Capítulo 9

El modelo lineal general, mixto y sus estimaciones máximo verosímiles restringidas

Puesto que el estimador ML esta sesgado para muestras pequeñas, el método conocido como máxima verosimilitud “restringida” (REML) propuesto por Patterson y Thompson (1971) da una descripción más adecuada para un modelo lineal o mixto. REML puede considerarse un intento de tomar en cuenta la “pérdida de grados de libertad” resultante de estimar los efectos fijos. La principal ventaja de este método es que mantiene el valor de los componentes de varianza en su área de parámetros (valores positivos). La varianza fenotípica estimada vía REML tiene la siguiente expresión matemática:

$$\sigma_p^2 = \frac{\sum_{i=1}^{n} \left(y_{ij} - \overline{y} \right)^2}{n - 1}$$

Y si el modelo tiene varios efectos fijos (Blasco, 2021):

$$\sigma_p^2 = \frac{\left(y - Xb \right)' \left(y - Xb \right)}{n - Rango\left(X \right)}$$

Donde:

X es la matriz de incidencia de los efectos fijos del modelo.

Esta definición de es insesgada, ya que está asociada a **“n-1”** grados de libertad.

La función de REML se puede expresar como:

$$L_{REML}\left(\sigma_w^2, \sigma_s^2 \mid SCw\ SCs\right) = \int L\left(\mu, \sigma_w^2, \sigma_s^2\right) d\mu$$

La solución de la integral, es lo que se conoce como función de máxima verosimilitud restringida. El Log (L_{REML}) para el modelo lineal general y data balanceada es:

$$Ln(L) = -\frac{1}{2}(sk-1)\ln(2\pi) - \frac{1}{2}\ln(sk) - \frac{1}{2}\ln s\,(k-1)\ln\sigma_w^2 - \frac{1}{2}(k-1)\ln\gamma - \frac{SCw}{2\sigma_w^2} - \frac{SCs}{2\gamma}$$

Derivando e igualando a cero:

$$\frac{\partial Ln(L)}{\partial \sigma_W^2} = \frac{-s(k-1)}{2\sigma_w^2} + \frac{SCw}{2\sigma_w^4} = 0$$

$$\frac{\partial Ln(L)}{\partial \gamma} = \frac{-(s-1)}{2\gamma} + \frac{SCs}{2\gamma^2} = 0$$

Cuyas soluciones únicas son:

$$\sigma_s^2 = \frac{CMs - CMw}{k_1} \qquad \sigma_w^2 = CMw$$

Por lo tanto, siempre que los componentes de varianza estén en el área de parámetros, los estimadores REML son idénticos a los de ANOVA (para el caso balanceado).

El método REML, se ha convertido en el método preferido de análisis de datos en cría de animales, debido a sus excelentes propiedades estadísticas. Las restricciones impuestas por REML garantizan que las estimaciones estén dentro del área de parámetros, es decir, que todas las variaciones sean positivas, que todas las estimaciones de correlación estén en el rango de -1 a +1, y que todas las correlaciones parciales sean consistentes entre sí. En términos estadísticos, esto es equivalente al requisito de que la matriz de covarianza estimada sea semidefinida positiva, es decir, que ninguno de sus autovalores sea negativo, (Meyer et al., 2004)

Ejemplo 8.

Usaremos el ejemplo número 3, para mostrar las soluciones REML.

Uso del SAS.

El programa SAS para obtener los resultados es:

```
data gallinas4;
input x y @@;
cards;
1 687 2 618 3 618 4 600 5 717
1 691 2 680 3 687 4 657 5 658
1 793 2 592 3 763 4 669 5 674
1 675 2 683 3 747 4 606 5 611
1 700 2 631 3 678 4 718 5 678
1 753 2 691 3 737 4 693 5 788
1 704 2 694 3 731 4 669 5 650
1 717 2 732 3 603 4 648 5 690
;
proc mixed method=REML;
class x;
model y = ;
random x;
run;
quit;
```

Y la salida del programa es:

Tabla 24
Método REML caso balanceado

The Mixed Procedure	
Covariance Parameter Estimates	
Cov Parm	**Estimate**
x	245.69
Residual	2333.9

Resultados idénticos a los encontrados con ANOVA. Calculando la varianza fenotípica vía REML y ML, obtenemos:

$$\sigma^2_{p(REML)} = 246 + 2334 = 2580$$
$$\sigma^2_{p(ML)} = 138 + 2334 = 2472$$

Comparando estas dos varianzas, notamos que la estimación **ML** estima una varianza fenotípica más pequeña, debido a que está asociado a n grados de libertad, por lo tanto, no se justifica su uso si no se tiene una muestra lo bastante grande, como para que el sesgo de la estimación sea insignificante.

REML caso no balanceado.

Para los casos no balanceados, las soluciones de las ecuaciones REML no tienen forma cerrada, es decir, que no se puede despejar directamente los parámetros por que las ecuaciones son no lineales, para estos casos debe usarse métodos numéricos como el de Newton rapson (Searle et al. 1992) o el algoritmo de la información promedio (Jhonson y Thompon 1995).

REML modelo mixto.

Para el modelo lineal mixto, en el caso balanceado los estimadores son idénticos a los encontrados por el método ANOVA, para el caso no balanceado, se usan métodos iterativos, los cuales pueden ser el método de Newton Rapson (Searle et al. 1992), el método de Fisher Score (Searle et al. 1992) o el algoritmo de la información promedio (Jhoson y Thompson 1995).