2 & \frac{1}{2}\sigma_a^2 & \frac{1}{8}\sigma_a^2 \\ \frac{1}{4}\sigma_a^2 & \frac{1}{4}\sigma_a^2 & \sigma_e^2 & \frac{1}{4}\sigma_a^2 & \frac{1}{8}\sigma_a^2 & \frac{1}{8}\sigma_a^2 & \frac{1}{8}\sigma_a^2 & \frac{1}{2}\sigma_a^2 \\ \frac{1}{4}\sigma_a^2 & \frac{1}{4}\sigma_a^2 & \frac{1}{4}\sigma_a^2 & \sigma_e^2 & \frac{1}{8}\sigma_a^2 & \frac{1}{8}\sigma_a^2 & \frac{1}{8}\sigma_a^2 & \frac{1}{8}\sigma_a^2 \\ \frac{1}{2}\sigma_a^2 & \frac{1}{8}\sigma_a^2 & \frac{1}{8}\sigma_a^2 & \frac{1}{8}\sigma_a^2 & \sigma_e^2 & \frac{1}{4}\sigma_a^2 & \frac{1}{16}\sigma_a^2 & \frac{1}{16}\sigma_a^2 \\ \frac{1}{2}\sigma_a^2 & \frac{1}{8}\sigma_a^2 & \frac{1}{8}\sigma_a^2 & \frac{1}{8}\sigma_a^2 & \frac{1}{4}\sigma_a^2 & \sigma_e^2 & \frac{1}{16}\sigma_a^2 & \frac{1}{16}\sigma_a^2 \\ \frac{1}{8}\sigma_a^2 & \frac{1}{2}\sigma_a^2 & \frac{1}{8}\sigma_a^2 & \frac{1}{8}\sigma_a^2 & \frac{1}{16}\sigma_a^2 & \frac{1}{16}\sigma_a^2 & \sigma_e^2 & \frac{1}{16}\sigma_a^2 \\ \frac{1}{8}\sigma_a^2 & \frac{1}{8}\sigma_a^2 & \frac{1}{2}\sigma_a^2 & \frac{1}{8}\sigma_a^2 & \frac{1}{16}\sigma_a^2 & \frac{1}{16}\sigma_a^2 & \frac{1}{16}\sigma_a^2 & \sigma_e^2 \end{bmatrix}$$

$$g' = [\sigma_a^2 \quad \frac{1}{4}\sigma_a^2 \quad \frac{1}{4}\sigma_a^2 \quad \frac{1}{4}\sigma_a^2 \quad \frac{1}{2}\sigma_a^2 \quad \frac{1}{2}\sigma_a^2 \quad \frac{1}{8}\sigma_a^2 \quad \frac{1}{8}\sigma_a^2]$$

Supondo-se que a variância fenotípica seja de 900 kg² e a variância genética aditiva de 207 kg², têm-se:

$$V = \begin{bmatrix} 900 & 51,75 & 51,75 & 51,75 & 103,5 & 103,5 & 25,875 & 25,875 \\ 51,75 & 900 & 51,75 & 51,75 & 25,875 & 25,875 & 103,5 & 25,875 \\ 51,75 & 51,75 & 900 & 51,75 & 25,875 & 25,875 & 25,875 & 103,5 \\ 51,75 & 51,75 & 51,75 & 900 & 25,875 & 25,875 & 25,875 & 25,875 \\ 103,5 & 25,875 & 25,875 & 25,875 & 900 & 51,75 & 14,9375 & 14,9375 \\ 103,5 & 25,875 & 25,875 & 25,875 & 51,75 & 900 & 14,9375 & 14,9375 \\ 25,875 & 103,5 & 25,875 & 25,875 & 14,9375 & 14,9375 & 900 & 14,9375 \\ 25,875 & 25,875 & 103,5 & 25,875 & 14,9375 & 14,9375 & 14,9375 & 900 \end{bmatrix}$$

$$g' = [207 \quad 51,75 \quad 51,75 \quad 51,75 \quad 103,5 \quad 103,5 \quad 25,875 \quad 25,875]$$

$$b' = g'V^{-1}$$

$$b' = [0,203962 \quad 0,034909 \quad 0,034909 \quad 0,036175 \quad 0,083251 \quad 0,083251 \quad 0,013835 \quad 0,013835]$$

Assim, pode-se obter a predição do valor genético do animal 17 baseando-se no seu desempenho e no de seus parentes por:

$$\hat{a}_{17} = \begin{bmatrix} 0,203962 & 0,034909 & 0,034909 & 0,036175 & 0,083251 & 0,083251 & 0,013835 & 0,013835 \end{bmatrix} \begin{bmatrix} -10,25 \\ -14,62 \\ -2,25 \\ -16,87 \\ -3,94 \\ -20,81 \\ 2,25 \\ 3,94 \end{bmatrix} = -5,27$$

Observe-se que a predição do valor genético do animal 17 quando baseada na informação exclusiva de seu desempenho foi de -2,36. Contudo, quando a informação do desempenho de seus parentes foi levada em consideração, a predição de seu valor genético caiu para -5,27. É obvio que isto aconteceu porque seus parentes apresentam desempenho inferior.

Ao se utilizar informações de parentes, na forma como mostrado agora, não foi considerado que estas informações de desempenho também podem estar associadas a outras, como por exemplo, o animal 28 é filho de 17 mas também é filho de 24. Observe-se que se for considerada a informação de desempenho do animal 24 para se predizer o valor genético de 17, o que acontecerá é um ajustamento da informação de 28 para o desempenho de sua mãe e, automaticamente, a predição para o valor genético de 17 também estará ajustada.

Pensando dessa forma para predizer o valor genético de 17 podemos considerar a informação de seu próprio desempenho e dos animais 16, 18, 19, 28, 29, 31, 34, 22, 24 e 26. Nesse caso a predição seria dada como segue.

$$\hat{a}_{17} = \begin{bmatrix} b_1 & b_2 & b_3 & b_4 & b_5 & b_6 & b_7 & b_8 & b_9 & b_{10} & b_{11} \end{bmatrix} \begin{bmatrix} \hat{E}_{17} \\ \hat{E}_{16} \\ \hat{E}_{18} \\ \hat{E}_{19} \\ \hat{E}_{28} \\ \hat{E}_{29} \\ \hat{E}_{31} \\ \hat{E}_{34} \\ \hat{E}_{22} \\ \hat{E}_{24} \\ \hat{E}_{26} \end{bmatrix} = \begin{bmatrix} 0,203835 & 0,034864 & 0,034864 & 0,036147 & 0,084382 & 0,083198 & 0,01401 & 0,01401 & -0,00161 & -0,0097 & -0,00161 \end{bmatrix} \begin{bmatrix} -10,25 \\ -14,62 \\ -2,25 \\ -16,87 \\ -3,94 \\ -20,81 \\ 2,25 \\ 3,94 \\ -1,12 \\ 2,25 \\ 10,13 \end{bmatrix}$$

$$\hat{a}_{17} = -5,3$$

Observe-se que os valores de b_9 , b_{10} e b_{11} são negativos. Isto acontece porque corrigem a predição do valor genético de 17.

Se para predição do valor genético de 17 for utilizada a informação de algum animal que não tenha qualquer relação com os parentes de 17 ou com o próprio 17, o valor de b correspondente a esse animal seria zero, assim essa informação em nada contribuiria para a predição do valor genético do animal 17.

Sendo assim, esse raciocínio pode ser expandido a todo o conjunto de dados e, assim, utilizar-se toda informação coletada para a predição dos valores genéticos. Nesse caso, a matriz V será a mesma para a predição dos valores genéticos de todos animais e os vetores g necessários para cada predição podem ser agrupados formando a matriz G , que é a matriz de variância e covariância genética aditiva entre os dados.

O produto de GV^{-1} produz uma matriz B em que cada linha corresponde a um vetor b^2 referente a cada valor genético a ser predito. Isto permite que todas as predições sejam obtidas simultaneamente, considerando toda informação disponível. A expressão que descreve tal situação é

$$\hat{a} = B\hat{\epsilon} = GV^{-1}\hat{\epsilon} = GV^{-1}(y - X\hat{\beta})$$

De maneira formal essas expressões são derivadas como segue. O que se deseja é minimizar $Var(\hat{a} - a)$, a variância do erro de predição. Isso equivale a minimizar $E[(\hat{a} - a)(\hat{a} - a)']$. Assim

$$\begin{aligned} E[(\hat{a} - a)(\hat{a} - a)'] &= E[(B\hat{\epsilon} - a)(B\hat{\epsilon} - a)'] \\ &= E[(B\hat{\epsilon} - a)(\hat{\epsilon}'B' - a')] \\ &= E(B\hat{\epsilon}\hat{\epsilon}'B' - B\hat{\epsilon}a' - a\hat{\epsilon}'B' + aa') \end{aligned}$$

Sendo a variância do erro de predição uma função quadrática de B a derivada dessa função, em relação a B , quando nula permite a estimação desejada.

$$\delta \left\{ E[(\hat{a} - a)(\hat{a} - a)'] \right\} / \delta B' = 2BE(\hat{\epsilon}\hat{\epsilon}') - 2E(a\hat{\epsilon}')$$

em que $E(\hat{\epsilon}\hat{\epsilon}')$ e $E(a\hat{\epsilon}')$ são, respectivamente, as matrizes de variância e covariâncias fenotípica e genética, V e G . Então,

$$BV = G$$

$$B = GV^{-1}.$$

Se não houver qualquer parentesco entre os indivíduos, então, $G = I\sigma_a^2$ e $V = I\sigma_\epsilon^2 = I(\sigma_a^2 + \sigma_a^2 + \sigma_\epsilon^2) = I\sigma_y^2$ e, por conseguinte, $B = Ib^2$. Nessa situação o vetor de predições dos valores genéticos se resume à expressão bem conhecida

$$\hat{a} = b^2(y - X\hat{\beta}),$$

em que a predição do valor genético é obtida pela regressão do valor genético em relação ao valor fenotípico, dada pelo produto do coeficiente de herdabilidade pela observação corrigida para os efeitos fixos de ambiente.

Todavia, se há parentesco entre os indivíduos, a matriz de variância genética passa a ser não-diagonal e dada por $G = A\sigma_a^2$, em que A é a matriz que descreve o parentesco entre os indivíduos, da mesma forma que a matriz de variância e covariância dos efeitos genéticos não-aditivos dada por $D = N\sigma_d^2$. Consequentemente, a matriz de variância e covariância fenotípica também será não diagonal e dada por $V = A\sigma_a^2 + N\sigma_d^2 + I\sigma_e^2$.

Nessa situação a predição do valor genético de um determinado indivíduo levará em conta as informações referentes aos seus parentes, além da sua própria, por meio de regressores específicos para cada tipo de informação. Esses regressores estão dispostos na matriz B , de tal forma que a i -ésima linha de B corresponde ao conjunto de regressores para as informações associadas ao i -ésimo indivíduo.

Deve-se observar que não apenas as informações dos parentes são utilizadas na predição do valor genético de determinado indivíduo. As informações referentes àqueles indivíduos não aparentados ao i -ésimo indivíduo, mas que contribuíram na formação do valor genético de um parente dele, serão regredidas para ajustar a predição de seu valor genético. Isso acontece, por exemplo, se dispomos das informações de touros, de seus filhos e das vacas que produziram tais filhos. Nesse caso as informações das vacas são computadas para ajustar a predição dos valores genéticos dos touros, mesmo não sendo aparentadas com eles.

Além dessa vantagem, a utilização da estrutura de parentesco permite a predição de valores genéticos de indivíduos nos quais não se fez observações. Para que isso seja possível é necessário apenas o uso de uma matriz Z , de incidência de valores genéticos, com dimensão linha igual ao número de observações e dimensão coluna igual ao número de indivíduos a serem avaliados, de tal forma que a matriz B de regressores passa a ser dada por

$$B = GZ'V^{-1}.$$

Da mesma forma, é necessário o uso da matriz Z se a característica envolvida apresenta medidas repetidas.

MELHORANDO A ESTIMAÇÃO DOS EFEITOS AMBIENTAIS IDENTIFICÁVEIS

O fato da matriz V não apresentar estrutura diagonal, em função do parentesco entre os indivíduos, torna mais acurada a predição dos valores genéticos, entretanto, leva à necessidade de revisão da forma como foram estimados os efeitos ambientais. As pressuposições feitas para o estabelecimento dos estimadores dos efeitos fixos não consideraram a estrutura não diagonal de V . Se tal fato for levado em consideração torna-se necessário ponderar a soma de quadrados da parte aleatória das observações pela estrutura de correlações entre elas. Essa estrutura está descrita na matriz V . Assim, o que deve ser minimizado é $E(\epsilon'V^{-1}\epsilon)$ e não $E(\epsilon'\epsilon)$, então,

$$\begin{aligned} E(\epsilon'V^{-1}\epsilon) &= E[(y - X\beta)'V^{-1}(y - X\beta)] \\ &= E(y'V^{-1}y - 2\beta'X'V^{-1}y + \beta'X'V^{-1}X\beta) \end{aligned}$$

$$\begin{aligned} \delta[E(\varepsilon'V^{-1}\varepsilon)]/\delta\beta' &= \delta[E(y'V^{-1}y - 2\beta'X'V^{-1}y + \beta'X'V^{-1}X\beta)]/\delta\beta' \\ &= -2X'V^{-1}y + 2X'V^{-1}X\beta \\ &\quad -X'V^{-1}y + X'V^{-1}X\beta = 0 \\ X'V^{-1}X\beta &= X'V^{-1}y. \end{aligned}$$

Considerando que sejam feitas adequadas restrições se $X'V^{-1}X$ não tem posto completo, então,

$$\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}y$$

Agora a estimação dos efeitos fixos de ambiente se processa, coerentemente, por meio do método de Quadrados Mínimos Generalizados (GLS) e não Quadrados Mínimos Ordinários (OLS) como antes.

Assim, o procedimento para avaliação genética consistirá na aplicação do método GLS, produzindo Melhores Estimadores Lineares Não Viesados (BLUE) dos efeitos fixos de ambiente, e em seguida, usando tais estimadores, ajustar as observações e regredi-las obtendo-se os Melhores Preditores Lineares Não Viesados (BLUP) dos valores genéticos.

$$\begin{aligned} \hat{\beta} &= (X'V^{-1}X)^{-1}X'V^{-1}y \\ \hat{a} &= GZ'V^{-1}(y - X\hat{\beta}). \end{aligned}$$

As dificuldades na aplicação dessas equações residem na obtenção da inversa de V .

USANDO AS EQUAÇÕES DE MODELOS MISTOS

Uma forma alternativa de obtenção de Melhores Estimadores Lineares Não Viesados (BLUE) para os efeitos fixos de ambiente e de Melhores Preditores Lineares Não Viesados (BLUP) para os valores genéticos pode ser estabelecida por meio da maximização da função densidade de probabilidade conjunta das observações e dos valores genéticos, como segue.

O modelo inicialmente definido

$$y_{ij} = f_i + a_{ij} + d_{ij} + E_{ij},$$

agora é reescrito como

$$y_{ij} = f_i + a_{ij} + e_{ij},$$

Sendo $e_{ij} = d_{ij} + E_{ij}$.

Na forma matricial, tem-se

$$y = X\beta + Za + e,$$

cujos elementos já foram descritos.

Assume-se que y , a e e apresentam a seguinte distribuição

$$\begin{bmatrix} y \\ a \\ e \end{bmatrix} \sim \left\{ \begin{bmatrix} X\beta \\ 0 \\ 0 \end{bmatrix}; \begin{bmatrix} V & ZG & R \\ GZ' & G & 0 \\ R & 0 & R \end{bmatrix} \right\}$$

em que $V = ZGZ' + R$, sendo $G = A\sigma_a^2$ e $R = I\sigma_e^2$.

Observe-se que nesta definição despreza-se a estrutura não diagonal da matriz de variância e covariância dos efeitos genéticos não aditivos, o que em princípio não terá grande importância, a menos que a estrutura de parentesco apresente quantidade considerável de irmãos completos.

A função densidade de probabilidade conjunta das observações e dos valores genéticos, $f(y, a)$, pode ser escrita como o produto da função densidade de probabilidade de y , dado que se conhece a , pela função densidade de probabilidade de a .

$$f(y, a) = f(y|a)f(a)$$

$$f(y, a) = \frac{1}{\sqrt{(2\pi)^n |R|}} e^{-\frac{1}{2}[(y-X\beta-Za)'R^{-1}(y-X\beta-Za)]} \cdot \frac{1}{\sqrt{(2\pi)^n |G|}} e^{-\frac{1}{2}[(a-0)'G^{-1}(a-0)]}$$

Para se proceder a maximização pode-se aplicar a transformação logarítmica para facilitar a manipulação algébrica. Assim,

$$L = \frac{1}{2} [2n \ln(2\pi) - \ln|R| + \ln|G| + y'R^{-1}y - 2y'R^{-1}X\beta + 2y'R^{-1}Za + 2\beta'X'R^{-1}Za + \beta'X'R^{-1}X\beta + a'Z'R^{-1}Za + a'G^{-1}a]$$

Derivando-se em relação a β e a tem-se

$$\begin{bmatrix} \delta L / \delta \beta \\ \delta L / \delta a \end{bmatrix} = \begin{bmatrix} -X'R^{-1}y + X'R^{-1}X\beta + X'R^{-1}Za \\ -Z'R^{-1}y + Z'R^{-1}X\beta + Z'R^{-1}Za + G^{-1}a \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} X'R^{-1}X\beta + X'R^{-1}Za \\ Z'R^{-1}X\beta + (Z'R^{-1}Z + G^{-1})a \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix}$$

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & (Z'R^{-1}Z + G^{-1}) \end{bmatrix} \begin{bmatrix} \beta \\ a \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix}$$

Essas são as Equações de Modelos Mistos (MME) que permitem a obtenção de BLUE para os efeitos fixos de ambiente e BLUP para os valores genéticos.

A solução desse sistema pode ser obtida por absorção resultando em

$$\hat{\beta} = \left\{ X' \left[R^{-1} - R^{-1}Z(Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1} \right] X \right\}^{-1} X' \left[R^{-1} - R^{-1}Z(Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1} \right] y$$

$$\hat{a} = (Z'R^{-1}Z + G^{-1})^{-1} ZR^{-1} (y - X\hat{\beta}).$$

Essas equações fornecem os mesmos resultados de

$$\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}y$$

$$\hat{a} = GZ'V^{-1}(y - X\hat{\beta}),$$

como pode ser visto em MARTINS et al. (1993).

Na Tabela 12.6 são apresentadas as predições dos valores genéticos e a classificação dos animais do exemplo de ilustração.

TABELA 12.6. Predição do valor genético de todos os animais utilizando modelo animal

NÚMERO	VALOR GENÉTICO	CLASSE	NÚMERO	VALOR GENÉTICO	CLASSE
1	-4.75	44	25	0.66	22
2	3.66	9	26	6.02	3
3	2.98	12	27	0.53	24
4	-1.18	34	28	-2.26	40
5	-2.00	38	29	-5.98	48
6	0.07	26	30	4.34	5
7	-1.64	36	31	-0.44	30
8	1.85	18	32	-1.61	35
9	1.33	20	33	4.16	6
10	0.48	25	34	2.13	16
11	1.22	21	35	7.35	1
12	-0.07	27	36	6.30	2
13	-0.55	32	37	3.83	7
14	3.02	11	38	2.18	15
15	-0.64	33	39	-4.89	46
16	-4.14	43	40	-0.26	28
17	-5.37	47	41	-2.19	39
18	-2.27	41	42	-1.92	37
19	-4.83	45	43	1.92	17
20	4.61	4	44	2.19	14
21	3.82	8	45	0.55	23
22	2.56	13	46	-0.55	31
23	3.32	10	47	-3.44	42
24	1.39	19	48	-0.35	29

CONSIDERAÇÕES FINAIS

A diferença entre os modelos

$$y = X\beta + \varepsilon \text{ e } y = X\beta + Za + e$$

está apenas no fato de que no primeiro o valor genético não é considerado de forma explícita. Contudo, sendo um efeito aleatório é considerado na matriz de variância e covariância das observações dada por $V = ZGZ' + R$. Aliado a isto há ortogonalidade entre a parte fixa e a parte aleatória das observações, de tal sorte que a estimação dos efeitos fixos independe da forma como os efeitos aleatórios são considerados, se decompostos como $Za + e$ ou na forma compacta como ε .

A vantagem na utilização do segundo modelo reside no fato de não ser necessário obter a inversa de V e, além disso, as inversas das matrizes de variâncias e covariâncias genética e residual são facilmente obtidas porque R é diagonal e $G^{-1} = A^{-1}\sigma_a^{-2}$, e existem algoritmos eficientes para obtenção direta da inversa de A , a matriz de parentesco.

Uma dificuldade pode persistir se a solução for obtida por absorção, porque será necessário obter-se $(Z'R^{-1}Z + G^{-1})^{-1}$ que, apesar de apresentar rápida convergência em métodos iterativos de obtenção de inversas, é, frequentemente, de grandes dimensões. Todavia, tais métodos iterativos podem ser usados diretamente no sistema de Equações de Modelos Mistos para obtenção dos vetores β , de efeitos fixos de ambiente, e a , de valores genéticos.

Se o interesse é apenas a predição dos valores genéticos de touros, a definição de um modelo de reprodutor, com esse efeito tomado como aleatório, pode ser feita e as MME fornecerão BLUP dos valores genéticos dos touros, com expressiva redução do esforço computacional (MRODE, 1996). Porém, neste caso deve-se atentar para o fato de que tais predições não estarão ajustadas para os valores genéticos das vacas com as quais os touros foram acasalados.

Para que as predições dos valores genéticos dos touros sejam corrigidas para os valores genéticos das vacas com eles acasaladas, tendo-se mesmo assim um menor esforço computacional, torna-se necessário a utilização do Modelo Animal Reduzido (RAM) que consiste numa manipulação algébrica do Modelo Animal, em que as equações referentes aos indivíduos que não têm filhos são absorvidas naquelas dos indivíduos que têm descendentes (MARTINS et al., 1997). Neste caso, são obtidas predições dos valores genéticos de touros e vacas.

A metodologia de modelos mistos é robusta em suas aplicações, no sentido de que mesmo quando as condições dos dados coletados não satisfazem plenamente as premissas envolvidas na sua concepção, os resultados obtidos são pouco distorcidos. Isto, contudo, não nos exime da responsabilidade de coletar e informar corretamente os dados de campo, os quais são a matéria prima no processo de avaliação genética.

A despeito dessa breve explanação sobre o uso das MME para predições de valores genéticos pode-se observar sua importância como ferramenta para orientar as decisões de seleção nos rebanhos.

FONTES DE REFERÊNCIA

- HENDERSON, C.R. Applications of linear models in animal breeding. Guelph, Canadá, University of Guelph Press, 1984. 462p.
- LOPES, P.S.; MARTINS, E.N.; SILVA, M.A.; RAGGI, L.A. Métodos de resolução de sistemas de equações lineares. Viçosa, MG, UFV, Imprensa Universitária, 1993. 55p.
- MARTINS, E.N.; LOPES, P.S.; SILVA, M.A.; REGAZZI, A.J. Modelo linear misto. Viçosa, MG, UFV, Imprensa Universitária, 1993. 46p.
- MARTINS, E.N. Uso de modelos mistos no melhoramento animal. In: SIMPÓSIO INTERNACIONAL DE PRODUÇÃO DE NÃO-RUMINANTES, 1994. Maringá, PR. *Anais...* Maringá, Sociedade Brasileira de Zootecnia, 1994. p.25-51.
- MARTINS, E.N.; LOPES, P.S.; SILVA, M.A.; TORRES JUNIOR, R.A.A. Uso de modelos mistos na avaliação genética animal. Viçosa, MG, UFV, Imprensa Universitária, 1997. 100p.
- MRODE, R.A. Linear models for the prediction of animal breeding values. Wallingford, UK, CAB International, 1996. 187p.