

21. НАИЛУЧШИЙ ЛИНЕЙНЫЙ НЕСМЕЩЕННЫЙ ПРОГНОЗ

Мощным средством анализа популяций сельскохозяйственных животных является метод BLUP (см. также [55,122,136,140-144,149, 150, 152,154]). BLUP - это аббревиатура английского выражения:

- **Best** - наилучший;
- **Linear** - линейный;
- **Unbiased** - несмещенный;
- **Prediction** - прогноз.

Методология BLUP была разработана американским ученым Ч.Хендерсоном во второй половине XX века. В основе BLUP - *уравнения смешанной модели (Mixed Model Equations, ММЕ)*, которые относительно просты для построения и решения.

Отличительной особенностью BLUP является то, что, во-первых, он хорошо приспособлен для анализа несбалансированных данных; во-вторых, очень гибкий и универсальный; в-третьих, позволяет *одновременно* получать наилучшие линейные несмещенные оценки (BLUE) фиксированных эффектов и наилучший линейный несмещенный прогноз (BLUP) рандомизированных эффектов.

Благодаря своим замечательным свойствам, метод BLUP нашел широкое применение в племенном животноводстве. Так, в настоящее время не только *национальные* системы оценки племенной ценности молочного скота экономически развитых стран, но и *интернациональная* оценка, проводимая Центром INTERBULL, базируются на теории BLUP.

Следует отметить, что BLUP используют не только в разведении сельскохозяйственных животных, но и в других сферах: в экономике, астрономии, социологии, спорте (прогноз счета в футбольных играх) и т.п.

21.1. Понятие «племенная ценность»

Животное получает от своих родителей не генотипы, а гены. Генотип особи создается из комбинаций половины набора генов отца и половины набора генов матери, фиксируется при зачатии и остается константным в течение всей ее жизни (не считая соматических мутаций). Какие гены получит потомок от родителей – дело случая.

В каждом новом поколении генотипы образуются заново. Поэтому для характеристики наследственных качеств животного нужна такая мера, которая бы относилась не к генотипам, а к генам. Такой мерой является *средний эффект гена*.

Средний эффект гена - это усредненное отклонение от популяционного среднего тех особей, которые получили определенный ген от одного из родителей, при условии, что ген, полученный от другого родителя, как бы случайно выбран из пула генов.* Большинство хозяйственно полезных признаков - полигенные. Поэтому суммарный средний эффект генов, которые имеет животное и которые может передать своему потомству, отождествляют с *генетической ценностью* (или *племенной ценностью* - Breeding Value, BV). Но прямое измерение генетической ценности невозможно. Её оценивают по фенотипу самого животного и/или его родственников. Полученное значение называют *оценкой племенной ценности* (Estimated Breeding Value, EBV) животного. Она должна максимально коррелировать с *истинной племенной ценностью* (BV), под которой подразумевают генетическую ценность животного.

Так как средний эффект генов родителей определяет средний эффект генов потомка, то племенная ценность животного характеризуется генами, которые оно может передать своему потомству. Отсюда, оценка племенной ценности – это попытка *предсказать* средний эффект *выборки* генов, которые данное животное передаст потомству.

Выше отмечалось, что генетическая ценность животного (G) включает аддитивный эффект генов (A), эффекты доминирования (D) и эпистаза (I). Эффекты доминирования и эпистаза не передаются потомству. Поэтому для целей селекции наиболее важно знать аддитивный эффект генов, так как только он передается потомству. Сумма по всем локусам всех аддитивных генетических эффектов есть *аддитивная генетическая ценность* (A) животного. Аддитивную генетическую ценность по одному

* Для пояснения допустим, что все гаметы, несущие ген A_1 , случайно объединяются с гаметами всей популяции. Тогда для полученных таким путем генотипов усредненное отклонение от популяционного среднего и есть средний эффект гена A_1 . Средний эффект гена зависит как от качества самого гена, так и от его частоты в популяции.

признаку любой особи можно определить как аддитивный эффект всех генов, влияющих на этот признак.

Аддитивную генетическую ценность (A) называют также *аддитивной племенной ценностью* животного. Так как родитель передает потомку только половину своих генов, то часто используется термин «*передающая способность*» (Transmitting Ability, TA) или «*предсказанная разность*» (Predicted Difference, PD). Передающая способность - это значение среднего аддитивного эффекта генов, которые будут переданы потомству:

$$TA = PD = \left(\frac{1}{2}\right) \times BV = \left(\frac{1}{2}\right) \times A.$$

Если обозначить истинную аддитивную племенную ценность отца как A_{sire} (= сумме аддитивных эффектов всех генов отца), то аддитивная генетическая ценность отцовских гамет равна в среднем $\frac{1}{2}A_{\text{sire}}$ (передается только половина генов). Однако *каждая* гамета содержит *различную* случайную выборку генов (т.н. «*менделевская выборка*») и будет, следовательно, иметь аддитивную генетическую ценность, отличную от $\frac{1}{2}A_{\text{sire}}$. Фактически аддитивная генетическая ценность гамет, продуцируемых отцом, имеет нормальное распределение со средним равным $\frac{1}{2}A_{\text{sire}}$ и дисперсией $\frac{1}{4}\sigma_A^2$ (то есть $\sim N(\frac{1}{2}A_{\text{sire}}, \frac{1}{4}\sigma_A^2)$). Аналогично, аддитивная генетическая ценность гамет, продуцируемых матерью, имеет нормальное распределение со средним равным $\frac{1}{2}A_{\text{dam}}$ и дисперсией $\frac{1}{4}\sigma_A^2$. Исходя из этого, *истинная* племенная ценность потомка (A_{off}) может быть выражена следующим образом:

$$A_{\text{off}} = \frac{1}{2}A_{\text{sire}} + \frac{1}{2}A_{\text{dam}} + RA,$$

где RA - рандомизированный эффект спаривания, обусловленный как случайностью выборки генов отца и матери (менделевская выборка) при образовании гамет, так и случайностью при слиянии гамет в зиготу.

Значения RA имеют нормальное распределение со средним равным нулю и дисперсией $\frac{1}{2}\sigma_A^2$:

$$RA \sim N(0, \frac{1}{2}\sigma_A^2).$$

Менделевской выборкой объясняется тот факт, что полные братья и сестры являются генетически не идентичными.

На основе представленных выше уравнений, варианса (Var или σ^2) аддитивной генетической ценности животных в популяции может быть разложена на вариансу племенной ценности в отцовском поколении, вариансу племенной ценности в поколении матери и вариансу эффектов менделевской выборки:

$$\begin{aligned}\text{Var}(A_{\text{off}}) &= \text{Var}(\frac{1}{2}A_{\text{sire}}) + \text{Var}(\frac{1}{2}A_{\text{dam}}) + \text{Var}(RA) = \\ &= \frac{1}{4}\sigma_A^2 + \frac{1}{4}\sigma_A^2 + \frac{1}{2}\sigma_A^2 = \\ &= \sigma_A^2.\end{aligned}$$

Из отношения следует, что варианса племенной ценности потомства текущего поколения (σ_A^2) определяется на 25% различиями в племенных ценностях отцов, на 25% - различиями в племенных ценностях матерей и на 50% генетической комбинаторикой при образовании родительских гамет и в дальнейшем - зигот. Поэтому от родителей с высокой аддитивной племенной ценностью можно с большей вероятностью ожидать потомства с высокой аддитивной генетической ценностью и, следовательно, с более высокой продуктивностью (т.к. аддитивная генетическая ценность является частью фенотипической ценности).

Оценка племенной ценности (EBV) - это количественная характеристика наследственных качеств животного. Она выражает - насколько в среднем продуктивность будущих потомков животного будет выше или ниже среднего по популяции. Племенная ценность может быть выражена в отклонениях или в процентах от средней продуктивности по популяции, или в сравнении со средней племенной ценностью животных той же самой племенной категории, или относительно какой-либо базовой генетической группы.

21.2. Точность, надежность, ошибка

Так как истинная племенная ценность животных *прогнозируется*, то EBV имеет свою ошибку, которая равна EBV минус истинная BV. Если $EBV = \hat{A}$, а $BV = A$, то ошибка прогноза

будет $(\hat{A}-A)$. Несмещенность прогноза состоит в том, что *усредненная* ошибка прогноза по всем животным равна нулю, т.е. $E(\hat{A}-A)=0$ (здесь E – символ математического ожидания).

Точность. *Точность* прогноза есть корреляция между истинной BV и EBV

$$r_{A\hat{A}} = \frac{\sigma_{A\hat{A}}}{\sigma_A \sigma_{\hat{A}}}.$$

Из теории селекции следует, что $\sigma_{A\hat{A}} = \sigma_{\hat{A}}^2$. Тогда формула точности прогноза редуцируется до

$$r_{A\hat{A}} = \frac{\sigma_{\hat{A}}}{\sigma_A}.$$

Точность прогноза - важное понятие в селекции животных. Чем точнее произведена оценка племенной ценности, тем более точно можно отобрать животных с самой высокой истинной племенной ценностью.

Надежность. Квадрат точности есть *надежность* (достоверность) оценки BV . Надежность можно также определить как доля аддитивной генетической вариации, которая объясняется EBV , то есть

$$r_{A\hat{A}}^2 = \frac{\sigma_{\hat{A}}^2}{\sigma_A^2}.$$

Преобразование формулы дает:

$$\sigma_{\hat{A}}^2 = r_{A\hat{A}}^2 \sigma_A^2.$$

Таким образом, *варианса оценок племенной ценности равна произведению надежности на вариацию истинной племенной ценности* (аддитивную генетическую вариацию). Так как аддитивная генетическая вариация более или менее константна, то формула указывает на важность точности. Большая точность - это большая вариация и рассеяние оценок племенной ценности животных в популяции и, следовательно, большая возможность для дифференциации генетически превосходных, посредственных и плохих животных.

Если $r_{\hat{A}\hat{A}}=0$, то варианса \hat{A} тоже равна нулю и, таким образом, все животные в популяции имеют идентичные EBV. Очевидно, что в этом случае эффективный отбор невозможен. Если точность равна 1, то варианса \hat{A} равна генетической вариансе. В этом *смысл* племенной оценки, так как в этом случае BV прогнозируется безошибочно и $\hat{A}=A$ (EBV=BV).

Ошибка. Варианса ошибки прогноза (Prediction Error Variance, PEV) равна $(1-r_{\hat{A}\hat{A}}^2)\sigma_A^2$. Ошибка прогноза имеет нормальное распределение со средней равной 0 и вариансой $(1-r_{\hat{A}\hat{A}}^2)\sigma_A^2$:

$$\text{ошибка прогноза} \sim N(0, (1-r_{\hat{A}\hat{A}}^2)\sigma_A^2).$$

Корень квадратный из PEV дает *стандартную ошибку прогноза* (Standard Error Prediction, SEP):

$$\begin{aligned} \text{SEP} &= \sqrt{\text{PEV}} \\ &= \sigma_A \sqrt{1-r_{\hat{A}\hat{A}}^2}. \end{aligned}$$

Таким образом, для снижения ошибки прогноза необходимо использовать такие доступные методы, которые максимизируют $r_{\hat{A}\hat{A}}$ при имеющемся количестве информации. Другими словами, методы должны наиболее эффективно использовать данные.

Источниками информации для оценки племенной ценности могут быть родители и более отдаленные предки, сам пробанд, боковые родственники и потомки. Эти данные поступают в определенной временной последовательности. В процессе оценки их интегрируют в *комбинированную* племенную ценность (для повышения точности прогноза). На заключительном этапе оценки комбинированной племенной ценности по отдельным признакам с помощью экономических весов интегрируют в *селекционный индекс* по комплексу признаков.

21.3. Модель для оценки по собственным показателям

Для BLUP требуется статистическая модель, которая описывала бы наблюдения. Модель - самая важная часть анализа. Результаты применения BLUP по неполной или некорректной модели могут быть хуже, чем применение другой техники

расчетов по корректной модели. Следует стремиться использовать лучшую модель в пределах физических возможностей компьютеров. В то же время, исследователь должен доказать, что выбранная им для анализа статистическая модель правильная.

Ниже рассматриваются простые модели. Это сделано для того, чтобы показать методологию BLUP, а также некоторые возможности и различия между моделями.

Предположим, что имеются следующие данные продуктивности трех молочных коров (собственные показатели):

№ коровы	№ лактации	Год отела	Удой, кг
1	1	2002	5630
	2	2003	6470
2	1	2003	5810
	2	2004	6290
3	1	2004	5760

Для этих данных статистическая модель имеет вид:

$$y_{ijklm} = L_i + Y_j + C_k + e_{ijklm},$$

где y_{ijklm} - удой за лактацию;

L_i - эффект номера лактации (коровы имеют тенденции с увеличением числа отелов давать больше молока);

Y_j - эффект года отела (в разные годы условия среды различны: может измениться кормление, климат может стать более влажным, или может быть построен новый скотный двор);

C_k - эффект коровы;

e_{ijklm} - влияние неучтенных факторов (остаток, ошибка).

Остаток (e_{ijklm}) включает эффекты факторов, которые не были включены в модель. Допускается, что эти эффекты, рандомизированные для каждого наблюдения, и что их среднее равно нулю. Обычно влияние остаточных эффектов значительное. Поэтому стараются уменьшить их путем подбора модели.

Эффект коровы (C_k) считают, как правило, рандомизированным. Он может быть разделен на две части:

$$C_k = A_k + P_k,$$

где A_k - аддитивный генетический эффект коровы;

P_k - перманентный (постоянный) эффект среды; этот эффект связан с каждой лактацией коровы, но не является генетическим и не передается потомству (далее допускается, что он равен нулю).

21.4. Уравнения смешанной модели

В основе BLUP - уравнения смешанной модели (ММЕ), в формировании которых можно выделить следующие этапы.

1 этап. Составление списка элементов модели и суммарной продуктивности по каждому элементу:

Элементы модели	Наблюдения	Сумма	Символ
L_1 – 1-ая лактация	5630, 5810, 5760	17200	$Y_{1..}$
L_2 – 2-ая лактация	6470, 6290	12760	$Y_{2..}$
Y_1 – 2002 год отела	5630	5630	$Y_{.1.}$
Y_2 – 2003 год отела	6470, 5810	12280	$Y_{.2.}$
Y_3 – 2004 год отела	6290, 5760	12050	$Y_{.3.}$
C_1 – коровы № 1	5630, 6470	12100	$Y_{.1.}$
C_2 – корова № 2	5810, 6290	12100	$Y_{.2.}$
C_3 – корова № 3	5760	5760	$Y_{.3.}$

Проверка:

$$\begin{aligned}
 Y_{1..} + Y_{2..} &= Y_{.1.} + Y_{.2.} + Y_{.3.} = Y_{..1.} + Y_{..2.} + Y_{..3.} \\
 17200 + 12760 &= 5630 + 12280 + 12050 = 12100 + 12100 + 5760 \\
 29960 &= 29960 = 29960.
 \end{aligned}$$

2 этап. Формирование уравнений *наименьших квадратов*.

Сначала для каждой суммы записывают модель (без остатков):

- для L_1

$$\begin{aligned}
 y_{1..} &= \begin{matrix} 5630 & L_1 + Y_1 + C_1 \\ +5810 & +L_1 + Y_2 + C_2 \\ +5760 & +L_1 + Y_3 + C_3 \end{matrix} \\
 &= 17200 = 3L_1 + 1Y_1 + 1Y_2 + 1Y_3 + 1C_1 + 1C_2 + 1C_3;
 \end{aligned}$$

- для L_2

$$\begin{aligned}
 y_{2..} &= \begin{matrix} 6470 & L_2 + Y_2 + C_1 \\ +6290 & +L_2 + Y_3 + C_2 \end{matrix} \\
 &= 12760 = 2L_2 + 1Y_2 + 1Y_3 + 1C_1 + 1C_2;
 \end{aligned}$$

⋮

- для C_1

$$\begin{aligned} y_{..1} &= \begin{matrix} 5630 \\ +6470 \end{matrix} = \begin{matrix} L_1 + Y_1 + C_1 \\ +L_2 + Y_2 + C_1 \end{matrix} \\ &= 12100 = 1L_1 + 1L_2 + 1Y_1 + 1Y_2 + 2C_1; \end{aligned}$$

и т.п.

Полученные уравнения переносят в таблицу:

	L_1	L_2	Y_1	Y_2	Y_3	C_1	C_2	C_3	Сумма
$y_{1..}$	3	0	1	1	1	1	1	1	17200
$y_{2..}$	0	2	0	1	1	1	1	0	12760
$y_{.1.}$	1	0	1	0	0	1	0	0	5630
$y_{.2.}$	1	1	0	2	0	1	1	0	12280
$y_{.3.}$	1	1	0	0	2	0	1	1	12050
$y_{..1}$	1	1	1	1	0	2	0	0	12100
$y_{..2}$	1	1	0	1	1	0	2	0	12100
$y_{..3}$	1	0	0	0	1	0	0	1	5760

Числа из таблицы записывают в матричной форме для построения уравнений, которые известны как *уравнения наименьших квадратов* (LS-уравнения):

$$\begin{bmatrix} 3 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 2 & 0 & 1 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 2 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 2 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 2 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 & 2 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} L_1 \\ L_2 \\ Y_1 \\ Y_2 \\ Y_3 \\ C_1 \\ C_2 \\ C_3 \end{bmatrix} = \begin{bmatrix} 17200 \\ 12760 \\ 5630 \\ 12280 \\ 12050 \\ 12100 \\ 12100 \\ 5760 \end{bmatrix}.$$

3 этап. Формирование уравнений смешанной модели.

К коэффициентам уравнений для коров (C_k) суммируют отношение varianс

$$k = \sigma_e^2 / \sigma_c^2.$$

Предполагается, что varianсы σ_e^2 и σ_c^2 известны. Обычно их оценивают по исходным данным (считается, что varianсы во времени изменяются незначительно).

Если наследуемость удоя равна 0,25 и

$$\sigma_c^2 = \sigma_A^2 = h^2 \sigma_y^2,$$

а

$$\sigma_e^2 = \sigma_y^2 - \sigma_A^2 = \sigma_y^2 - h^2 \sigma_y^2 = (1 - h^2) \sigma_y^2,$$

то

$$\begin{aligned} k &= \sigma_e^2 / \sigma_c^2 = (1 - h^2) \sigma_y^2 / h^2 \sigma_y^2 = \\ &= 0,75 \sigma_y^2 / 0,25 \sigma_y^2 \\ &= 3. \end{aligned}$$

Подсуммирование $k=3$ приводит к формированию ММЕ-уравнений:

$$\begin{bmatrix} 3 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 2 & 0 & 1 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 2 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 2 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 2+3 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 & 2+3 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1+3 \end{bmatrix} \begin{bmatrix} L_1 \\ L_2 \\ Y_1 \\ Y_2 \\ Y_3 \\ C_1 \\ C_2 \\ C_3 \end{bmatrix} = \begin{bmatrix} 17200 \\ 12760 \\ 5630 \\ 12280 \\ 12050 \\ 12100 \\ 12100 \\ 5760 \end{bmatrix}.$$

4 этап. Решение ММЕ-уравнений.

ММЕ-уравнения могут быть представлены в следующем виде:

$$W b = t,$$

где

$$W = \begin{bmatrix} 3 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 2 & 0 & 1 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 2 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 2 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 5 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 & 5 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 4 \end{bmatrix};$$

$$b' = [L_1 \ L_2 \ Y_1 \ Y_2 \ Y_3 \ C_1 \ C_2 \ C_3];$$

$$t' = [17200 \ 1276 \ 5630 \ 12280 \ 12050 \ 12100 \ 12100 \ 5760].$$

Для решения ММЕ необходимо инвертировать матрицу W и умножить обе части уравнения на W^{-1} :

$$W^{-1}(W b) = W^{-1} t \quad \text{и}$$

$$(W^{-1} W) b = W^{-1} t,$$

но $W^{-1} W = I =$ единичная матрица, тогда

$$I b = W^{-1} t$$

$$b = W^{-1} t.$$

Проблема: Матрица W не инвертируемая, т.к. имеет место зависимость между ее строками:

строка 1 + строка 2 = строка 3 + строка 4 + строка 5.

Чтобы устранить эту зависимость, вводят какое-либо ограничение. Так как существует только одна зависимость в матрице W , то необходимо одно ограничение, например, оценка эффекта первой лактации равна нулю.

Для того чтобы $L_1 = 0$, необходимо к матрице W добавить одну строку и один столбец, которые покажут, какое решение необходимо сделать нулевым. Вектора b и t также следует расширить на одну строку. ММЕ-уравнения в окончательном виде:

$$\begin{bmatrix} 3 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 2 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 2 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 2 & 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 5 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 & 5 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 4 & 0 \\ \hline 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} L_1 \\ L_2 \\ Y_1 \\ Y_2 \\ Y_3 \\ C_1 \\ C_2 \\ C_3 \\ \hline ML \end{bmatrix} = \begin{bmatrix} 17200 \\ 12760 \\ 5630 \\ 12280 \\ 12050 \\ 12100 \\ 12100 \\ 5760 \\ \hline 0 \end{bmatrix}$$

↑
добавлено
к W
↑
добавлено
к b
↑
добавлено
к t

Новая матрица W больше не имеет зависимостей и может быть инвертирована:

$$\begin{bmatrix} \hat{L}_1 \\ \hat{L}_2 \\ \hat{Y}_1 \\ \hat{Y}_2 \\ \hat{Y}_3 \\ \hat{C}_1 \\ \hat{C}_2 \\ \hat{C}_3 \\ ML \end{bmatrix} = \begin{bmatrix} 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 1,000 \\ 0,000 & 1,167 & 0,167 & -0,500 & -0,667 & -0,167 & 0,000 & 0,167 & 1,000 \\ 0,000 & 0,167 & 1,315 & 0,093 & -0,074 & -0,315 & -0,037 & 0,019 & -1,000 \\ 0,000 & -0,500 & 0,093 & 0,870 & 0,370 & -0,093 & -0,148 & -0,093 & -1,000 \\ 0,000 & -0,667 & -0,074 & 0,370 & 1,037 & 0,074 & -0,148 & -0,259 & -1,000 \\ 0,000 & -0,167 & -0,315 & -0,093 & 0,074 & 0,315 & 0,037 & -0,019 & 0,000 \\ 0,000 & 0,000 & -0,037 & -0,148 & -0,148 & 0,037 & 0,259 & 0,037 & 0,000 \\ 0,000 & 0,167 & 0,019 & -0,093 & -0,259 & -0,019 & 0,037 & 0,315 & 0,000 \\ 1,000 & 1,000 & -1,000 & -1,000 & -1,000 & 0,000 & 0,000 & 0,000 & 0,000 \end{bmatrix} \begin{bmatrix} 17200 \\ 12760 \\ 5630 \\ 12280 \\ 12050 \\ 12100 \\ 12100 \\ 5760 \\ 0 \end{bmatrix} .$$

Решение ММЕ дает следующие оценки:

$$\begin{aligned} \hat{L}_1 &= 0; & \hat{Y}_1 &= 5622,7778; & \hat{C}_1 &= 7,2222; \\ \hat{L}_2 &= 595; & \hat{Y}_2 &= 5846,1111; & \hat{C}_2 &= -14,4444; \\ ML &= 0; & \hat{Y}_3 &= 5731,1111; & \hat{C}_3 &= 7,2222. \end{aligned}$$

Что означают эти оценки?

$\hat{L}_2 = 595$ - это оценка различий между L_2 и L_1 , т.е. коровы, которые имели вторую лактацию, давали на 595 кг молока больше, чем первотелки.

$\hat{Y}_1 = 5622,7778$ - это оценка $\mu + Y_1 + L_1$, т.е. оценка среднего удоя коров по первой лактации, которые отелились в 2002 году;

$\hat{Y}_2 = 5846,1111$ - это оценка $\mu + Y_2 + L_1$;

$\hat{Y}_3 = 5731,1111$ - это оценка $\mu + Y_3 + L_1$.

Следует заметить, что

$$\hat{Y}_2 - \hat{Y}_1 = 5846,1111 - 5622,7778 = 223,3333 \text{ кг},$$

это оценка

$$(\mu + Y_2 + L_1) - (\mu + Y_1 + L_1) = Y_2 - Y_1.$$

$\hat{C}_1 = 7,2222$ - это оценка племенной ценности (EBV) коровы № 1;

$\hat{C}_2 = -14,4444$ - EBV коровы № 2;

$\hat{C}_3 = 7,2222$ - EBV коровы № 3.

Для того чтобы узнать среднюю племенную ценность группы животных, необходимо знать модель, которая

использовалась, и процедуру расчетов. В этом примере в модели были учтены:

- эффект номера лактации (или возраста);
- эффект года отела и
- рандомизированный эффект коровы (племенная ценность).

Племенная ценность коров была рассчитана методом BLUP. Родственные отношения между коровами не учитывались. Для их учета необходимо использовать другую модель BLUP. Сравнение моделей, их оптимизация - это задача исследователей. В практической же селекции должны использоваться самые лучшие на данный момент времени модели.

21.5. Точность BLUP-оценок

При решении ММЕ необходимо инвертировать матрицу W (без или с ограничениями), т.е. получить матрицу W^{-1} . Диагональные элементы этой матрицы используют для расчета дисперсии ошибки прогноза генотипа животных - $\text{Var}(\hat{A} - A)$, где A - истинная аддитивная генетическая ценность (неизвестна); \hat{A} - прогноз аддитивной генетической ценности (=оценка племенной ценности, EBV).

Пусть D = диагональные элементы матрицы W^{-1} . Тогда для рассматриваемого примера получим вектор:

$$D' = (0 \ 1,1667 \ 1,3148 \ 0,8704 \ 1,0370 \ 0,3148 \ 0,2579 \ 0,3148 \ 0).$$

Интерес представляют диагональные элементы для \hat{C}_1 , \hat{C}_2 и \hat{C}_3 , в частности $(0,3148 \ 0,2593 \ 0,3148)$. Остальные элементы используют для расчетов стандартной ошибки (SEP) оценок \hat{L}_1 , \hat{L}_2 , \hat{Y}_1 , \hat{Y}_2 и \hat{Y}_3 .

Для коровы № 1:

$$\text{Var}(\hat{A} - A) = 0,3148 \sigma_e^2;$$

$$\text{SEP} = 0,3148^{0,5} \sigma_e = \sqrt{(1 - r_{AA}^2) \sigma_A^2}.$$

Тогда

$$0,3148 \sigma_e^2 = (1 - r_{AA}^2) \sigma_A^2;$$

$$0,3148 \frac{\sigma_e^2}{\sigma_A^2} = 1 - r_{AA}^2;$$

$$0,3148k = 1 - r_{AA}^2 \Rightarrow r_{AA}^2 = 1 - 0,3148k;$$

$$r_{AA} = \sqrt{1 - 0,3148k} = \sqrt{1 - 0,3148 \times 3} = \sqrt{0,0556} = 0,2358.$$

Для коровы № 2:

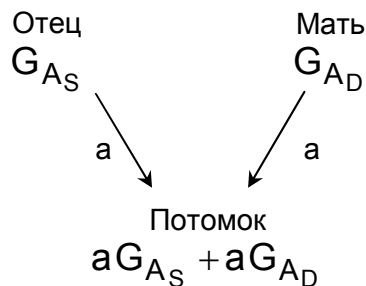
$$r_{AA} = \sqrt{1 - 0,2593 \times 3} = \sqrt{0,2221} = 0,4713.$$

Обе коровы имели по две лактации, но их r_{AA} различны, т.к. у коровы № 1 по первой лактации не было сверстниц.

21.6. Оценка по потомству

Проверку по потомству применяют для прогноза аддитивного генотипа по признакам, которые контролируются только у самок (например, молочная продуктивность).

Ниже, на рисунке иллюстрируется связь родителей с потомком:



где a = родство между родителем и потомком; обычно равно $\frac{1}{2}$, если родители не инбредны.

Проблема заключается в прогнозе аддитивной генетической ценности *производителя* (A_S) по продуктивности потомства.

Продуктивность потомка есть

$$y = \frac{1}{2} A_S + \frac{1}{2} A_D + E.$$

Средняя продуктивность n потомков есть

$$\bar{y} = \frac{1}{2} A_S + \frac{1}{2} \bar{A}_D + \bar{E}.$$

Допустим, что

- матери случайным образом выбраны из популяции и
- \bar{A}_D одно и то же для всех быков.

\bar{E} - это усредненная ошибка, которая также считается случайной. \bar{E} должно приближаться к нулю, особенно когда число потомков (n) увеличивается.

Пусть имеются следующие данные о продуктивности дочерей трех быков в двух стадах (количество молочного жира):

Бык	Стадо 1		Стадо 2		Итого	
	дочерей	сумма	дочерей	сумма	дочерей	сумма
1	4	480	6	840	10	1320
2	50	6250	-	-	50	6250
3	50	7250	50	6750	100	14000
Итого	104	13980	56	7590	160	21570

Эти данные могут быть описаны следующей моделью:

$$y_{ijk} = \mu + N_i + S_j + e_{ijk},$$

где μ - общая средняя;

N_i - эффект стада;

S_j - эффект отца, равный $\frac{1}{2} A_S$;

e_{ijk} - остаточный эффект.

Для составления уравнений смешанной модели (ММЕ) необходимо выполнить известные этапы:

1 этап. Составление списка элементов модели и расчет суммарной продуктивности по каждому элементу:

Элементы модели	Дочерей	Сумма	Символ
μ	180	21570	$y_{...}$
N_1 - стадо 1	104	13980	$y_{1..}$
N_2 - стадо 2	56	7590	$y_{2..}$
S_1 - бык №1	10	1320	$y_{.1.}$
S_2 - бык №2	50	6250	$y_{.2.}$
S_3 - бык №3	100	14000	$y_{.3.}$

2 этап. Запись уравнения для каждого элемента модели:

	μ	N_1	N_2	S_1	S_2	S_3	Сумма
$y_{...}$	160	104	56	10	50	100	21570
$y_{1..}$	104	104	0	4	50	50	13980
$y_{2..}$	56	0	56	6	0	50	7590
$y_{.1.}$	10	4	6	10	0	0	1320
$y_{.2.}$	50	50	0	0	56	0	6250
$y_{.3.}$	100	50	50	0	0	100	14000

Оформление LS-уравнений в матричном виде:

$$\begin{bmatrix} 160 & 104 & 56 & 10 & 50 & 100 \\ 104 & 104 & 0 & 4 & 50 & 50 \\ 56 & 0 & 56 & 6 & 0 & 50 \\ 10 & 4 & 6 & 10 & 0 & 0 \\ 50 & 50 & 0 & 0 & 50 & 0 \\ 100 & 50 & 50 & 0 & 0 & 100 \end{bmatrix} \begin{bmatrix} \mu \\ H_1 \\ H_2 \\ S_1 \\ S_2 \\ S_3 \end{bmatrix} = \begin{bmatrix} 21570 \\ 13980 \\ 7590 \\ 1320 \\ 6250 \\ 14000 \end{bmatrix}.$$

3 этап. Преобразование LS-уравнений в ММЕ.

К коэффициентам уравнений для быков S_j суммируется k :

$$k = \sigma_e^2 / \sigma_s^2.$$

Вариансы σ_e^2 и σ_s^2 предполагаются известными. Обычно они оцениваются по исходным данным.

При оценке по потомству $\sigma_s^2 = 0,25\sigma_A^2$ и

$$h^2 = 4\sigma_s^2 / (\sigma_s^2 + \sigma_e^2)$$

$$h^2(\sigma_s^2 + \sigma_e^2) = 4\sigma_s^2$$

$$h^2\sigma_e^2 = (4 - h^2)\sigma_s^2$$

$$\sigma_e^2 = (4 - h^2)\sigma_s^2 / h^2.$$

Тогда
$$k = \frac{\sigma_e^2}{\sigma_s^2} = \frac{(4 - h^2)}{h^2}.$$

Если для количества жира $h^2 = 0,25$, то тогда $k = 15$.

Подсуммирование k приводит к формированию ММЕ:

$$\begin{bmatrix} 160 & 104 & 56 & 10 & 50 & 100 \\ 104 & 104 & 0 & 4 & 50 & 50 \\ 56 & 0 & 56 & 6 & 0 & 50 \\ 10 & 4 & 6 & 10+15 & 0 & 0 \\ 50 & 50 & 0 & 0 & 50+15 & 0 \\ 100 & 50 & 50 & 0 & 0 & 100+15 \end{bmatrix} \begin{bmatrix} \mu \\ H_1 \\ H_2 \\ S_1 \\ S_2 \\ S_3 \end{bmatrix} = \begin{bmatrix} 21570 \\ 13980 \\ 7590 \\ 1320 \\ 6250 \\ 14000 \end{bmatrix}.$$

4 этап. Решение ММЕ.

В ММЕ-уравнениях строка 1=строка 2+строка 3. Чтобы устранить эту зависимость, необходимо ввести ограничение.

Пусть $\hat{H}_1 = 0$. Тогда новая система уравнений ММЕ примет вид:

$$\begin{bmatrix} 160 & 104 & 56 & 10 & 50 & 100 & 0 \\ 104 & 104 & 0 & 4 & 50 & 50 & 1 \\ 56 & 0 & 56 & 6 & 0 & 50 & 0 \\ 10 & 4 & 6 & 25 & 0 & 0 & 0 \\ 50 & 50 & 0 & 0 & 65 & 0 & 0 \\ 100 & 50 & 50 & 0 & 0 & 115 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{H}_1 \\ \hat{H}_2 \\ \hat{S}_1 \\ \hat{S}_2 \\ \hat{S}_3 \\ LM \end{bmatrix} = \begin{bmatrix} 21570 \\ 13980 \\ 7590 \\ 1320 \\ 6250 \\ 14000 \\ 0 \end{bmatrix}.$$

После инверсии матрицы коэффициентов и умножения ее на вектор правой стороны системы ММЕ

$$\begin{bmatrix} \hat{\mu} \\ \hat{H}_1 \\ \hat{H}_2 \\ \hat{S}_1 \\ \hat{S}_2 \\ \hat{S}_3 \\ LM \end{bmatrix} = \begin{bmatrix} 0,036 & 0,000 & -0,011 & -0,012 & -0,028 & -0,027 & -1,000 \\ 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 1,000 \\ -0,011 & 0,000 & 0,034 & -0,004 & 0,009 & -0,005 & 1,000 \\ -0,012 & 0,000 & -0,004 & 0,046 & 0,009 & 0,012 & 0,000 \\ -0,028 & 0,000 & 0,009 & 0,009 & 0,037 & 0,021 & 0,000 \\ -0,027 & 0,000 & -0,005 & 0,012 & 0,021 & 0,034 & 0,000 \\ -1,000 & 1,000 & 1,000 & 0,000 & 0,000 & 0,000 & 0,000 \end{bmatrix} \begin{bmatrix} 21570 \\ 13980 \\ 7590 \\ 1320 \\ 6250 \\ 14000 \\ 0 \end{bmatrix},$$

получим искомые оценки:

$$\begin{aligned} \hat{\mu} &= 134,5413 = \widehat{\mu + H_1}; & \hat{S}_1 &= 0,2672 = 0,5\hat{G}_{AS_1} = ETA_1; \\ \hat{H}_1 &= 0; & \hat{S}_2 &= -7,3394 = 0,5\hat{G}_{AS_2} = ETA_2; \\ \hat{H}_2 &= -5,3487 = \widehat{H_2 - H_1}; & \hat{S}_3 &= 7,0722 = 0,5\hat{G}_{AS_3} = ETA_3. \end{aligned}$$

5 этап. Расчет точности ETA.

Необходимы диагональные элементы (w_{jj}) инвертированной матрицы коэффициентов для уравнений по быкам 1, 2 и 3. Они составили 0,0456, 0,0369, 0,0342 соответственно.

Точность оценки:

$$r_{AA} = \sqrt{1 - k \times (\text{диагональный элемент})}.$$

Бык	ETA	n	w_{jj}	r_{TI}^2	r_{TI}
1	+0,27	10	0,0456	0,316	0,562
2	-7,34	50	0,0369	0,446	0,668
3	+7,07	100	0,0342	0,488	0,698

Варианса ошибки прогноза:

$$\text{Var}(\hat{A} - A) = (\text{диагональный элемент в } W^{-1}) \sigma_e^2.$$

Пусть для количества молочного жира $\sigma_e^2 = 600 \text{ кг}^2$, тогда

Бык	ETA	n	w _{jj}	Var($\hat{A} - A$)	SEP = $\sqrt{\text{Var}(\hat{A} - A)}$
1	+0,27	10	0,0456	27,3749	5,23
2	-7,34	50	0,0369	22,1595	4,71
3	+7,07	100	0,0342	20,4923	4,53

С увеличением числа дочерей на быка SEP уменьшается, а $r_{\hat{A}\hat{A}}$ - повышается.

Число эффективных дочерей для j-го быка

$$= n_{.j} - \left(\frac{h n_{ij}^2}{\sum_{i=1}^h n_i} \right).$$

Для быков 1, 2 и 3 получим:

$$\text{Бык №1} = 10 - \frac{4^2}{104} - \frac{6^2}{56} = 9,2033 = n_{e_1}.$$

$$\text{Бык №2} = 50 - \frac{50^2}{104} = 25,9615 = n_{e_2}$$

$$\text{Бык №3} = 100 - \frac{50^2}{104} - \frac{50^2}{56} = 31,3187 = n_{e_3}.$$

Число эффективных дочерей на быка зависит от распределения их по стадам.

Точность и ошибку оценки быка можно определить приближенным методом, используя число эффективных дочерей:

$$r_{\hat{A}\hat{A}}^2 = 1 - k d_j$$

и

$$\text{SEP} = \sqrt{d_j \sigma_e^2},$$

где

$$d_j = \frac{1}{(n_{e_j} + k)}$$

Бык	n _e	Точная r _{AA} ²	Аппрок. r _{AA} ²	Точная SEP	Аппрок. SEP
1	9,2033	0,562	0,617	5,23	4,98
2	25,9615	0,668	0,796	4,71	3,83
3	31,3187	0,698	0,822	4,53	3,60

Приближение не очень хорошее, но иногда оно полезно для быстрого расчета точности оценки, которую можно ожидать.

Для получения наиболее точной оценки ЕТА быка необходимо, чтобы его дочери распределялись по многим стадам, и в каждом стаде должно быть достаточное число дочерей других быков (сверстниц).

Допустим, что все коровы в стаде были получены от одного быка, например 200 коров, т.е. $n_i = 200$ и $n_{i1} = 200$. Если этот бык имел потомков только в одном стаде, то число эффективных дочерей у этого быка будет $n_e = 200 - 200^2 / 200 = 0$. Следовательно, нельзя проверять и оценивать быков, когда в хозяйстве используется только один бык (нет потомства других быков).

21.7. Учет родственных связей

Допустим, что имеются следующие данные по трем коровам:

Корова	Мать	Отец	Жир, кг
1	-	А	150
2	-	В	120
3	1	С	130

Данные описываются следующей моделью:

$$y_{ij} = \mu + L_i + e_{ij}$$

где $L_i = G_A$ коровы и $\sigma_L^2 = \sigma_A^2$.

Также допустим, что $h^2 = 1/7$.

Тогда $k = (1 - h^2) / h^2 = 6$.

Уравнения смешанной модели для этого примера могут быть легко построены. Так, уравнения наименьших квадратов запишутся:

$$\begin{bmatrix} 3 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{L}_1 \\ \hat{L}_2 \\ \hat{L}_3 \end{bmatrix} = \begin{bmatrix} 400 \\ 150 \\ 120 \\ 130 \end{bmatrix}.$$

Следующее действие должно быть суммирование k к диагональным элементам уравнений для L_i .

Однако при этом будет проигнорирована информация о родстве между коровами: корова № 1 является матерью коровы № 3.

Ниже дана матрица родства для рассматриваемого примера:

	A	B	C	A - ? 1	B - ? 2	C - 1 3
A	1	0	0	0,5	0	0,25
B	0	1	0	0	0,5	0
C	0	0	1	0	0	0,5
1	0,5	0	0	1	0	0,5
2	0	0,5	0	0	1	0
3	0,25	0	0,5	0,5	0	1

Числа в этой таблице характеризуют аддитивное генетическое родство между животными анализируемой выборки. Обычно эти данные обозначают как «матрица A ». Чтобы учесть в анализе родство между животными, к диагональным элементам уравнений для L_i вместо k суммируется kA^{-1} .

В течение многих лет единственным способом получения A^{-1} было формирование матрицы A с последующим ее обращением. Однако компьютерные возможности позволяли обращать матрицы размерностью 200×200 или меньше. В 1975 году Ч.Хендерсон предложил простую процедуру создания матрицы A^{-1} без создания матрицы A . Эта процедура не лимитирует размеры матрицы A^{-1} . Изобретение этого метода было большим достижением в разведении животных.

Правила создания матрицы A^{-1} :

1. Составить списки всех животных и их родителей

Корова	Отец	Мать
1	A	-
2	B	-
3	C	1

2. Добавить к нему список родственников

A	-	-
B	-	-
C	-	-

3. Возможно три ситуации:

- а) известны оба родителя;
- б) известен только один родитель;
- с) родители неизвестны.

В зависимости от ситуации при формировании матрицы A^{-1} используются различные константы:

Элементы A^{-1}	Родители известны		
	оба (а)	один (б)	нет (с)
i=корова	-	-	-
j=отец	-	-	-
k=мать	-	-	-
i, i	2	$1\frac{1}{3}$	1
j, j	$\frac{1}{2}$	} $\frac{1}{3}^*$	0
k, k	$\frac{1}{2}$		0
i, j	-1	} $\frac{2}{3}^*$	0
i, k	-1		0
j, k	$\frac{1}{2}$	0	0

* - Зависит от того, какой родитель неизвестен

Для коровы № 1 отцом является бык А, а мать неизвестна, следовательно используются константы ситуации б. Они записываются в матрицу A^{-1} :

	A - ? 1	B - ? 2	C - 1 3	A	B	C
1	4/3			-2/3		
2						
3						
A	-2/3			1/3		
B						
C						

Следует отметить, что матрица A^{-1} симметричная. Поэтому все элементы строки переписываются в соответствующий столбец.

У коровы № 2 отцом является бык В, а мать неизвестна. Добавим соответствующие коэффициенты в матрицу A^{-1} :

	A - ? 1	B - ? 2	C - 1 3	A	B	C
1	4/3			-2/3		
2		4/3			-2/3	
3						
A	-2/3			1/3		
B		-2/3			1/3	
C						

Корова № 3 имеет обоих родителей, поэтому к элементам матрицы A^{-1} дополняются константы ситуации (а):

	A - ? 1	B - ? 2	C - 1 3	A	B	C
1	4/3+1/2		-1	-2/3		1/2
2		4/3			-2/3	
3	-1		2			-1
A	-2/3			1/3		
B		-2/3			1/3	
C	1/2		-1			1/2

Животные А, В, С имеют неизвестных родителей, поэтому используются константы ситуации (с), т.е. к соответствующим диагональным элементам добавляется 1:

	A - ? 1	B - ? 2	C - 1 3	A	B	C
1	4/3+1/2		-1	-2/3		1/2
2		4/3			-2/3	
3	-1		2			-1
A	-2/3			1/3+1		
B		-2/3			1/3+1	
C	1/2		-1			1/2+1

После небольшого преобразования, окончательная матрица A^{-1} будет иметь вид:

$$A^{-1} = \frac{1}{6} \begin{bmatrix} 11 & 0 & -6 & -4 & 0 & 3 \\ 0 & 8 & 0 & 0 & -4 & 0 \\ -6 & 0 & 12 & 0 & 0 & -6 \\ -4 & 0 & 0 & 8 & 0 & 0 \\ 0 & -4 & 0 & 0 & 8 & 0 \\ 3 & 0 & -6 & 0 & 0 & 9 \end{bmatrix}.$$

Если три коровы были бы неродственными, то тогда к диагональным элементам уравнений для L_i суммируется k , получая, тем самым, ММЕ-уравнения:

$$\begin{bmatrix} 3 & 1 & 1 & 1 \\ 1 & 1+6 & 0 & 0 \\ 1 & 0 & 1+6 & 0 \\ 1 & 0 & 0 & 1+6 \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{L}_1 \\ \hat{L}_2 \\ \hat{L}_3 \end{bmatrix} = \begin{bmatrix} 400 \\ 150 \\ 120 \\ 130 \end{bmatrix}.$$

Их решение (без ограничений) дает следующие оценки:

$$\begin{bmatrix} \hat{\mu} \\ \hat{L}_1 \\ \hat{L}_2 \\ \hat{L}_3 \end{bmatrix} = \begin{bmatrix} 0,389 & -0,056 & -0,056 & -0,056 \\ -0,056 & 0,151 & 0,008 & 0,008 \\ -0,056 & 0,008 & 0,151 & 0,008 \\ -0,056 & 0,008 & 0,008 & 0,151 \end{bmatrix} \begin{bmatrix} 400 \\ 150 \\ 120 \\ 130 \end{bmatrix}$$

и

$$\begin{aligned} \hat{\mu} &= 133,3333; \\ \hat{L}_1 &= +2,3810 \\ \hat{L}_2 &= -1,9048; \\ \hat{L}_3 &= -0,4762. \end{aligned}$$

Если коровы родственны между собой, то вместо k к диагоналям суммируется kA^{-1} . Для этого LS-уравнения необходимо расширить до следующих размеров:

$$\begin{bmatrix} 3 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{L}_1 \\ \hat{L}_2 \\ \hat{L}_3 \\ \hat{L}_A \\ \hat{L}_B \\ \hat{L}_C \end{bmatrix} = \begin{bmatrix} 400 \\ 150 \\ 120 \\ 130 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Суммирование kA^{-1} к уравнениям для L дает следующие ММЕ-уравнения:

$$\begin{bmatrix} 3 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 12 & 0 & -6 & -4 & 0 & 3 \\ 1 & 0 & 9 & 0 & 0 & -4 & 0 \\ 1 & -6 & 0 & 13 & 0 & 0 & -6 \\ 0 & -4 & 0 & 0 & 8 & 0 & 0 \\ 0 & 0 & -4 & 0 & 0 & 8 & 0 \\ 0 & 3 & 0 & -6 & 0 & 0 & 9 \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{L}_1 \\ \hat{L}_2 \\ \hat{L}_3 \\ \hat{L}_A \\ \hat{L}_B \\ \hat{L}_C \end{bmatrix} = \begin{bmatrix} 400 \\ 150 \\ 120 \\ 130 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

После инверсии матрицы коэффициентов и умножение на вектор правой части ММЕ

$$\begin{bmatrix} \hat{\mu} \\ \hat{L}_1 \\ \hat{L}_2 \\ \hat{L}_3 \\ \hat{L}_A \\ \hat{L}_B \\ \hat{L}_C \end{bmatrix} = \begin{bmatrix} 0,407 & -0,081 & -0,058 & -0,081 & -0,041 & -0,029 & -0,027 \\ -0,081 & 0,155 & 0,012 & 0,078 & 0,077 & 0,006 & 0,000 \\ -0,058 & 0,012 & 0,151 & 0,012 & 0,006 & 0,076 & 0,004 \\ -0,081 & 0,078 & 0,012 & 0,155 & 0,039 & 0,006 & 0,077 \\ -0,041 & 0,077 & 0,006 & 0,039 & 0,164 & 0,003 & 0,000 \\ -0,029 & 0,006 & 0,076 & 0,006 & 0,003 & 0,163 & 0,002 \\ -0,027 & 0,000 & 0,004 & 0,077 & 0,000 & 0,002 & 0,162 \end{bmatrix} \begin{bmatrix} 400 \\ 150 \\ 120 \\ 130 \\ 0 \\ 0 \\ 0 \end{bmatrix},$$

получим следующие оценки:

$$\begin{aligned} \hat{\mu} &= 133,0233; & \hat{L}_1 &= +2,1646; & \hat{L}_A &= +1,0823; \\ & & \hat{L}_2 &= -1,8605; & \hat{L}_B &= -0,9302; \\ & & \hat{L}_3 &= +0,6261; & \hat{L}_C &= -0,3041. \end{aligned}$$

Точность оценки коров 1, 2, 3 рассчитывается из

$$r_{TI} = \sqrt{1 - kd_i},$$

где d_i - диагональный элемент матрицы коэффициентов ММЕ.

$$\begin{aligned} r_{TI} \quad \text{для} \quad \hat{L}_1 & \text{ это } 0,2646; \\ & \hat{L}_2 \quad \text{это } 0,3066; \\ & \hat{L}_3 \quad \text{это } 0,2646. \end{aligned}$$

21.8. Учет оценки отца

Допустим, что *отцы* коров были ранее проверены по качеству потомства и их ЕТА по количеству молочного жира известна, например:

Бык	n	ЕТА
A	32	+5
B	109	-2
C	21	+1

Ч. Хендерсон предложил способ включения этой информации в оценку коров. Он заключается в следующем:

- к диагональным элементам матрицы коэффициентов уравнений для быков ($\hat{L}_A, \hat{L}_B, \hat{L}_C$) суммируют

$$\frac{n(1-r)}{4-h^2},$$

где n - число дочерей при оценке ЕТА отца;

r - коэффициент повторяемости;

h^2 - коэффициент наследуемости.

- в правую часть уравнений для быков записывают их дерегрессированные оценки, умноженные на $2(1-r)/(4-h^2)$

$$\frac{2(1-r)}{(4-h^2)} \times \left[\text{ЕТА} / \left(\frac{h^2}{4+(n-1)h^2} \right) \right].$$

Если $r=0,3$ и $h^2=1/7$, то эти элементы для числового примера будут следующими:

Бык	Диагонали	Правая часть
A	5,8	+107,1
B	19,8	-98,7
C	3,8	+17,4

Ниже дана окончательная система ММЕ:

$$\begin{bmatrix} 3 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 12 & 0 & -6 & -4 & 0 & 3 \\ 1 & 0 & 9 & 0 & 0 & -4 & 0 \\ 1 & -6 & 0 & 13 & 0 & 0 & -6 \\ 0 & -4 & 0 & 0 & 13,8 & 0 & 0 \\ 0 & 0 & -4 & 0 & 0 & 27,8 & 0 \\ 0 & 3 & 0 & -6 & 0 & 0 & 12,8 \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{L}_1 \\ \hat{L}_2 \\ \hat{L}_3 \\ \hat{L}_A \\ \hat{L}_B \\ \hat{L}_C \end{bmatrix} = \begin{bmatrix} 400 \\ 150 \\ 120 \\ 130 \\ 107,1 \\ -98,7 \\ 17,4 \end{bmatrix}.$$

После инверсии матрицы W и умножения на RHS

$$\begin{bmatrix} \hat{\mu} \\ \hat{L}_1 \\ \hat{L}_2 \\ \hat{L}_3 \\ \hat{L}_A \\ \hat{L}_B \\ \hat{L}_C \end{bmatrix} = \begin{bmatrix} 0,396 & -0,071 & -0,047 & -0,071 & -0,021 & -0,007 & -0,017 \\ -0,071 & 0,137 & 0,008 & 0,069 & 0,040 & 0,001 & 0,000 \\ -0,047 & 0,008 & 0,124 & 0,008 & 0,002 & 0,018 & 0,002 \\ -0,071 & 0,069 & 0,008 & 0,136 & 0,020 & 0,001 & 0,048 \\ -0,021 & 0,040 & 0,002 & 0,020 & 0,084 & 0,000 & 0,000 \\ -0,007 & 0,001 & 0,018 & 0,001 & 0,000 & 0,039 & 0,000 \\ -0,017 & 0,000 & 0,002 & 0,048 & 0,000 & 0,000 & 0,100 \end{bmatrix} \begin{bmatrix} 400 \\ 150 \\ 120 \\ 130 \\ 107,1 \\ -98,7 \\ 17,4 \end{bmatrix}$$

получим оценки искомых параметров:

$$\begin{aligned} \hat{\mu} &= 131,1746; & \hat{L}_1 &= +6,0673; & \hat{L}_A &= +9,5195; \\ & & \hat{L}_2 &= -3,0122; & \hat{L}_B &= -3,9838; \\ & & \hat{L}_3 &= +3,4212; & \hat{L}_C &= +1,5410. \end{aligned}$$

Все оценки, кроме $\hat{\mu}$, являются оценками племенной ценности (EBV) и более предпочтительны, чем ETA.

При включении информации об отце, точность оценки коров составила:

$$r_{TI} \quad \text{для} \quad \begin{aligned} \hat{L}_1 &- 0,4233; \\ \hat{L}_2 &- 0,5043; \\ \hat{L}_3 &- 0,42825. \end{aligned}$$