

Capítulo XIV

El modelo animal simple: una metodología para los genetistas

José Atilio Aranguren Méndez
Rafael María Román Bravo

El Dr. C.R. Henderson dedicó su carrera a investigar el problema de la estimación de los efectos fijos, al cual llamó el “Mejor Estimador Lineal Insesgado” (BLUE) y a la predicción de los efectos aleatorios, a la cual llamó “Mejor Predicción Lineal Insesgada” (BLUP), bajo un modelo mixto. El origen de esta metodología se basó en una pregunta que les hiciera en el año 1949, el profesor Mood en una clase, y la cual se puede plantear como “Dado un puntaje de 130 en una prueba de coeficiente intelectual (IQ), ¿Cuál es el estimador de máxima verosimilitud del verdadero valor del IQ de ese individuo? Sabemos que la media del verdadero IQ y del IQ de la prueba, se asume normalmente distribuido con media 100, la varianza del error en la prueba es 25 y la varianza de del verdadero IQ es 225”.

Henderson (1973), razonó este problema sobre la base de lo que había aprendido sobre la estimación de la Más Probable Habilidad Productiva (MPHP) de una vaca y plantea de hecho la respuesta correcta 127; el profesor no aceptó la metodología y sugirió a la clase que deberían maximizar la distribución conjunta del verdadero IQ y del IQ de la prueba, para las variaciones en el verdadero IQ (Henderson, 1973). De hecho, el usó la sugerencia del profesor para derivar una metodología que le permitió estimar los efectos fijos y predecir los aleatorios, que resultan de una ligera modificación a las ecuaciones de cuadrados mínimos, en la parte correspondiente al efecto aleatorio. Se había demostrado que la metodología es equivalente a la recuperación de la información interbloque, en el análisis de bloques incompletos, en donde tratamiento es fijo y los bloques son aleatorios (Henderson *et al.*, 1959). Las aplicaciones de esta metodología han sido detalladas en forma amplia para la evaluación de toros (Henderson, 1973, 1974).

Sin embargo, por limitaciones de tipo técnico y teórico, no podía ser usada esta metodología en gran escala en la evaluación de animales. No fue sino hasta 1976, cuando se descubrió una vía rápida para invertir la matriz de relación, aun sin obtener la matriz de relación, lo que fue logrado a partir de un listado de individuos ordenados de forma que la identificación de la progenie sea inferior a la de sus padres (Henderson, 1976). Este descubrimiento es probablemente uno de los más grandes legados de Henderson (Van Vleck, 1998).

MODELO ANIMAL IGNORANDO LAS RELACIONES GENÉTICAS

La base conceptual del Modelo Animal en estas condiciones fue explicada en detalle por Van Vleck (1993). Supongamos que consideramos un modelo en el que no existe otro efecto fijo que el de la media μ , y que los animales se usan como único criterio de clasificación. Las ecuaciones para estimar μ y predecir el efecto del animal se construyen en la forma usual como construimos las ecuaciones ordinarias de cuadrados mínimos (OLS), considerando el efecto animal como un efecto fijo; luego, agregamos al bloque de animales la constante λ , la cual es el cociente de la varianza residual sobre la varianza genética aditiva. Así tendríamos el miembro izquierdo (LHM) de las ecuaciones del modelo mixto (MME) (Henderson, 1989). En este caso, debemos construir un vector, formado por el número total de observaciones, seguido de los valores fenotípicos para los animales como están en el LHM, lo que comúnmente se conoce como el miembro derecho (RHM) de las MME. En estas condiciones el modelo matemático que aplica es:

$$y_i = \mu + a_i + e_i$$

donde:

μ constante común a todas las observaciones; a_i = Es el efecto del genotipo aditivo del $i^{\text{ésimo}}$ animal, $i=1, \dots, n$ y e_i = Es el efecto de los factores ambientales no controlados, que afectan los registros del animal. Se asume que los a_i y los e_i son independientes, de manera que la covarianza entre ellos es $Cov(a_i, e_i) = 0$ Para ignorar las relaciones debemos asumir que la covarianza genética entre cualquier animal es $Cov(a_i, a_j) = 0$; finalmente bajo normalidad asumimos que los errores son independientes de manera que la covarianza entre ellos es $Cov(e_i, e_j) = 0$. De esta manera tenemos:

$$E(a_i) = 0; E(a_i^2) = \sigma_a^2; E(e_i) = 0 \text{ y } E(e_i^2) = \sigma_e^2$$

$$\text{Sabemos que el índice de herencia del carácter es: } h^2 = \frac{\sigma_a^2}{\sigma_y^2}$$

De esta manera tenemos: $\sigma_a^2 = h^2 \sigma_y^2$ y $\sigma_e^2 = (1-h^2) \sigma_y^2$

$$\text{Por lo tanto: } \lambda = \frac{\sigma_e^2}{\sigma_a^2} \rightarrow \frac{(1-h^2)\sigma_y^2}{h^2\sigma_y^2} \rightarrow \frac{(1-h^2)}{h^2}$$

Como señalamos inicialmente, las MME se construyen como para el caso de la estimación por OLS, pero se diferencian de ellas en que agregamos λ a la diagonal.

ECUACIONES OLS

$$\begin{array}{lcl}
 \text{Par./Est.} & : & \hat{\mu} \quad \hat{\alpha}_1 \quad \hat{\alpha}_2 \quad \hat{\alpha}_3 \quad \hat{\alpha}_4 \\
 \mu & : & n\hat{\mu} + \hat{\alpha}_1 + \hat{\alpha}_1 + \hat{\alpha}_1 + \hat{\alpha}_1 = y \\
 a_1 & : & \hat{\mu} + \hat{\alpha}_1 = y_1 \\
 a_2 & : & \hat{\mu} + \hat{\alpha}_1 = y_2 \\
 a_3 & : & \hat{\mu} + \hat{\alpha}_1 = y_3 \\
 a_4 & : & \hat{\mu} + \hat{\alpha}_1 = y_4
 \end{array} \tag{1}$$

Esas ecuaciones tienen propiedades interesantes: son simétricas, esto es, el elemento en la posición (i, j) es idéntico al de la posición (j, i) en este caso todos 0's, señalando que ningún animal i contribuye en nada a la estimación del animal j ; si sumamos las ecuaciones para los 4 animales, obtenemos la ecuación para la estimación de la media, lo mismo pasa por columnas. De forma similar si sumamos los elementos del RHS para los cuatro animales, obtenemos el total de observaciones de la ecuación para μ . Esto implica que ese sistema de ecuaciones es linealmente dependiente y por lo tanto para obtener los estimadores de los \hat{a}_i debemos imponer una restricción sobre los parámetros a estimar, normalmente sobre a_i aunque también lo podemos hacer sobre μ .

Ese sistema de ecuaciones, en general, lo podemos representar por conveniencia en forma matricial.

$$\begin{array}{cccccc|c|c}
 n_{..} & n_{.1} & n_{.2} & n_{.3} & n_{.4} & & \hat{\mu} & y_{..} \\
 n_{1.} & n_{1.} & 0 & 0 & 0 & & \hat{a}_1 & y_{1.} \\
 n_{2.} & 0 & n_{2.} & 0 & 0 & * & \hat{a}_2 & y_{2.} \\
 n_{3.} & 0 & 0 & n_{3.} & 0 & & \hat{a}_3 & y_{3.} \\
 n_{4.} & 0 & 0 & 0 & n_{4.} & & \hat{a}_4 & y_{4.}
 \end{array} = \tag{1}$$

Observe sin embargo que en este ejemplo sólo tenemos un registro por animal de manera que los coeficientes $n_{i.}$ son todos 1's y $n_{..}$ es en este caso 4, el número de observaciones para estimar μ . De esa manera, las ecuaciones en forma matricial serían:

$$\begin{array}{cccccc|c|c}
 4 & 1 & 1 & 1 & 1 & & \hat{\mu} & y \\
 1 & 1 & 0 & 0 & 0 & & \hat{a}_1 & y_1 \\
 1 & 0 & 1 & 0 & 0 & * & \hat{a}_2 & y_2 \\
 1 & 0 & 0 & 1 & 0 & & \hat{a}_3 & y_3 \\
 1 & 0 & 0 & 0 & 1 & & \hat{a}_4 & y_4
 \end{array} = \tag{2}$$

Resulta conveniente particionar las ecuaciones anteriores de la forma siguiente:

$$\begin{array}{c|c|c} \begin{array}{c} 4 \\ \dots \\ 1 \\ 1 \\ 1 \\ 1 \end{array} & \begin{array}{c} : \\ : \\ : \\ : \\ : \end{array} & \begin{array}{c} 1 \\ \dots \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \end{array} \\ \hline & & \begin{array}{c} \hat{\mu} \\ \dots \\ \hat{a}_2 \\ \hat{a}_3 \\ \hat{a}_4 \\ \hat{a}_5 \end{array} \\ \hline & & \begin{array}{c} = \\ = \\ = \\ = \\ = \end{array} & \begin{array}{c} y \\ \dots \\ y_1 \\ y_2 \\ y_3 \\ y_4 \end{array} \end{array}$$

Sea: $x' = [1 \ 1 \ 1 \ 1]$, el vector de incidencia de la media; entonces $x'x = 4$, la primera matriz en las OLS, sea Z la matriz de incidencia de los animales, entonces $x'Z$ es el vector hilera que aparece como la submatriz a la derecha de $x'x$, de la misma manera $Z'x$ es el vector columna que aparece debajo de $x'x$ finalmente tenemos la submatriz $Z'Z$ en este caso una matriz identidad, de orden el número de animales I_4 . En el RHS tendríamos como primer elemento el producto $x'y$ y debajo de este la suma de las observaciones para los 4 animales $Z'y$. Esto es:

$$\begin{bmatrix} x'x & x'Z \\ Z'x & Z'Z \end{bmatrix} * \begin{bmatrix} \hat{\mu} \\ \hat{a} \end{bmatrix} = \begin{bmatrix} x'y \\ Z'y \end{bmatrix}$$

Retornando a las ecuaciones OLS en [2], para obtener las MME lo que requerimos es agregar a las OLS en la diagonal de la sub-matriz $Z'Z$ la constante λ . Esto es:

$$\begin{bmatrix} 4 & 1 & 1 & 1 & 1 \\ 1 & 1+\lambda & 0 & 0 & 0 \\ 1 & 0 & 1+\lambda & 0 & 0 \\ 1 & 0 & 0 & 1+\lambda & 0 \\ 1 & 0 & 0 & 0 & 1+\lambda \end{bmatrix} * \begin{bmatrix} \hat{\mu} \\ \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \hat{a}_4 \end{bmatrix} = \begin{bmatrix} y \\ y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} \quad [3]$$

Ahora tenemos un hecho importante; la matriz sigue siendo simétrica, pero ahora, para el caso a una sola vía de clasificación; el hecho de haber sumado λ a la diagonal, hace que la suma de las hileras de animales no sumen a la de la hilera para la media; en otras palabras, el sistema de ecuaciones es linealmente independiente y no requerimos imponer restricciones en los efectos para obtener una solución. La solución al sistema requeriría obtener la inversa del LHM en [2] y multiplicar este por el RHM:

$$\begin{bmatrix} \hat{\mu} \\ \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \hat{\alpha}_3 \\ \hat{\alpha}_4 \end{bmatrix} = \begin{bmatrix} 4 & 1 & 1 & 1 & 1 \\ 1 & 1+\lambda & 0 & 0 & 0 \\ 1 & 0 & 1+\lambda & 0 & 0 \\ 1 & 0 & 0 & 1+\lambda & 0 \\ 1 & 0 & 0 & 0 & 1+\lambda \end{bmatrix}^{-1} * \begin{bmatrix} y \\ y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}$$

INTERPRETACIÓN DE LOS BLUP'S

Para ilustrar tomemos de [2], el producto de la hilera de cualquier animal digamos α_4 , y el vector de incógnitas; tendríamos para ese animal:

$$\hat{\mu} + \hat{\alpha}_4 (1 + \lambda) = y_4$$

Sustituyamos el valor de λ por lo que representa en función del cociente σ_e^2 / σ_a^2 . Pero el segundo término a la izquierda de esa expresión es:

$$1 + \frac{(1-h^2)}{h^2} = \frac{h^2 - (1-h^2)}{h^2} \rightarrow \frac{1}{h^2}$$

$$\hat{\mu} + \frac{1}{h^2} \hat{\alpha}_4 = y_4.$$

Por lo tanto, despejando para $\hat{\alpha}_4$:

$$\frac{1}{h^2} \hat{\alpha}_4 = (y_4 - \hat{\mu})$$

Finalmente tenemos:

$$\hat{\alpha}_4 = h^2 (y_4 - \hat{\mu}) \quad [4]$$

Ahora es aparente la razón de agregar el cociente σ_e^2 / σ_a^2 a la diagonal del segmento de animales, lo que tenemos en [4], es lo que Henderson (1974), llamó estimación por Máxima Verosimilitud y que posteriormente acuña como BLUP del efecto genético del animal y que Van Vleck (1993) señala que en algunos casos, es equivalente a un índice de selección, sólo que acá se usa el estimador de la media $\hat{\mu}$ en lugar del parámetro real μ . Así mismo, en el término h^2 , en el caso de índice de selección se asume que los componentes de varianza son conocidos, en el caso de BLUP, se usan los valores de los parámetros estimados con los datos.

Según Littel *et al.* (2006), BLUP, es una suerte de contracción hacia la media general, basada en los componentes de varianza. La idea básica de esta estimación por contracción es que la media del efecto aleatorio regresa hacia la media general y que el grado de contracción dependerá de la magnitud de la varianza. Una varianza grande en el efecto aleatorio, resultará en muy poca contracción hacia μ . Así mismo, señala que los estimadores muy arriba o muy debajo de μ son regresados hacia μ consistente con la magnitud relativa de σ_a^2 en relación a σ_y^2 ; de esta manera los valores extremos, son atenuados por la variabilidad subyacente, con lo cual se reduce el riesgo de interpretar los datos de manera incorrecta (Robinson, 1991).

DEMOSTRACIÓN DE QUE $\hat{\mu} = \bar{y}$, EL BLUE de μ

Hagamos el producto matricial de la primera hilera del LHM en [3], multiplicado por el vector de incógnitas; esto es, obtengamos la ecuación para estimar μ . Seguidamente sustituycamos los valores de los \hat{a}_i , encontrados para todos los animales como en [4]. Con un poco de manipulación algebraica:

$$4\hat{\mu} + \hat{a}_1 + \hat{a}_1 + \hat{a}_1 + \hat{a}_1 = y.$$

$$4\hat{\mu} + h^2(y_1, -\hat{\mu}) + h^2(y_2, -\hat{\mu}) + h^2(y_3, -\hat{\mu}) + h^2(y_4, -\hat{\mu}) = y$$

$$4\hat{\mu} + h^2(y_1, -\hat{\mu}) + (y_2, -\hat{\mu}) + (y_3, -\hat{\mu}) + (y_4, -\hat{\mu}) = y$$

$$4\hat{\mu} + h^2[y_1 + y_2 + y_3 + y_4, -4\hat{\mu}] = y$$

$$4\hat{\mu} + h^2[y, -4\hat{\mu}] = y$$

$$4\hat{\mu} + h^2 y, -4h^2\hat{\mu} = y,$$

$$4\hat{\mu} - 4h^2\hat{\mu} = y, -h^2 y,$$

$$4\hat{\mu}(1 - h^2) = y(1 - h^2)$$

$$\hat{\mu} = \frac{y}{4} \rightarrow \bar{y}$$

El cual es el Mejor Estimador Lineal no sesgado de la media poblacional BLUE de μ .

MODELO ANIMAL SIMPLE CONSIDERANDO TODAS LAS RELACIONES

Sea el modelo en forma matricial: $y = X\beta + Za + e$

dónde:

y = El vector de observaciones de dimensión $n \times 1$

β = El vector con los efectos fijos de dimensión $q \times 1$

e = El vector de los residuales con dimensión $n \times 1$

x = La matriz de incidencia de los efectos fijos, asociando registros con efectos fijos y de dimensión $n \times q$

z = La matriz de incidencia de los efectos aleatorios, asociando animales con registros y de dimensión $n \times r$

Se asume que:

$$E(a) = 0; E(e) = 0; E(y) = E[X\beta + Za + e] \text{ y } E(y) = X\beta$$

Además, los efectos residuales, que incluyen los efectos ambientales aleatorios y los errores de medición, junto con los efectos no aditivos, son independientes uno de otro, con varianza σ_e^2 . Por lo tanto:

$$V(e) = I\sigma_e^2, \text{ usualmente llamada } R \text{ y } V(a) = A\sigma_a^2, \text{ usualmente llamada } G$$

Los efectos aleatorios del modelo son independientes por lo tanto: $Cov(a, e) = 0$. En consecuencia, $V(y) = ZGZ' + R$.

Las MME para el modelo anterior pueden ser representadas en forma matricial como:

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + A*\lambda \end{bmatrix} * \begin{bmatrix} \hat{\beta} \\ \hat{a} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix} \tag{5}$$

La matriz $X'X$ contiene el número de observaciones asociado con la media y cada uno de los niveles de los efectos fijos, si es que la primera está incluida en el modelo; $X'Z$ asocia para cada efecto fijo los efectos aleatorios; la sub-matriz $Z'Z$, es una matriz identidad aumentada por tantas columnas e hileras de 0's como animales sin registros existan, normalmente los individuos de la población base, A^{-1} es la inversa de la matriz de relación A , siendo esta lo que es llamada el numerador de la matriz de la relación entre todos los animales de la población y λ como definimos en la parte anterior es el cociente σ_e^2 / σ_a^2 . Las 4 submatrices conforman lo que es la matriz de coeficientes de las MME.

Los elementos $\hat{\beta}$ y \hat{a} conforman el vector de los términos desconocidos a estimar y predecir respectivamente.

El RHM, está formado por el vector columna que tiene los sub-vectores $X'y$ cuya función es totalizar los totales de cada nivel de los efectos fijos; además, contiene el sub-vector con los totales de los registros por animal (para este caso simplemente el valor de los registros de cada animal). Ejemplo numérico.

Animal	Padre	Madre	Sexo	y
1
2
3	1	2	1	4,0
4	.	2	1	4,1
5	3	.	2	3,7
6	1	5	2	4,2

La matriz X tiene tantas columnas como niveles tenga cada efecto fijo, para evitar problemas de estimabilidad ignoraremos la columna para la media, de manera que no tengamos que imponer restricciones en este ejemplo particular, de esta manera en el vector de soluciones, obtendremos como solución la media de cada sexo. Para la construcción de X simplemente teniendo dos niveles, habrá que definir 2 columnas y asociar los registros con cada columna, en el lugar que un registro tenga valores asociados con ese nivel coloque un 1, el resto lo completa con 0's, proceda así hasta agotar todos los niveles.

$$X = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}$$

La matriz Z , se define en una forma similar, pero observaremos la columna de animales y el vector de observaciones, construiremos tantas columnas como animales con registros tengamos en la data y de acuerdo con las columnas de animal y observación colocaremos un 1 para cada animal asociado con el registro los demás elementos se colocan a 0's. De esta manera, tendremos una matriz identidad de orden igual al número de registros, sin embargo debemos agregar a la izquierda de esta matriz tantas columnas de 0's, como individuos base exista, en este caso dos.

$$Z = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

El vector de observaciones es simplemente: $y = \begin{bmatrix} 4,0 \\ 4,1 \\ 3,7 \\ 4,2 \end{bmatrix}$

$$X'X = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

$$X'Z = \begin{bmatrix} 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

$$Z'X = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}$$

$$Z'Z = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Ahora, los productos de las matrices de incidencia del efecto fijo y el efecto aleatorio multiplicado por y , para formar parte del RHM, serán:

$$X'y = \begin{bmatrix} 8,1 \\ 7,9 \end{bmatrix} \quad Z'y = \begin{bmatrix} 0 \\ 0 \\ 4,0 \\ 4,1 \\ 3,7 \\ 4,2 \end{bmatrix}$$

De esta manera ensamblando esas cuatro sub-matrices, correspondientes al LHM, tal como se señala en [5], el vector de incógnitas y los sub-vectores en el RHM, tendríamos lo que se conoce como las ecuaciones ordinarias de cuadrados mínimos.

$$\begin{bmatrix} 2 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} * \begin{bmatrix} \hat{s}_1 \\ \hat{s}_2 \\ \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \hat{a}_4 \\ \hat{a}_5 \\ \hat{a}_6 \end{bmatrix} = \begin{bmatrix} 8,1 \\ 7,9 \\ 0 \\ 0 \\ 4,0 \\ 4,1 \\ 3,7 \\ 4,2 \end{bmatrix}$$

Observado detenidamente esas ecuaciones vemos que tienen la debilidad de que no podremos estimar el mérito genético de los animales sin registros \hat{a}_1 y \hat{a}_2 . Tendríamos que eliminarlos para obtener la solución. Haciendo eso tendríamos, además que imponer una restricción sobre uno de los efectos en el sistema de ecuaciones, pues tal como están son linealmente dependientes.

Para obtener las MME, que son simplemente una modificación de las OLS, requerimos de la inversa de la matriz de relación A^{-1} , que pudiéramos obtener aplicando las reglas de Henderson (Van Vleck, 1993).

$$A^{-1} = \begin{bmatrix} 2 & 1/2 & -1 & 0 & 1/2 & -1 \\ 1/2 & 11/6 & -1 & -2/3 & 0 & 0 \\ -1 & -1 & 7/3 & 0 & -2/3 & 0 \\ 0 & -2/3 & 0 & 4/3 & 0 & 0 \\ 1/2 & 0 & -2/3 & 0 & 11/6 & -1 \\ -1 & 0 & 0 & 0 & -1 & 2 \end{bmatrix} \tag{6}$$

Asumiendo que $h^2 = 0,6$ para ese carácter, tenemos:

$$\lambda = \frac{(1-h^2)}{h^2} \rightarrow \frac{(1-0,6)}{0,6} \rightarrow \lambda = 2/3$$

$$\lambda^* A^{-1} = \begin{bmatrix} 4/3 & 1/3 & -2/3 & 0 & 1/3 & -2/3 \\ 1/3 & 11/9 & -2/3 & -4/9 & 0 & 0 \\ -2/3 & -2/3 & 14/9 & 0 & -4/9 & 0 \\ 0 & -4/9 & 0 & 8/9 & 0 & 0 \\ 1/3 & 0 & -4/9 & 0 & 11/9 & -2/3 \\ -2/3 & 0 & 0 & 0 & -2/3 & 4/3 \end{bmatrix} \quad [7]$$

Después de agregar $\lambda^* A^{-1}$ al segmento de animales en las OLS tenemos las MME.

$$\begin{bmatrix} 2 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1,3333 & 0,3333 & -0,6667 & 0 & 0,3333 & -0,6667 \\ 0 & 0 & 0,3333 & 1,2222 & -0,6667 & -0,4445 & 0 & 0 \\ 1 & 0 & -0,6667 & -0,6667 & 2,5555 & 0 & -0,4445 & 0 \\ 1 & 0 & 0 & -0,4445 & 0 & 1,8889 & 0 & 0 \\ 0 & 1 & 0,3333 & 0 & -0,4445 & 0 & 2,2222 & -0,6667 \\ 0 & 1 & -0,6667 & 0 & 0 & 0 & -0,6667 & 2,3333 \end{bmatrix} * \begin{bmatrix} \hat{s}_1 \\ \hat{s}_2 \\ \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \hat{a}_4 \\ \hat{a}_5 \\ \hat{a}_6 \end{bmatrix} = \begin{bmatrix} 8,1 \\ 7,9 \\ 0 \\ 0 \\ 4,0 \\ 4,1 \\ 3,7 \\ 4,2 \end{bmatrix} \quad [8]$$

Para este caso simple, la solución de este sistema se puede realizar con muy pocos individuos, por inversión directa o por medio de un procedimiento iterativo como por ejemplo, uno de los más simples, el procedimiento de Gauss-Seidel. En casos de aplicaciones reales que involucran miles de animales, se requiere obligatoriamente recurrir a métodos iterativos, en los cuales hay en la actualidad un desarrollo considerable.

$$\begin{bmatrix} \hat{s}_1 \\ \hat{s}_2 \\ \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \hat{a}_4 \\ \hat{a}_5 \\ \hat{a}_6 \end{bmatrix} = b \begin{bmatrix} 2 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1,3333 & 0,3333 & -0,6667 & 0 & 0,3333 & -0,6667 \\ 0 & 0 & 0,3333 & 1,2222 & -0,6667 & -0,4445 & 0 & 0 \\ 1 & 0 & -0,6667 & -0,6667 & 2,5555 & 0 & -0,4445 & 0 \\ 1 & 0 & 0 & -0,4445 & 0 & 1,8889 & 0 & 0 \\ 0 & 1 & 0,3333 & 0 & -0,4445 & 0 & 2,2222 & -0,6667 \\ 0 & 1 & -0,6667 & 0 & 0 & 0 & -0,6667 & 2,3333 \end{bmatrix}^{-1} \begin{bmatrix} 8,1 \\ 7,9 \\ 0 \\ 0 \\ 4,0 \\ 4,1 \\ 3,7 \\ 4,2 \end{bmatrix}$$

La inversa de la matriz de coeficientes de las MME, la cual requerimos para obtener su diagonal, con el propósito de calcular los errores típicos de la predicción, exactitudes y confiabilidades es ilustrada a continuación:

$$\begin{bmatrix} \hat{s}_1 \\ \hat{s}_2 \\ \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \hat{\alpha}_3 \\ \hat{\alpha}_4 \\ \hat{\alpha}_5 \\ \hat{\alpha}_6 \end{bmatrix} = \begin{bmatrix} 1,4369 & 0,4465 & -0,3838 & -0,7475 & -0,9373 & -0,9366 & -0,4607 & -0,4324 \\ 0,4465 & 1,6764 & -0,5355 & -0,2859 & -0,5894 & -0,3037 & -1,1521 & -1,2006 \\ -0,3828 & -0,5355 & 1,2776 & 0,0310 & 0,5556 & 0,2099 & 0,3706 & 0,7004 \\ -0,7475 & -0,2859 & 0,0310 & 1,4895 & 0,7487 & 0,7462 & 0,3425 & 0,2292 \\ -0,9373 & -0,5894 & 0,5556 & 0,7487 & 1,2021 & 0,6724 & 0,5969 & 0,5819 \\ -0,9366 & -0,3037 & 0,2099 & 0,7462 & 0,6724 & 1,2009 & 0,3245 & 0,2828 \\ -0,4607 & -1,1521 & 0,3706 & 0,3425 & 0,5969 & 0,3245 & 1,3258 & 0,9785 \\ -0,4324 & -1,2006 & 0,7004 & 0,2292 & 0,5819 & 0,2828 & 0,9785 & 1,4228 \end{bmatrix} \begin{bmatrix} 8,1 \\ 7,9 \\ 0 \\ 0 \\ 4,0 \\ 4,1 \\ 3,7 \\ 4,2 \end{bmatrix} *$$

La solución por inversión directa de la matriz de coeficientes LHM de las MME, y el producto de esta por el RHM, nos da la solución; los dos primeros elementos del vector nos da los BLUE's de los sexos, y los elementos 3 a 8, son los BLUP's de los seis animales.

$$\begin{bmatrix} \hat{s}_1 \\ \hat{s}_2 \\ \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \hat{a}_4 \\ \hat{a}_5 \\ \hat{a}_6 \end{bmatrix} = \begin{bmatrix} 4,0571 \\ 3,9592 \\ 0,0652 \\ -0,0284 \\ -0,0302 \\ 0,0160 \\ -0,1005 \\ 0,0961 \end{bmatrix}$$

La inspección de las MME nos indica que la solución para cada sexo está dada por las expresiones:

$$\hat{s}_1 = \frac{1}{2} [8,1 - \hat{a}_3 - \hat{a}_4] \rightarrow 4,0571$$

$$\hat{s}_2 = \frac{1}{2} [7,9 - \hat{a}_5 - \hat{a}_6] \rightarrow 3,9592$$

Las cuales reflejan directamente los promedios de cada sexo (por cuanto se eliminó la ecuación para la media) ajustadas por el efecto genético aditivo de los animales que se sumen para obtener el total de cada sexo.

DESCOMPOSICIÓN DE LOS BLUP'S DE LOS ANIMALES

El análisis de los BLUP's de los animales ha sido presentado en detalle (Van Raden & Wiggans, 1991); Mrode (1992); Van Vleck (1993) y la lectura de ellos nos lleva a la conclusión de que la solución de un animal, depende de la cantidad de información que aporten los animales relacionados con él. En general podemos identificar tres casos: a) Individuos de la población base, sin registros pero con progenie; b) Individuos

con registros, con uno o ambos progenitores conocidos, pero sin progenie y c) Individuos con registros, con uno o ambos progenitores conocidos y con progenie.

El aporte de cada fuente de información es más fácil de entender si despejamos, para cada animal en las MME dadas en [8]; en otras palabras, tomando el producto de la matriz de coeficientes por las incógnitas, igualándola con el RHM y posteriormente despejando en términos de los efectos, tal como si fuésemos a resolver el sistema de ecuaciones en forma iterativa. Los casos de animales sin registros son fáciles de seguir; pero en los otros casos, las manipulaciones algebraicas son más complejas y requerirán además del uso de factores de ponderación.

INDIVIDUOS DE LA POBLACIÓN BASE CON PROGENIE

Este es el caso por ejemplo del animal 1 y 2; para estos, su BLUP, es función de los registros de su progenie, ajustado por el BLUP del otro progenitor que interviene en el apareamiento para generar la progenie en cuestión. Es por lo tanto acá donde se capta que el modelo animal toma en cuenta cualquier efecto de la selección. La contribución de la progenie al valor del progenitor está dada por el término CP_i y este debe ser dividido por la diagonal en las MME para el animal en cuestión:

$$CP_i = \sum \frac{1}{2} \lambda g [2\hat{a}_j - \hat{a}_{progenitor}] \quad [9]$$

$$\hat{a}_i = \frac{CP_i}{diagonal\ MME} \quad [10]$$

λ es como hemos dicho, el cociente de la varianza ambiental sobre la varianza genética aditiva; g toma valores de 1 ó $\frac{2}{3}$, dependiendo de si el otro progenitor que interviene en el apareamiento para generar la progenie es conocido o no respectivamente. \hat{a}_j es el valor genético de la progenie y $\hat{a}_{progenitor}$, es el merito genético del progenitor apareado con el i ésimo individuo para generar la progenie.

BLUP (α_1, λ)

Tomemos la ecuación del animal 1 de las MME:

$1,3333\hat{a}_1 + 0,3333\hat{a}_2 - 0,6667\hat{a}_3 + 0,3333\hat{a}_5 - 0,6667\hat{a}_6 = 0$. Esta ecuación en términos de los coeficientes de inversa de la matriz de relación, la sub-matriz del segmento de animales y λ . es:

$$(0 + 2\lambda)\hat{\alpha}_1 + (1/2)\lambda\hat{\alpha}_2 - (1/2)\lambda\hat{\alpha}_3 + (1)\lambda\hat{\alpha}_5$$

Despejando para $\hat{\alpha}_1$.

$$(2\lambda)\hat{\alpha}_1 = -(1/2)\lambda\hat{\alpha}_2 + \lambda\hat{\alpha}_3 - (1/2)\lambda\hat{\alpha}_5 + (1)\lambda\hat{\alpha}_6$$

Agrupando términos en función de los apareamientos que generan la progenie:

$$\hat{\alpha}_1 = \frac{1}{2\lambda} [(\lambda\hat{\alpha}_3 - (1/2)\lambda\hat{\alpha}_2) + (\lambda\hat{\alpha}_6 - (1/2)\lambda\hat{\alpha}_5)]$$

$$\hat{\alpha}_1 = \frac{1}{2\lambda} [(\lambda(\hat{\alpha}_3 - (1/2)\hat{\alpha}_2) + \lambda(\hat{\alpha}_6 - (1/2)\hat{\alpha}_5))]$$

Luego de factorizar λ se convierte en:

$$\hat{a}_1 = \frac{1}{2} [\hat{a}_3 - (1/2)\hat{a}_2 + \hat{a}_6 - (1/2)\hat{a}_5]$$

En resumen, al BLUP del animal 1, contribuyen los valores genéticos de sus progenies 3 y 6, ajustado el primero por el valor genético del animal 2 (madre de 3) y el segundo ajustado por el valor genético del animal 5 (madre de 6). Sustituimos los valores de la solución para animales:

$$\hat{a}_1 = \frac{1}{2} [(-0,0302 + \frac{1}{2}(-0,0284)) + (0,0961 - \frac{1}{2}(-0,1005))] = 0,0652$$

En resumen, al BLUP del animal 1, contribuyen los valores genéticos de sus progenies 3 y 6, ajustado el primero por el valor genético del animal 2 (madre de 3) y el segundo ajustado por el valor genético del animal 5 (madre de 6). Resultados idénticos serían obtenidos aplicando [9]. Primero encontramos la contribución de la progenie y luego dividimos por la diagonal de las MME para ese animal.

BLUP (a_2).

La ecuación para el animal 2 es:

$0,3333\hat{a}_1 + 1,2222\hat{a}_2 - 0,6667\hat{a}_3 - 0,4444\hat{a}_4 = 0$; Expresando como antes en términos de la matriz de relación, la matriz correspondiente a los animales y λ , tenemos:

$$(1/2)\lambda\hat{a}_1 + (0 + \frac{11}{6}\lambda\hat{a}_2 - \lambda\hat{a}_3 - \frac{2}{3}\lambda\hat{a}_4 = 0$$

$$\frac{11}{2}\lambda\hat{a}_2 = (\lambda\hat{a}_3 - (1/2)\lambda\hat{a}_1) + \frac{2}{3}\lambda\hat{a}_4$$

$$\hat{a}_2 = \frac{6}{11\lambda} (\lambda\hat{a}_3 - (1/2)\lambda\hat{a}_1 + \frac{2}{3}\lambda\hat{a}_4$$

$$\hat{a}_2 = \frac{6}{11} (\hat{a}_3 - (1/2)\hat{a}_1) + \frac{2}{3}\hat{a}_4$$

$$\hat{a}_2 = \frac{1}{11/9} [(2/3)(-0,302 - (1/2)0,0652) + (2/3)(2/3)0,0160] = 0,0284$$

Resultados idénticos se obtendrían aplicando [9], con ligeros cambios.

ANIMALES SIN PROGENIE PERO CON REGISTROS PROPIOS

Un ejemplo de este tipo de animales son el 4 y el 6; a los BLUP's de este contribuyen dos términos; el primero es el promedio los méritos genéticos de los dos progenitores, el segundo, es la contribución debida al registro propio pero ajustado por los efectos fijos y el muestreo mendeliano.

Definamos la covarianza entre el registro propio y el muestreo mendeliano como:

$$k = \frac{dh^2}{(dh^2 + (1-h^2))}$$

Donde d son elementos que toman valores de $\frac{1}{2}$, $\frac{3}{4}$ ó 1 según si ambos, uno o ningún progenitor es conocido. Para el caso del animal con ambos progenitores conocidos:

BLUP $((a_6))$.

$$k = \frac{(0,50)(0,6)}{(0,50)(0,6) + (1-0,6)} \rightarrow 0,4286$$

$$\hat{a}_6 = (1/2)(\hat{a}_1 + \hat{a}_5) + k[(y_6 - \hat{s}_2 - (1/2)(\hat{a}_1 + \hat{a}_5))]$$

$$\hat{a}_6 = (1/2)(0,0652 + (-0,1005)) + 0,4286[(4,2 - 3,9522 - (1/2)(0,0652 + (-0,1005))] \rightarrow 0,0981$$

ANIMALES CON REGISTROS Y PROGENIE

Para esos casos, hay tres contribuciones a su BLUP: primero, la mitad del aporte de los progenitores (AP); segundo, registro corregido por los efectos fijos (RC) y tercero, contribución de la progenie (CP). Los términos AP, RC y CP; deben ser ponderados por los factores w_1, w_2 y w_3 .

El numerador de w_1 es $2\lambda g$ donde $g=1, \frac{1}{3}$ ó $\frac{1}{2}$ si ambos, uno o ningún progenitor es conocido respectivamente; el numerador de w_2 es el número de registros para la progenie n_i y el numerador para w_3 es igual a $(1/2)\lambda g$, con $g=1$ ó $\frac{1}{3}$ cuando el otro progenitor de la progenie es conocido o desconocido respectivamente. El denominador de los 3 es la suma de los 3 numeradores. La suma de w_1, w_2 y w_3 debe ser 1.

Para el caso del animal 3, con ambos progenitores son conocidos tenemos:

$$n_1 = 2\lambda g \rightarrow 2\left(\frac{2}{3}\right)1 = \frac{4}{3}$$

$$n_2 = n_i \rightarrow 1$$

$$n_3 = (1/2)\lambda g \quad \frac{1}{2}\left(\frac{2}{3}\right)\left(\frac{2}{3}\right) = \frac{2}{9}$$

El denominador es:

$$d = n_1 + n_2 + n_3 \rightarrow \frac{23}{9}$$

$$w_1 = \frac{4/3}{23/9} \rightarrow \frac{12}{23}$$

$$w_2 = \frac{1}{23/9} \rightarrow \frac{9}{23}$$

$$w_3 = \frac{2/9}{23/9} \rightarrow \frac{18}{207}$$

$$(w_1 + w_2 + w_3) = \left(\frac{12}{23} + \frac{9}{23} + \frac{18}{207} \right) \rightarrow 1$$

$$\hat{a}_3 = w_1(AP) + w_2(RC) + w_3(CP)$$

Nota: para el cálculo de la contribución de la progenie CP se debe aplicar el término entre corchetes dado en [9].

$$\hat{a}_3 = w_1(1/2)(\hat{a}_1 + \hat{a}_2) + w_2(y_3 - \hat{s}_1) + w_3(2\hat{a}_5)$$

$$\hat{a}_3 = \frac{12}{23} \left(\frac{1}{2} (0,0652 + (-0,0284)) \right) + \frac{9}{23} (4,0 - 4,0571) + \frac{18}{207} (2(-0,1005)) \rightarrow -0,03022$$

ESTIMACIÓN DE LA CONFIABILIDAD EXACTITUD Y ERROR TÍPICO DE PREDICCIÓN DE LOS BLUP'S

Las predicciones de los valores genéticos de los animales deben ir acompañados de una medida de variabilidad, que es llamada error típico de la predicción (SEP) y la cual es la raíz cuadrada de la varianza del error de predicción (PEV). Para el cálculo de la PEV, se requieren los elementos de la diagonal de la inversa de las MME, correspondiente al bloque de animales, $d_{i,i}$, la PEV, puede ser considerada como la fracción de la varianza genética aditiva no tomada en cuenta por la predicción.

$$PEV = C_{22} \sigma_e^2 \rightarrow (1 - r^2) \sigma_a^2$$

Para el $i^{\text{ésimo}}$ animal:

$$d_{i,i} \sigma_e^2 = (1 - r^2) \sigma_a^2$$

$$d_{i,i} \frac{\sigma_e^2}{\sigma_a^2} = (1 - r^2)$$

$$d_{i,i} \lambda = 1 - r^2$$

$$r^2 = 1 - d_{i,i} \lambda$$

BLUPs, elementos de la diagonal, confiabilidades y exactitudes de las evaluaciones

animal	BLUP	d ⁱⁱ	r ²	r
\hat{a}_1	0,0652	1,2776	0,1482	0,3850
$\hat{\alpha}_2$	-0,0284	1,4895	0,0699	0,0836
\hat{a}_3	-0,0302	1,2021	0,1986	0,4456
\hat{a}_4	0,0160	1,2009	0,1994	0,4466
\hat{a}_5	0,1005	1,3528	0,1161	0,3408
\hat{a}_6	0,0961	1,4228	0,0515	0,2269

r^2 es el cuadrado de la correlación entre el mérito genético real y el mérito genético predicho, en el ganado lechero se refiere generalmente como “confiabilidad”. La exactitud r , es la raíz cuadrada de la confiabilidad y normalmente es usado en ganado de carne.

$$r = \sqrt{1 - d^{i,i} \lambda}$$

El error típico de la predicción está dado por: $SEP = \sqrt{d^{i,i} \sigma_e^2}$

Los BLUP's de los animales predicen el mérito genético de cada uno de ellos para la sumatoria de todos los *loci* que intervienen en un carácter, por tal razón, en general, las evaluaciones de los animales se publican en términos de PTA o DEP, los cuales son ½ BLUP, por cuanto los progenitores transmiten a la progenie la mitad de sus genes, no sus genotipos. De cualquier forma esos valores de predicción se usan para comparar animales en la población donde fueron predichos

Por ejemplo:

$$\hat{a}_1 - \hat{a}_3 = 0,0652 - (-0,0302) \rightarrow 0,0954$$

Sería de esperar que el valor genético de la progenie del macho 1 sea 0,0954 unidades superior a la progenie del macho 3.

La correlación entre los BLUP's obtenidos en la predicción y sus verdaderos valores nos dan una idea del grado de confianza que debemos tener en esos valores, a medida que el número de animales incrementa considerablemente, la correlación tenderá al valor 1, para cada animal, sin llegar a 1. De la misma manera, animales con poca progenie, tienden a tener valores bajos en esta correlación.

REFERENCIAS BIBLIOGRÁFICAS

- Henderson CR. 1973. Sire Evaluation and Genetic Trends. In: Proc Anim Breed Genetics Symp in Honour J.L. Lush. Am Soc Anim Sci. Blackburgh, Champaign, Illinois. pp 10-41.
- Henderson CR. 1974. General flexibility of linear model techniques for sire evaluation. Symposium: New sire evaluations. J Dairy Sci. 57 (8): 963-972.
- Henderson CR. 1976. A simple method for computing the inverse of the relationship matrix used in prediction of breeding values. Biometrics 32: 69-83.
- Henderson CR. 1988. Theoretical basis and computational methods for a number of different animal models. J Dairy Sci. 71(Suppl. 2): 1-16.
- Henderson CR, Kempthorne O, Searle SR, von Krosigk CM. 1959. The estimation of environmental and genetic trends from records subject to culling. Biometrics. 13:192-218.
- Litell RC, Milliken GA, Stroup WW, Wolfinger ES, Shabenberger O. 2006. SAS for Mixed Models. 2nd Edition. SAS Institute Inc., Cary, NC, USA pp. 206-242.
- Mrode RA. 1996. Linear Models for the Prediction of Animal Breeding Values. CAB, International, Wallingford, U.K. 187 pp.
- Robinson GK. 1991. That BLUP is a Good Thing: The Estimation of Random Effects. Stat Sci. 6 (1): 15-33.
- Van Raden PM, Wiggans GR. 1991. Derivation, calculation, and use of national animal model information. J Dairy Sci. 74: 2737-2746.
- Van Vleck LD. 1993. Selection Index and Introduction to Mixed Model Methods. CRC Press, Inc. Boca Raton, Fl, U.S.A. 481 pp.
- Van Vleck LD. 1998. Charles Roy Henderson, 1911-1989: a brief biography. J Anim Sci. 76: 2959-2961.