ESTIMACION DE PARAMETROS GENETICOS A PARTIR DE OPTIMAS PREDICCIONES LINEALES INSESGADAS

E. Manfredi (1)

Recibido: 30/9/82
Aceptado: 14/2/83

INTRODUCCION

La estimación de parámetros genéticos, específicamente heredabilidad y correlaciones genética, es imprescindible para evaluar animales y predecir progresos genéticos posibles.

La heredabilidad (h^2) se define como la (σ^2_G)ión entre la varianza genética aditiva (σ^2_G) y la varianza fenotípica (σ^2_F):

$$h^2 = \frac{\sigma^2_G}{\sigma^2_F}$$

La correlación genética entre dos caracteres es (Searle y Rounsville, 1974):

$$r_{i,j} = \frac{\sigma_{i,j}}{(\sigma^2_i \sigma^2_j)^{0,5}} = \frac{\dfrac{\sigma^2_{i+j} - \sigma^2_i - \sigma^2_j}{2}}{(\sigma^2_i \ \sigma^2_j)^{0,5}}$$

en donde

$r_{i,j}$: correlación genética entre caracteres i y j

$\sigma_{i,j}$: covarianza genética entre caracteres i y j

$\sigma^2_i, \sigma^2_j, \sigma^2_{i+j}$: varianzas genéticas para los caracteres i, j e (i + j) respectivamente.

Estas definiciones muestran que la estimación de heredabilidad y correlaciones genéticas se basa en estimaciones de varianzas genéticas aditivas para caracteres individuales y sumas de pares de caracteres individuales.

Las varianzas genéticas aditivas pueden estimarse a su vez mediante un análisis de medio hermanos (hermanas) paternos (maternas) (Van Vleck, 1979):

$$\hat{\sigma}^2_G = 4\hat{\sigma}^2_p$$

en donde

$\hat{\sigma}^2_G$: varianza genética aditiva.

$\hat{\sigma}^2_p$: varianza asociada al efecto padre.

Es entonces obvia la conclusión de que una estimación del componente de varianza asociado a los padres en un análisis de registros de medio hermanas paternas, proveerá buenos estimadores de variantes genéticas que podrán ser utilizados con éxito en cálculos de heredabilidades y correlaciones genéticas.

Aquí reside la importancia que los analistas de datos ganaderos asignan a la estimación de componentes de varianza y que ha generado numerosos estudios sobre el tema.

Fisher (1925) propuso el método de análisis de varianza que permite partir a ésta en distintos componentes igualando sumas de cuadrados a sus esperanzas matemáticas. Cochran (1939) extendió esta metodología a la situación, muy frecuente en ganadería, de datos desbalanceados. Henderson (1953) presentó tres métodos de estimación de compo-

---

(1) Cátedra de Zootecnia General, Departamento de Zootecnia, Facultad de Agronomía, Universidad de Buenos Aires, Av. San Martín 4453, (1417) Buenos Aires, Argentina.

nentes de varianza: el Método I, aplicable a modelos totalmente aleatorios; el Método III, para modelos mixtos; y el Método II, también aplicable a ciertos modelos mixtos, pero más fácil de computar que el Método III.

Recientemente, Henderson (1980) presentó un nuevo método de estimación de componentes de varianza, que según el autor, es útil para cualquier modelo (ventaja sobre el Método II) y es de cómputo relativamente simple (ventaja sobre el Método III). Este último método se aplicó en este estudio para estimar heredabilidades y correlaciones genéticas en producción de leche, grasa y proteína sobre ganado Holando.

MATERIALES Y METODOS

1. Datos

Se procesaron 18.416 registros obtenidos en los archivos maestros del Laboratorio de Procesamiento de registros lecheros de la Universidad de Cornell, Estados Unidos de Norte América. Los registros aceptados incluían información sobre producción de leche, grasa y proteína en vacas de primera lactancia, correspondiente al período de julio de 1977 a junio de 1980.

2. Análisis

2. 1. Modelo

El mo[illegible] $\mathbf{y} = \mathbf{T}\mathbf{t} + \mathbf{A}\mathbf{a} + \mathbf{P}\mathbf{p} + \mathbf{e}$ omponentes d[illegible]

$$\mathbf{y} = \mathbf{T}\mathbf{t} + \mathbf{A}\mathbf{a} + \mathbf{P}\mathbf{p} + \mathbf{e}$$

en donde

- $\mathbf{y}$: vector de variables aleatorias observadas (producciones de leche, grasa o proteína por lactancia).
- $\mathbf{T}, \mathbf{A}, \mathbf{P}$: matrices de incidencia conocidas, correspondientes a combinaciones año-tambo-época de parto, edad al parto y padre, respectivamente.
- $\mathbf{e}$: vector de residuales asociados a cada registro.
- $\mathbf{t}$ y $\mathbf{a}$: vectores de efectos fijos desconocidos para combinaciones año-tambo-época de parto y edad de parto respectivamente.
- $\mathbf{p}$: vector de variables aleatorias desconocidas que representan mitades de valores genéticos aditivos de los padres.

Los supuestos del modelo fueron:

$$E \begin{bmatrix} \mathbf{y} \\ \mathbf{p} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{T}\mathbf{t} + \mathbf{A}\mathbf{a} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}$$

y

$$V \begin{bmatrix} \mathbf{y} \\ \mathbf{p} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_N \sigma_e^2 + \mathbf{P}\mathbf{P}'\sigma_p^2 & \mathbf{S}\sigma_p^2 & \mathbf{I}_N \sigma_e^2 \\ & \mathbf{I}_s + \sigma_p^2 & \mathbf{0} \\ & & \mathbf{I}_N \sigma_e^2 \end{bmatrix}$$

en donde

- E y V: esperanza matemática y matriz de varianza-covarianza, respectivamente.
- N: número total de observaciones.
- s: número de padres.

2. 2. Método

El método utilizado consiste en igualar sumas de cuadrados de aproximaciones a óptimas predicciones lineales insesgadas de los efectos padres a sus esperanzas matemáticas.

Las óptimas predicciones lineales insesgadas fueron descriptas por Henderson (1959). La secuencia de pasos seguidos se detalla a continuación, incluyéndose la prueba matemática del método para el modelo en cuestión. Es de destacar que Henderson (1980) hace una presentación general describiendo el método pero no prueba sus resultados.

Tampoco lo hace Hudson (1980) en el único trabajo anterior al presente en donde se utiliza esta metodología.

En la subsección 2.2.4. pueden encontrarse algoritmos útiles para programar el método en una computadora.

2. 2. 1. Absorción de los efectos correspondientes a las combinaciones año-tambo-época a partir de las ecuaciones de mínimos cuadrados generalizados

En esta etapa del cálculo, p es considerado como un vector de efectos fijos tal que:

$$V(\underset{\sim}{y}) = \underset{\sim}{I}_N \sigma_e^2$$

Entonces:

$$\underset{\sim}{T}'\underset{\sim}{T}\underset{\sim}{t}^o + \underset{\sim}{T}'\underset{\sim}{A}\underset{\sim}{a}^o + \underset{\sim}{T}'\underset{\sim}{P}\underset{\sim}{p}^o = \underset{\sim}{T}'\underset{\sim}{y} \quad \text{(I)}$$
$$\underset{\sim}{A}'\underset{\sim}{T}\underset{\sim}{t}^o + \underset{\sim}{A}'\underset{\sim}{A}\underset{\sim}{a}^o + \underset{\sim}{A}'\underset{\sim}{P}\underset{\sim}{p}^o = \underset{\sim}{A}'\underset{\sim}{y} \quad \text{(II)}$$
$$\underset{\sim}{P}'\underset{\sim}{T}\underset{\sim}{t}^o + \underset{\sim}{P}'\underset{\sim}{A}\underset{\sim}{a}^o + \underset{\sim}{P}'\underset{\sim}{P}\underset{\sim}{p}^o = \underset{\sim}{P}'\underset{\sim}{y} \quad \text{(III)}$$

son las ecuaciones de mínimos cuadrados generalizados, en donde t^o, a^o y p^o representan vectores de soluciones no únicas para combinaciones años-tambo-época, edad al parto y padres respectivamente.

De (I), destacando que (T' T) es diagonal:

$$\underset{\sim}{t}^o = (\underset{\sim}{T}'\underset{\sim}{T})^{-1} (\underset{\sim}{T}'\underset{\sim}{y} - \underset{\sim}{T}'\underset{\sim}{A}\underset{\sim}{a}^o - \underset{\sim}{T}'\underset{\sim}{P}\underset{\sim}{p}^o) \quad \text{(IV)}$$

Reemplazando t^o de (IV) en (II) y (III), y sacando factor común $\underset{\sim}{a}^o$ y $\underset{\sim}{p}^o$ se tiene:

$$\underset{\sim}{A}'\underset{\sim}{M}\underset{\sim}{A}\underset{\sim}{a}^o + \underset{\sim}{A}'\underset{\sim}{M}\underset{\sim}{P}\underset{\sim}{p}^o = \underset{\sim}{A}'\underset{\sim}{M}\underset{\sim}{y} \quad \text{(V)}$$
$$\underset{\sim}{P}'\underset{\sim}{M}\underset{\sim}{A}\underset{\sim}{a}^o + \underset{\sim}{P}'\underset{\sim}{M}\underset{\sim}{P}\underset{\sim}{p}^o = \underset{\sim}{P}'\underset{\sim}{M}\underset{\sim}{y} \quad \text{(VI)}$$

en donde

$$\underset{\sim}{M} = \underset{\sim}{I} - T\,(\underset{\sim}{T}'\underset{\sim}{T})^{-1}\,\underset{\sim}{T}'$$

2. 2. 2. Absorción de los efectos de edad

Esta segunda absorción se realizó debido a que el método exije la absorción de todos los efectos fijos del modelo.

De (V):

$$\underset{\sim}{a}^o = (\underset{\sim}{A}'\underset{\sim}{M}\underset{\sim}{A})^{-} (\underset{\sim}{A}'\underset{\sim}{M}\underset{\sim}{y} - \underset{\sim}{A}'\underset{\sim}{M}\underset{\sim}{P}\underset{\sim}{p}^o)$$

Reemplazando $\underset{\sim}{a}^o$ en (VI) y después de algebra elemental se tiene:

$$\underset{\sim}{P}'\underset{\sim}{B}\underset{\sim}{P}\underset{\sim}{p}^o = \underset{\sim}{P}'\underset{\sim}{B}\underset{\sim}{y} \quad \text{(VII)}$$

en donde

$$\underset{\sim}{B} = \underset{\sim}{M} - \underset{\sim}{M}\underset{\sim}{A}\,(\underset{\sim}{A}'\underset{\sim}{M}\underset{\sim}{A})^{-}\,\underset{\sim}{A}'\underset{\sim}{M}$$

2. 2. 3. Obtención de aproximaciones a óptimas predicciones lineales insesgadas, sus formas cuadráticas y las esperanzas de sus formas cuadráticas

Hasta aquí se aplicaron procedimientos de mínimos cuadrados generalizados corrientes. En esta subsección se sigue a Henderson (1980).

Siendo **p** un vector de efectos aleatorios, se sumó a cada elemento diagonal de $\underset{\sim}{P}'\underset{\sim}{B}\underset{\sim}{P}$ en (VII) la relación $\alpha = \sigma_e^2 / \sigma_p^2$ (estimada en trabajos previos). Esta técnica está descripta por Henderson *et al.* (1959) y permite obtener óptimas predicciones lineales insesgadas de los efectos aleatorios del modelo.

Usualmente, α es un número grande comparado con los elementos no diagonales de $\underset{\sim}{P}'\underset{\sim}{B}\underset{\sim}{B}$. Por ejemplo, para producción de leche α tenía un valor de 15, en tanto que los elementos no diagonales de $\underset{\sim}{P}'\underset{\sim}{B}\underset{\sim}{P}$ nunca superaban el valor 1. Bajo estas circunstancias, Henderson (1980) propone obtener soluciones a las ecuaciones ignorando los relativamente pequeños elementos no diagonales tal que:

$$\underset{\sim}{D}\tilde{\underset{\sim}{p}} = \underset{\sim}{r}$$

en donde

$\underset{\sim}{D}$: matriz diagonal cuyos elementos diagonales son iguales a los de $(\underset{\sim}{P}'\underset{\sim}{B}\underset{\sim}{P} + \underset{\sim}{I}\alpha)$.

$\tilde{p}$: vector de soluciones aproximadas a las ecuaciones originales.

$r = P' By$: vector de miembros derechos de la igualdad después de la absorción de efectos fijos.

Entonces:

$$\tilde{p} = D^{-1} r$$

con forma $\tilde{p} = D^{-1} r$:

$$\tilde{p}'\tilde{p} = r' D^{-1} D^{-1} r = r' D^{-2} r \quad \text{(VIII)}$$

y esperanza:

$$\begin{aligned} E\left[\tilde{p}'\tilde{p}\right] &= E\left[r' D^{-2} r\right] = \\ &= E\left[y' B' PD^{-2} P' By\right] = \\ &= E\left[y' Ly\right] = \\ &= \mathrm{tr}\ (LV) + E\left[y'\right] L\, E\left[y\right] \quad \text{(IX)} \end{aligned}$$

en donde

L: $B' PD^{-2} P'B$

V: matriz de varianza-covarianza de las observaciones

tr: suma de los elementos diagonales

Nótese que:

$$E\left[y'\right] L\ E\left[y\right] = (Tt + Aa)' L\, (Tt + Aa) =$$

$$= t' T' LTt + t' T' LAa + a' A' LTt +$$

$$+ a' A' TAa \quad \text{(X)}$$

Pero $LT = T' L = O$ pues $TM = MT' = O$, de modo que los tres primeros términos de (X) son O.

Además:

$$\begin{aligned} a' A' TAa &= a' A' B' PD^{-2} BAa = \\ &= a' A' B' PD^{-2} \\ &\quad (MA - MA\,(A'MA)^{-} A'MA)\, a \end{aligned}$$

y siendo M simétrica es idempotente, este cuarto término de (X) es O.

Entonces, (IX) es:

$$\begin{aligned} E\left[\tilde{p}'\tilde{p}\right] &= \mathrm{tr}\,(LV) = \\ &= \mathrm{tr}\,(B'PD^{-2} P'B)\,(PP'\sigma^2_p + I\sigma^2_e) \end{aligned}$$

y utilizando tr (PQ) = tr (QP) (Searle, 1971), resulta:

$$\begin{aligned} E\left[\tilde{p}'\tilde{p}\right] &= \mathrm{tr}\,(P'B'PD^{-2} P'BP\sigma^2_p + \\ &\quad + P'BB'PD^{-2} \sigma^2_e) \end{aligned}$$

Finalmente, considerando que **B** = **B'** = **B**^{2} y tomando **C** = **P'BP**, que no es otra que la matriz de coeficientes de los miembros izquierdos de las ecuaciones después de la doble absorción realizada, se tiene:

$$E\left[\tilde{p}'\tilde{p}\right] = \mathrm{tr}(CD^{-2} C\sigma^2_p + CD^{-2}\sigma^2_e) \text{(XI)}$$

2. 2. 4. Estimación del componente de varianza para padres

Al igualar la forma cuadrática (VIII) con su esperanza (XI) puede estimarse σ^2_p como:

$$\hat{\sigma}^2_p = \frac{\mathrm{tr}\,(r' D^{-2} r) - \hat{\sigma}^2_e\ \mathrm{tr}\,(CD^{-2})}{\mathrm{tr}\,(CD^{-2} C)}$$

que se calculó fácilmente tomando:

$$\mathrm{tr}\,(r' D^{-2} r) = \sum_{k=1}^{s} \left(\frac{r_k}{d_{kk}}\right)^2$$

$$\mathrm{tr}\,(CD^{-2}) = \sum_{k=1}^{s} \frac{c_{kk}}{d^2_{kk}}$$

CUADRO 2: Varianzas (en la diagonal) y covarianzas (bajo la diagonal) genéticas aditivas.

Producción de:	Leche	Grasa	Proteína
Leche	179.060		
Grasa'	759	280	
Proteína	1.018	51	160

CUADRO 3: Heredabilidades (en la diagonal) y correlaciones genéticas (bajo la diagonal).

Producción de:	Leche	Grasa	Proteína
Leche	0,213		
Grasa	0,428	0,239	
Proteína	0,792	0,801	0,189

Otros autores (Henderson, 1980; Hudson, 1980) compararon estimaciones obtenidas con este método y con el Método III de Henderson, encontrando similares resultados para los dos procedimientos.

CONCLUSIONES

Se verificó que los valores de heredabilidad para producciones de leche, grasa y proteína oscilan entre 0,20 y 0,25.

También se confirmó que las correlaciones genéticas entre aquellos caracteres son altas, lo que sugiere la posibilidad de seleccionar simultáneamente por producción de leche, grasa y proteína en los rebaños lecheros.

El método de estimación, desarrollado como alternativa a procedimientos óptimos tales el de maxima verosimilitud, no ofrece ventajas técnicas sobre otros métodos 'de momentos", ya que la cuadrática igualada a su esperanza es elegida arbitrariamente.

Sin embargo, en análisis de datos ganaderos con numerosos niveles por factor ofrece un compromiso entre optimización y costos de computación.

BIBLIOGRAFIA

1) Butcher, K. R.; F. D. Sargent and J. E. Legates, 1967. Estimates of genetic parameters milk constituents and yields. *J. Dairy Sci.*, 50: 185.

2) Cochran, W. G., 1939. The use of the analysis of variance in enumeration by sampling. *J. Am. Statistical Ass.*, 34: 492-510.

3) Fisher, R. A., 1925. Statistical methods for research workers. Oliver and Boyd, Edinburgh.

4) Henderson, C. R.; 1953. Estimation of variance and covariance components. *Biometrics*, 9: 226-252.

5) Henderson, C. R.; O. Kempthorne; S. R. Searle and C. N. Von Krosig, 1959. Estimation of invoromental and genetics trends from records subject to culling. *Biometrics*, 15: 192.

6) Henderson, C. R., 1980. A simple method for unbiased estimation of variance components in the mixed models. Presentado en el '72 nd. American Society of Animal Science Annual Meeting. Cornell University, Ithaca, New York. 13 p.

7) Hudson, J., 1980. Estimation of components of variance by Method 3 and Henderson's New Method. Sujeto a publicación. (J. Dairy Sci.).

8) Rothschild, M. F., 1978. Estimation of population parameters for first and second lactation milk production of Holstein dairy cattle free of selection bias. Inédito.

9) Searle, S. R., 1971. Linear Models, Wiley, New York, 532 p.

10) Searle, S. R. and T. R. Rounsville, 1974. A note on estimating covariance components. *Amer. Stat.* 28: 67.

11) Van Vleck, L. D.; L. H. Wadell and C. R. Henderson, 1961. Components of variance associated with milk and fat records of artificially sired Holstein daughters. *J. Anim. Sci.*, 20: 812.

12) Van Vleck, L. D., 1966. Heritability estimates of milk production with different numbers of records per sire by herd subclass. *J. Dairy Sci.* 49: 53.

13) Van Vleck, L. D., 1979. Summary of methods for estimating genetic parameters using simple statistical models. Dept. of Animal Science, Cornell University, Ithaca, New York, 62 p.

14) Wilcox, C. J.; S. N. Gaunt and B. R. Farthing, 1971. Genetic interrelationships of milk composition and yield. Interregional Publication of Northeast and Southeast Agr. Exp. Sta. Southern Cooperative Series, Bull. 155. University of Florida, Gainesville.