



ESTADÍSTICA APLICADA AL **MEJORAMIENTO GENÉTICO ANIMAL**

José Raúl Pérez

ESTADÍSTICA APLICADA AL
**MEJORAMIENTO
GENÉTICO ANIMAL**



M.A. ALFREDO LEÓN VETHENCOURT
Rector

DR. RAMÓN RINCÓN GONZÁLEZ
Vicerrector

ING. ÓSCAR URDANETA
Secretario

DRA. LETICIA PRIETO
Decana de Posgrado e Investigación

DRA. TANIA MÉNDEZ
Decana de la Facultad de Ciencias Políticas, Administrativas y Sociales

DR. CÉZAR GARCÍA
Decano de la Facultad de Ingeniería

DR. RÓMULO RINCÓN
Decano de la Facultad de Ciencias Agropecuarias

ESP. HOSGLAS SANCHEZ
Directora del Fondo Editorial

ESTADÍSTICA APLICADA AL
**MEJORAMIENTO
GENÉTICO ANIMAL**

JOSÉ RAÚL PÉREZ

2024

Pérez, José Raúl

Estadística aplicada al mejoramiento genético animal. / José Raúl Pérez —1a. ed. —Maracaibo, Venezuela: Fondo Editorial Universidad Rafael Urdaneta, 2024.

78 p.: il.; 24 cm.
Incluye bibliografía.

ISBN: 978-980-8121-01-8

1. Genética animal - Estadística. 2.— Mejoramiento de las especies - Análisis estadístico. 3. Mejoramiento de las especies – Animales de granja – Estadística. II. Título.

LC: SF156.5 P9449 2024

Estadística aplicada al mejoramiento genético animal

Depósito Legal ZU2024000367

ISBN 978-980-8121-01-8

2024 © José Raúl Pérez

© Universidad Rafael Urdaneta, Fondo Editorial



Avenida 2 “El Milagro”, entrada autónoma de la Universidad Rafael Urdaneta, luego de la Biblioteca del Estado. Maracaibo, Venezuela. Teléf. (58) (261) - 2000887, Fax (58)(261) 2000868. Web:

Producción Editorial

Hosglas Sanchez, Gestión Editorial

Mibsay Contreras, Diseño y diagramación

Mileyda Briceño, Catalogación y Clasificación

Yannin Dávila, Portada



Esta obra está bajo una licencia internacional Creative Commons Atribución-NoComercial-SinDerivadas 4.0. <https://creativecommons.org/licenses/by-nc-nd/4.0/>

PRÓLOGO

Uno de los propósitos de la ciencia es describir, mejorar y predecir los procesos y fenómenos naturales del mundo en que vivimos; constituye, así, una de las actividades humanas más importantes, porque sus resultados contribuyen de manera decisiva al mejoramiento de nuestra existencia.

Este libro está escrito en esta dirección, suministrando una metodología aplicada al mejoramiento animal de nuestros rebaños. Su aplicación permitirá la toma de decisiones dentro de la empresa agropecuaria, conducente a obtener una mejora animal aplicando los principios de la genética aplicada animal y vegetal. La obra es resultado de la formación en los cursos universitarios de las ilustres Universidades Rafael Urdaneta y de la Maestría en la Universidad del Zulia cuyo aporte puede ser utilizado tanto en la enseñanza de la genética cuantitativa como en los procesos de investigación en la mejora animal. El Ingeniero en Producción Animal, José Raúl Pérez González, ha sabido ordenar la metodología descrita de manera que resulte didáctica y comprensible por quienes hagan uso de ella. Está diseñado para presentar varios procedimientos básicos usados en la genética cuantitativa. Un esfuerzo se ha hecho en cubrir las áreas aplicables a plantas y animales.

Dentro de cada procedimiento el modelo estadístico se da primero, y entonces el significado genético de los varios componentes del

modelo, las fórmulas estadísticas por computar, los procedimientos que presentan son apoyados con un ejemplo didáctico y pedagógico.

Finalmente, expreso mi sincero reconocimiento al autor de la obra por constituir uno de sus primeros logros intelectuales que muy seguramente, sin temor a equivocarme, conforman una plataforma para producir herramientas que soporten la producción animal y vegetal en nuestro País.

OSIRIS CASTEJÓN SANDOVAL, M.V

AGRADECIMIENTOS

Quiero agradecer a los profesores: Rafael María Román, Egar Sánchez, Rómulo Rincón, Atilio Aranguren, Néstor Montiel y muy especialmente a Osiris Castejón al que, con su dedicación, conocimiento, guía y aportes a la enseñanza en nuestro bello país, fue posible mi formación profesional en esta área, un honor aprender de ustedes.

Un agradecimiento especial a mis padres, hermanos y a toda mi familia que me apoyo en el escrito de estas páginas.

JOSÉ RAÚL PÉREZ

INDICE GENERAL

PRÓLOGO	7
AGRADECIMIENTOS	9
ÍNDICE TABLAS	13
INTRODUCCIÓN	15
CAPÍTULO 1	
Parámetros Genéticos	17
CAPÍTULO 2	
Análisis de regresión y sus estimaciones mínimo cuadráticas	21
CAPÍTULO 3	
El modelo lineal general	25
CAPÍTULO 4	
El modelo lineal mixto para dos factores	35
CAPÍTULO 5	
Modelo lineal mixto con interacción para dos factores	41
CAPÍTULO 6	
El modelo lineal general, mixto y sus estimaciones máximo verosímiles	45

CAPÍTULO 7

Modelo mixto para n factores aleatorios y fijos y data desbalanceada	51
--	----

CAPÍTULO 8

Modelos aleatorios con efectos anidados	57
---	----

CAPÍTULO 9

El modelo lineal general, mixto y sus estimaciones máximo verosímiles restringidas	63
--	----

CAPÍTULO 10

Criterios y métodos de selección	69
----------------------------------	----

BIBLIOGRAFÍA	81
---------------------	----

ÍNDICE TABLAS

Tabla 1. <i>Data de peso al año de toros y su progenie</i>	22
Tabla 2. <i>Método de regresión por mínimos cuadrados</i>	24
Tabla 3. <i>ANOVA de un factor caso balanceado</i>	27
Tabla 4. <i>ANOVA de un factor caso desbalanceado</i>	30
Tabla 5. <i>Datos de gallo padre</i>	31
Tabla 6. <i>Método ANOVA caso balanceado</i>	33
Tabla 7. <i>ANOVA 2 factores caso balanceado</i>	36
Tabla 8. <i>Datos de los distintos operadores</i>	38
Tabla 9. <i>Método ANOVA, caso balanceado modelo mixto</i>	40
Tabla 10. <i>ANOVA 2 factores con interacción caso balanceado</i>	42
Tabla 11. <i>Datos de raciones en distintos cerdos</i>	42
Tabla 12. <i>Método ANOVA, modelo mixto con interacción balanceado</i>	44
Tabla 13. <i>Método ML caso balanceado</i>	49
Tabla 14. <i>Método ML, caso balanceado modelo mixto</i>	50
Tabla 15. <i>Método ML, modelo mixto con interacción balanceado</i>	50
Tabla 16. <i>Datos de pesos al destete en vacunos</i>	53
Tabla 17. <i>Método ML para modelo mixto no balanceado</i>	54
Tabla 18. <i>Prueba de hipótesis para efectos fijos</i>	54
Tabla 19. <i>LSMEANS para sexo</i>	55
Tabla 20. <i>Datos de padres y madres de gallinas White rock</i>	59
Tabla 21. <i>Cuadrado del Anova</i>	60
Tabla 22. <i>Método ANOVA modelo anidado balanceado</i>	61
Tabla 23. <i>Método ML modelo anidado balanceado</i>	62
Tabla 24. <i>Método REML caso balanceado</i>	66
Tabla 25. <i>Datos de toros</i>	74
Tabla 26. <i>Salida índice de selección</i>	79

INTRODUCCIÓN

La crianza de animales ha contribuido al mejoramiento de la producción ganadera desde hace muchos años. Cuando se cuenta con registros de producción, para dirigir la selección y determinar el cambio, los resultados son prometedores. La combinación de la cría (genética), nutrición y manejo produce un éxito notable en las ganaderías que aplican de manera correcta estos principios, por lo tanto, es de esperarse que se obtenga un desempeño mayor.

La cría de animales, es la aplicación del conocimiento científico al mejoramiento genético de las especies de granja, la cual suministra, las bases que deben guiar a un productor, a mejorar las características fenotípicas de los rebaños. El objetivo que persiguen los productores es convertir subproductos animales (comestibles o no) en alimento nutritivo para los seres humanos, por lo tanto, el mejoramiento animal tiene gran importancia para cumplir este fin.

En zootecnia, es de vital interés, mejorar las características de importancia económica de los animales de un rebaño particular. Cuando se seleccionan los mejores fenotipos para un rasgo particular, se espera que las características de la siguiente generación de animales, sean superiores. A medida que se desarrollaron métodos de estimación de parámetros más eficientes, la estimación de los parámetros genéticos se ha hecho más precisa, por lo tanto, tiene sentido estimar parámetros estadísticos según lo dictaminen los datos y no según algún método “favorito”.

Este libro trata sobre el análisis estadístico aplicado al mejoramiento genético animal, donde se explica de manera detallada como calcular los parámetros genéticos de manera manual y con el paquete de análisis estadístico SAS (Statistical Analysis System), con el objeto, de poder realizar un mejoramiento genético en los animales de granja de manera eficiente.

Los temas a tratar en las siguientes páginas son todos relacionados a la estimación de parámetros (de regresión o de componentes de varianza) en diferentes modelos y estructuras de datos (modelos para diseños balanceados y desbalanceados), por lo tanto, el lector debe tener nociones de álgebra matricial y estadística para la comprensión completa del texto

CAPITULO 1

PARÁMETROS GENÉTICOS

La varianza es el elemento más importante de las valoraciones genéticas en los animales, ya que es el dato fundamental con el que se realiza un correcto mejoramiento genético. En estadística se conoce la varianza como una medida de dispersión, que indica cuanto se desvían los datos con respecto del promedio, y tiene la siguiente fórmula matemática:

$$\sigma^2 = \frac{SC_{total}}{GL_{total}}$$

Donde:

SC_{total} = suma de cuadrados totales.

GL_{total} = grados de libertad totales.

σ^2 = varianza.

La suma de los cuadrados es simplemente una suma de números al cuadrado, donde se suma todos los valores elevados al cuadrado y se le resta un factor de corrección. Esto en términos matemáticos es:

$$SC = \sum_{i=1}^n (y_i - \bar{y})^2 = y_1^2 + \dots + y_n^2 - \frac{(\sum y)^2}{n}$$

Donde:

n = tamaño de la muestra.

Los grados de libertad se definen como la cantidad de información con la cual se calcula la varianza, y para varianzas insesgadas es igual a restarle uno al número de datos, en términos matemáticos esto es:

$$GL = n - 1$$

Por lo tanto, la varianza queda definida como:

$$\sigma^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n - 1}$$

Es decir, la varianza es un cociente entre la suma de los cuadrados y los grados de libertad.

La varianza σ^2 expresa la desviación de los datos con respecto de la media al cuadrado, ante este problema una medida de dispersión más apropiada es la desviación estándar o típica, la cual es simplemente la raíz cuadrada de la varianza, en términos matemáticos se expresa de la siguiente forma:

$$\sigma = \sqrt{\sigma^2}$$

Donde:

σ = desviación estándar

La variabilidad en los registros productivos en los animales, es lo que brinda la oportunidad de mejorar, es decir, el hecho de que existan animales que produzcan más que otros, es lo que nos permite realizar un correcto mejoramiento genético. El conocimiento de la proporción de la variabilidad que es de origen genético es, entonces, un parámetro de mucho interés para los genetistas, ya que, indica si las diferencias observadas en los fenotipos de un grupo de animales se pueden heredar. Este parámetro se conoce como heredabilidad y toma la siguiente expresión matemática:

$$h^2 = \frac{\sigma_a^2}{\sigma_p^2}$$

Donde:

σ_a^2 = Varianza genética aditiva.

σ_p^2 = varianza fenotípica.

Mientras la heredabilidad sea más alta, más se hereda el carácter de un grupo de animales a sus hijos, por lo tanto, es un indicativo de que tan buena es la respuesta a la selección. Según Gutiérrez (2010) la heredabilidad puede ser alta, moderada o baja, y solo puede tomar valores entre cero y uno.

Según Becker (1986) la varianza entre toros σ_s^2 es $\frac{1}{4}$ de la varianza genética aditiva, por lo tanto, cuando se calcula la varianza σ_s^2 , y sabiendo que $\sigma_s^2 + \sigma_w^2 = \sigma_p^2$ se puede calcular la heredabilidad con la siguiente formula:

$$h^2 = \frac{4\sigma_s^2}{\sigma_s^2 + \sigma_w^2}$$

Otro parámetro de interés en mejora genética es llamado Índice de Constancia (repetibilidad), el cual indica cuánto de la varianza total entre medidas de un mismo animal es debida al efecto de factores que son permanentes (por lo tanto, lleva una varianza ambiental permanente (σ_e^2) en su fórmula).

En general este índice es un indicador del número de registros, necesarios para aplicar la selección. Su expresión matemática es:

$$R = \frac{\sigma_G^2 + \sigma_e^2}{\sigma_p^2}$$

Donde:

σ_g^2 = varianza genética total.

CAPÍTULO 2

ANÁLISIS DE REGRESIÓN Y SUS ESTIMACIONES MÍNIMO CUADRÁTICAS

Los estimadores para los parámetros genéticos se pueden obtener con el método de mínimos cuadrados. Este método escoge los estimadores que reducen al máximo la suma de cuadrados del error.

El modelo de regresión lineal simple es:

$$y_i = \alpha + b_1 x_i + \varepsilon_i$$

Donde:

y = Registros de los hijos (variable respuesta)

α = intercepto

b_1 = Coeficiente de regresión

x = Registros del padre (variable de regresión)

ε = Error residual

Las ecuaciones normales de mínimos cuadrados en álgebra matricial son Montgomery (2005):

$$(X'X)b = (X'y)$$

Cuya solución única es:

$$b = (X'X)^{-1}(X'y)$$

El coeficiente de regresión padre-hijo es la mitad de la heredabilidad, por lo tanto, para obtener la estimación completa usamos la siguiente formula de Becker (1986):

$$h^2 = 2.(b)$$

Cabe recordar que el valor encontrado es una estimación, por lo tanto, debe ser tomada con un tamaño de muestra adecuado.

Ejemplo 1.

Se tiene una data de peso al año de toros Nelore y su progenie, se quiere saber si seleccionar para este rasgo tendrá beneficios en el incremento del peso de las siguientes generaciones, los datos se muestran en la Tabla 1:

Tabla 1
Data de peso al año de toros y su progenie

Peso al año de toros	Peso al año de progenie
1006 libras	956 libras
1095 libras	916 libras
1018 libras	944 libras
1029 libras	1036 libras
1011 libras	1054 libras
1004 libras	923 libras
1005 libras	938 libras
1124 libras	1003 libras
964 libras	899 libras
1048 libras	941 libras

Para los datos del ejemplo 1, las ecuaciones normales de mínimos cuadrados son:

$$\begin{bmatrix} 10 & 10304 \\ 10304 & 10637364 \end{bmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 9610 \\ 9906944 \end{pmatrix}$$

La solución viene de:

$$\begin{pmatrix} \alpha \\ b \end{pmatrix} = \begin{bmatrix} 10 & 10304 \\ 10304 & 10637364 \end{bmatrix}^{-1} \begin{pmatrix} 9610 \\ 9906944 \end{pmatrix}$$

Y, la solución es:

$$\begin{pmatrix} \alpha \\ b \end{pmatrix} = \begin{pmatrix} 715.208 \\ 0.2385 \end{pmatrix}$$

Por lo tanto, el índice de herencia es:

$$h^2 = 2(0.238) = 0.47$$

La heredabilidad del peso al año es de 47%, es decir, que las diferencias observadas en los pesos al año, son ocasionadas por efectos genéticos aditivos en un 47%, por lo tanto, un 53% es de origen ambiental. Cabe esperar un avance de la selección para este rasgo.

Uso SAS (*Statistical Analysis System*)

```
data pesos;
input Ppadre Phijos;
datalines;
1006      956
1095      916
1018      944
1029      1036
1011      1054
1004      923
1005      938
1124      1003
964       899
1048      941
;
```

```

proc glm data=pesos;
title Cálculo de la heredabilidad;
model Phijos= Ppadre/ ssl xpx i solution;
run;
quit;

```

La salida es:

Tabla 2
Método de regresión por mínimos cuadrados

Parameter	Estimate	Standar Error	t value	Pr> t
Intercept	715.2082455	393.6355915	1.82	0.1068
Ppadre	0.2385401	0.3816606	0.63	0.5494

Resultados idénticos a los encontrados de manera manual.

CAPÍTULO 3

EL MODELO LINEAL GENERAL

En general, los componentes de varianza se estiman de los datos, por lo tanto, los diferentes métodos que se pueden aplicar pueden generar valores diferentes de la heredabilidad o del índice de constancia.

El modelo estadístico lineal, para una vía de clasificación es:

$$Y_{ij} = \mu + s_i + \varepsilon_{ij}$$

Dónde:

μ = Media o promedio del rebaño

Y_{ij} = Variable respuesta o rasgo particular

s_i = Efecto del i-esimo semental

ε_{ij} = Efecto aleatorio denominado error ambiental

La solución de las ecuaciones normales de mínimos cuadrados para este modelo según Montgomery (2005) es:

$$\begin{aligned}\hat{\mu} &= \bar{y} \\ \hat{s} &= \bar{y}_{i.} - \bar{y}_{..}\end{aligned}$$

La suma de cuadrados para este modelo lineal se consigue de la siguiente manera, llamada *descomposición de la suma de cuadrados* de Montgomery (2005):

$$\begin{aligned}
\sum \sum (y_{ij} - \bar{y}_{..})^2 &= \sum \sum \left[(\bar{y}_{i.} - \bar{y}_{..}) + (y_{ij} - \bar{y}_{i.}) \right]^2 \\
&= \sum \sum \left[(\bar{y}_{i.} - \bar{y}_{..})^2 + (y_{ij} - \bar{y}_{i.})^2 + 2(\bar{y}_{i.} - \bar{y}_{..})(y_{ij} - \bar{y}_{i.}) \right] \\
&= r \sum (\bar{y}_{i.} - \bar{y}_{..})^2 + \sum \sum (y_{ij} - \bar{y}_{i.})^2 + \sum \sum (\bar{y}_{i.} - \bar{y}_{..})(y_{ij} - \bar{y}_{i.})
\end{aligned}$$

Los productos cruzados son cero, por lo tanto:

$$\sum (\bar{y}_{i.} - \bar{y}_{..}) = y_{i.} - n\bar{y}_{..} = y_{i.} - \frac{ny_{i.}}{n} = 0$$

El resultado es la suma de cuadrados para cada fuente de variación:

$$\sum \sum (y_{ij} - \bar{y}_{..})^2 = r \sum (\bar{y}_{i.} - \bar{y}_{..})^2 + \sum \sum (y_{ij} - \bar{y}_{i.})^2$$

Los grados de libertad, para cada fuente de variación son {Becker, 1986):

Total = N-1

Padres = S - 1

Hijos = N - S

Las varianzas o cuadrados medios son un cociente entre las sumas de cuadrados y los grados de libertad.

A partir de estas fórmulas, se puede construir el cuadrado del ANOVA (Castejón, 2008):

Tabla 3
ANOVA de un factor caso balanceado

Fuente de variación	Suma de cuadrados	Grados de libertad	Varianzas (CM)	Esperanza matemática de las varianzas (CM)
Reproductores (S)	$\sum \frac{y_{j.}^2}{k} - \frac{(\sum y_{..})^2}{n}$	s-1	$\frac{SC_s}{gl_s}$	$\sigma_w^2 + k_1 \cdot \sigma_s^2$
Hijos (w)	$SC_{total} - SC_s$	n-s	$\frac{SC_w}{gl_w}$	σ_w^2
Total	$\sum y_{ij}^2 - \frac{(\sum y_{..})^2}{n}$	n-1		

Donde K es el número de réplicas del diseño.

Los componentes de varianza para este modelo lineal, se estiman calculando el valor esperado de las varianzas (Montgomery 2005):

$$E(CMs) = \frac{1}{s-1} E \left[\frac{1}{2} \sum_{i=1}^s \left(\sum_{j=1}^K \mu + s_i + e_{ij} \right)^2 - \frac{1}{2} \sum_{i=1}^s \left(\sum_{j=1}^K \mu + s_i + e_{ij} \right)^2 \right]$$

Cuya solución es:

$$E(CMs) = \sigma_w^2 + k_1 \sigma_s^2$$

Los estimadores basados en el método ANOVA aplicado a conjuntos de datos balanceados poseen muchas propiedades interesantes: son insesgados y de mínima varianza. Sin embargo, si el conjunto de datos es desbalanceado, como sería el caso de tener observaciones perdidas o simplemente que la cantidad de animales no sea suficiente para correr un diseño balanceado, el ANOVA pierden algunas de sus propiedades (Searle et al., 1992).

El Dr. Charles Henderson (Henderson, 1953) propuso tres métodos análogos al ANOVA para calcular los componentes de la varianza

en datos desbalanceados, a los que denomino métodos I, II y III de Henderson, de los cuales, el III es el más usado, porque es aplicada a datos muy desbalanceados y modelos mixtos.

La imposibilidad de obtener una data balanceada (debido a que generalmente cada padre, tiene diferente número de crías), imposibilita obtener sumas de cuadrados que estén libres de prejuicios por los métodos balanceados (ANOVA), por lo tanto, en la práctica los parámetros se estiman con métodos específicos para datos desbalanceados.

Descripciones del Método III de Henderson para un modelo con “n” variables de clasificación y data desbalanceada pueden encontrarse en Henderson (Henderson, 1953), Searle (Searle et al., 1992), Harvey (Harvey, 1960) e Steel y Torrie (Steel y Torrie, 1985). Este método consiste en construir las ecuaciones normales de mínimos cuadrados para los datos, e imponer alguna restricción para romper la dependencia lineal del sistema de ecuaciones, con lo cual se calculan sumas de cuadrados ajustadas por constantes. Entre las restricciones que se pueden usar está (otra forma es usar una inversa generalizada de $X'X$ y obtener $b = (X'X)^{-} - (X'y)$ tal como lo hace el GLM del SAS):

$$\sum s = 0$$

Para calcular las constantes de ajuste, de las sumas de cuadrados, se debe obtener los elementos inversos de las ecuaciones normales de mínimos cuadrados:

$$\sum C_{ij} Y_i = \hat{c}_i$$

Dónde:

C_{ij} = Elemento inverso para la hilera i y la columna j de la matriz inversa.

$\hat{c}_i$ = Constante estimada

Existen varios métodos para obtener una inversa o inversa generalizada de una matriz, sin embargo, en este trabajo no los discutiremos. Estos métodos pueden ser consultados en los textos de Searle (1982) y Harville (1997).

La suma de cuadrados para S, se puede obtener resolviendo la siguiente expresión matemática:

$$SC_s = B' Z^{-1} B$$

Dónde:

B = Vector de las constantes estimadas

Z^{-1} = Inversa del segmento de la inversa de la matriz de varianza-covarianza

La suma de cuadrados total es:

$$SCT = y'y - FC$$

Donde:

$$FC = \frac{(\sum y_{..})^2}{n}$$

Y, la suma de cuadrados del residual es:

$$SCe = y'y - R(\mu, si)$$

Y, $R(\mu, si)$ viene dado por:

$$R(\mu, s_i) = b' \times (X'y)_r$$

Por lo tanto, la suma de cuadrados del factor es:

$$SC_s = SCT - SCe$$

El coeficiente K (para casos desbalanceados) se estima de la siguiente manera:

$$k_1 = \frac{1}{s-1} \left(n - \frac{\sum n_i^2}{n} \right)$$

Con esto se construye el cuadrado del ANOVA para el método III:

Tabla 4
ANOVA de un factor caso desbalanceado

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrados medios	Esperanza matemática de las varianzas
Reproductores (s)	$SCT - SCe = R(\mu, s) - R(\mu)$	s-1	$\frac{SC_s}{gl_s}$	$\sigma_w^2 + k_1 \cdot \sigma_s^2$
Hijos (w)	$y'y - R(\mu, s)$	n-s	$\frac{SC_w}{gl_w}$	σ_w^2
Total	$y'y - FC$	n-1		

Los componentes de varianza se obtienen igualando las varianzas con sus esperanzas matemáticas (Searle et al., 1992):

$$CM_w = \sigma_w^2$$

$$CM_s = \sigma_w^2 + k_1 \cdot \sigma_s^2$$

La solución de este sistema de ecuaciones es:

$$\sigma_s^2 = \frac{CM_s - CM_w}{k_1}$$

$$\sigma_w^2 = CM_w$$

Ejemplo 2.

Se busca calcular, el índice de herencia del peso (gramos) de gallinas ponedoras, después de un determinado período. Se escogieron 5 reproductores al azar y se pesó su progenie a las 8 semanas de vida, se tomaron 8 observaciones por gallo (padre) y los datos de la progenie se muestran en la siguiente tabla:

Tabla 5
Datos de gallo padre

Padre				
1	2	3	4	5
687	618	618	600	717
691	680	687	657	658
793	592	763	669	674
675	683	747	606	611
700	631	678	718	678
753	691	737	693	788
704	694	731	669	650
717	732	603	648	690

Modelo lineal general, ejemplo 2.

Sumas de cuadrados:

$$FC = \frac{27331^2}{40} = 18674589$$

$$SC_s = \frac{5720^2 + 5321^2 + 5564^2 + 5466^2 + 5260^2}{8} - 18674589 = 18691786 - 18674589 = 17197 \text{ (padres)}$$

$$SC_{total} = 687^2 + \dots + 690^2 - \frac{27331^2}{40} = 98884$$

$$SCW = 98884 - 17197 = 81687 \text{ (hijos)}$$

Grados de libertad:

$$\begin{aligned} GLs &= 5 - 1 = 4 \\ GL_{total} &= 40 - 1 = 39 \\ GLw &= 40 - 5 = 35 \end{aligned}$$

Cuadrados medios:

$$\begin{aligned} CMs &= \frac{17197}{4} = 4299.25 \\ CMw &= \frac{81687}{35} = 2333.9 \end{aligned}$$

Igualando los cuadrados medios, a sus valores esperando y resolviendo el sistema de ecuaciones, obtenemos:

$$\begin{aligned} \sigma_w^2 &= 2333.896 \\ \sigma_s^2 &= \frac{4299.25 - 2333.9}{8} = 245.668 \end{aligned}$$

Otra manera es resolver el siguiente sistema de ecuaciones:

$$P^{-1} \begin{pmatrix} CMs \\ CMw \end{pmatrix} = \begin{bmatrix} \sigma_w^2 \\ \sigma_s^2 \end{bmatrix}$$

Donde:

$$P = \begin{bmatrix} 1 & k \\ 1 & 0 \end{bmatrix}$$

Por lo tanto:

$$\begin{bmatrix} \sigma_w^2 \\ \sigma_s^2 \end{bmatrix} = \begin{bmatrix} 1 & k \\ 1 & 0 \end{bmatrix}^{-1} \begin{pmatrix} 4299.4 \\ 2333.896 \end{pmatrix} = \begin{pmatrix} 2333.896 \\ 245.688 \end{pmatrix}$$

Lo que lleva al cálculo del índice de herencia:

$$h^2 = \frac{4(245.688)}{2333.896 + 245.688} = 0.38$$

3.2 Uso del SAS.

```

data gallinas;
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 GLM;
class x;
model y = x/ ss3;
random x;
run;
proc varcomp method=type1;
class x;
model y = x;
run;

```

Tabla 6
Método ANOVA caso balanceado

Type 1 Analysis of Variance				
Source	DF	Sum of Squares	Mean Square	Expected Mean Square
x	4	17198	4299.400000	Var (Error) + 8 Var (x)
Error	35	81686	2333.896429	Var (Error)
Corrected total	39	98884	.	.
Type 1 Estimates				
Variance Component			Estimate	
Var (x)			245.68795	
Var (Error)			2333.9	

Se puede notar resultados muy parecidos a los expuestos de manera manual (las diferencias son por el redondeo).

CAPÍTULO 4

EL MODELO LINEAL

MIXTO PARA DOS FACTORES

Cuando se asume la existencia de factores fijos, el modelo estadístico lineal es conocido como modelo mixto. En zootecnia consideramos un efecto fijo a cualquier factor medible que pueda causar variabilidad en los registros de los animales, por ejemplo: época del año, edad, número de lactancia entre otros. Cuando se emplea un modelo mixto el efecto de los factores fijos es removido de la estimación de los factores aleatorios, lo que genera, estimadores ajustados a las condiciones de cada sistema de producción animal.

Los métodos descriptos anteriormente pueden ser aplicados para la obtención de parámetros en el modelo mixto, en general el modelo estadístico lineal de 2 factores (sin interacción) es:

$$Y_{ijk} = \mu + s_i + b_j + \varepsilon_{ijk}$$

Dónde:

μ = Media o promedio del rebaño

Y_{ijk} = Variable respuesta o rasgo particular

b_j = Efecto del j-esimo factor fijo

s_i = Efecto del i-esimo semental

ε_{ijk} = Efecto aleatorio denominado error ambiental

Con el método ANOVA la esperanza de los CM (para los efectos aleatorios) son:

$$E(CM_w) = \sigma_w^2$$

$$E(CM_s) = \sigma_w^2 + b k \sigma_s^2$$

Los componentes de varianza para los efectos aleatorios (Searle 1992), pueden obtenerse igualando las varianzas con sus valores esperados:

$$\sigma_w^2 = E(CM_w)$$

$$\sigma_s^2 = \frac{CM_s - CM_w}{bk}$$

El cuadrado del ANOVA para este caso es:

Tabla 7
ANOVA 2 factores caso balanceado

FV	SC	GL	CM	E (CM)
Fijo	$\sum sk(\bar{y}_{i..} - \bar{y} \dots)^2$	b-1	CMb	
Padres	$\sum bk(\bar{y}_{.j.} - \bar{y} \dots)^2$	s-1	CMs	$E(CM_s) = \sigma_w^2 + b k \sigma_s^2$
Hijos	$\sum \sum \sum (y_{ijk} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y} \dots)^2$	Sbk- s-b+1	CMw	$E(CM_w) = \sigma_w^2$
Total	$\sum \sum \sum (y_{ijk} - \bar{y} \dots)^2$	n-1		

El método III de Henderson es el método ANOVA para estimar los componentes de la varianza utilizando formas cuadráticas basadas en **R (b)**. Sin tener que saber o formular qué elementos de **b** son efectos fijos y que son aleatorios, la generalización de **var (b)** resulta

útil al considerar lo que necesitamos para cualquier forma de método de estimación ANOVA, al saber, el valor esperado de una forma cuadrática de la suma de cuadrados reducida (Searle et al., 1992).

En el método III las sumas esperadas de cuadrados de la forma **E (R (b2 I b1))** para nunca involucran **b1**. Por lo tanto, siempre que formulemos **b1** para incluir siempre los efectos fijos de nuestro modelo, **E (R (b2 I b1))** nunca incluye efectos fijos. Por este medio, el Método III evita la deficiencia del método I al no ser adecuado para modelos mixtos. Y por la naturaleza general de **R (b2 I b1)**, no hay restricción, como la hay con el método II, de tener que prescindir de interacciones fijas por aleatorias (Searle et al., 1992). Esto hace que el método III se perfecto para ajustar datos a cualquier modelo lineal, sea fijo, aleatorio o mixto (si no existe interacción el método II de Henderson es aplicable, y por su mayor sencillez, es recomendable). La aplicación de modelos lineales en la estimación de los valores de cría, pueden encontrarse en Román y Aranguren (Román y Aranguren, 2014) y Yáñez y Verde (Yáñez y Verde, 2014).

Ejemplo 3.

Se busca estimar el índice de constancia de la medida de suavidad en el musculo del pecho en pavos de raza Broad. Diez medidas de suavidad fueron medidas en el musculo del pavo, como las medidas fueron tomadas por operadores diferentes, ese efecto fue considerado como fijo. Los datos son:

Tabla 8
Datos de los distintos operadores

Pavo (1)	Operador (2)																		
		y	1	2	y	1	2	y	1	2	y	1	2	y	1	2	y	1	2
1	1	2.3	2	1	2.2	3	1	2.1	4	1	2.3	5	1	2.1	6	1	2		
1	2	2.2	2	2	3.1	3	2	2.6	4	2	2.4	5	2	2	6	2	1.9		
1	1	2.2	2	1	2.7	3	1	2.7	4	1	2.1	5	1	2.2	6	1	2.6		
1	2	3	2	2	2.2	3	2	2	4	2	2.6	5	2	2	6	2	2.5		
1	1	2.4	2	1	2.5	3	1	1.9	4	1	2	5	1	2	6	1	2.3		
1	2	2.8	2	2	2.6	3	2	1.9	4	2	1.7	5	2	1.8	6	2	2		
1	1	2.6	2	1	2.9	3	1	1.8	4	1	2.4	5	1	2	6	1	2.5		
1	2	2.5	2	2	2.3	3	2	2.3	4	2	2.1	5	2	1.7	6	2	2.2		
1	1	2.2	2	1	2.2	3	1	2.1	4	1	2.5	5	1	1.8	6	1	2		
1	2	2.3	2	2	3	3	2	2.2	4	2	2.7	5	2	1.7	6	2	1.7		

Modelo lineal mixto, ejemplo 3.

Sumas de cuadrados:

$$FC = \frac{(135.6)^2}{60} = 306.45$$

$$SC_{animal} = \frac{(24.5)^2 + \dots + (21.7)^2}{10} - \frac{(135.6)^2}{60} = 2.6$$

$$SC_{operador} = 0.002$$

$$SC_{total} = 2.3^2 + \dots + 1.7^2 - \frac{(135.6)^2}{60} = 7.064$$

$$SC_w = 7.064 - 2.6 - 0.002 = 4.46$$

Grados de libertad:

$$GL_{total} = 60 - 1 = 59$$

$$GL_{operador} = 2 - 1 = 1$$

$$GL_{animal} = 6 - 1 = 5$$

$$GL_w = 60 - 6 - 2 + 1 = 53$$

Cuadrados medios:

$$CM_{animal} = \frac{2.6}{5} = 0.52$$

$$CM_w = \frac{4.46}{53} = 0.084$$

Componentes de varianza:

$$\sigma_{animal}^2 = \frac{0.520 - 0.084}{10} = 0.0436$$

$$\sigma_w^2 = 0.084$$

Y, el índice de constancia es:

$$R = \frac{0.0436}{0.084 + 0.0436} = 0.34$$

Uso del SAS.

```
data pavos1;
input pavo operador y @@;
datalines;
1 1 2.3 2 1 2.2 3 1 2.1 4 1 2.3 5 1 2.1 6 1 2.0
1 2 2.2 2 2 3.1 3 2 2.6 4 2 2.4 5 2 2.0 6 2 1.9
1 1 2.2 2 1 2.7 3 1 2.7 4 1 2.1 5 1 2.2 6 1 2.6
1 2 3.0 2 2 2.2 3 2 2.0 4 2 2.6 5 2 2.0 6 2 2.5
1 1 2.4 2 1 2.5 3 1 1.9 4 1 2.0 5 1 2.0 6 1 2.3
1 2 2.8 2 2 2.6 3 2 1.9 4 2 1.7 5 2 1.8 6 2 2.0
1 1 2.6 2 1 2.9 3 1 1.8 4 1 2.4 5 1 2.0 6 1 2.5
1 2 2.5 2 2 2.3 3 2 2.3 4 2 2.1 5 2 1.7 6 2 2.2
1 1 2.2 2 1 2.2 3 1 2.1 4 1 2.5 5 1 1.8 6 1 2.0
1 2 2.3 2 2 3.0 3 2 2.2 4 2 2.7 5 2 1.7 6 2 1.7

;

proc varcomp method=type1;
class pavo operador;
model y=operador pavo / fixed=1;
run;
quit;
```

Tabla 9
Método ANOVA, caso balanceado modelo mixto

Dependent Variable: y				
Type 1 Analysis of Variance				
Source	DF	Sum of Squares	Mean Square	Expected Mean Square
operador	1	0.002667	0.002667	Var(Error)+Q(ope- rador)
pavo	5	2.596000	0.519200	Var(Error)+10 Var (pavo)
Error	53	4.465333	0.084252	Var(Error)
Corrected total	59	7.064000	.	.
Type 1 Estimates				
Variance Component			Estimate	
Var (pavo)			0.04349	
Var (Error)			0.08425	

Se puede notar, que los resultados son idénticos a los encontrados de manera manual.

CAPÍTULO 5

MODELO LINEAL MIXTO CON INTERACCIÓN PARA DOS FACTORES

El nuevo modelo aditivo lineal, que toma en cuenta la interacción es:

$$Y_{ijk} = \mu + s + b_j + (sb)_{ij} + \varepsilon_{ijk}$$

Dónde:

μ = Media o promedio del rebaño

Y_{ijk} = Variable respuesta o rasgo particular

b_j = Efecto del j-esimo factor fijo

sb = componente de interacción

s_i = Efecto del i-esimo semental.

ε_{ijk} = Efecto aleatorio denominado error ambiental

Y, los componentes de la varianza, usando el método ANOVA son:

$$\begin{aligned}\sigma_w^2 &= CM_w \\ \sigma_s^2 &= \frac{CMs - CMsb}{bk} \\ \sigma_{sb}^2 &= \frac{CMsb - CMw}{k}\end{aligned}$$

Y la SC de interacción es:

$$SC\ Interaccion(s \times b) = \sum \sum k(\bar{y}_{ij} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2$$

La nueva SCw es:

$$SCw = SCTotales - SCs - SCb - SCsb$$

El nuevo cuadrado del ANOVA es:

Tabla 10
ANOVA 2 factores con interacción caso balanceado

FV	GL	SC	CM
Fijo	b-1	SCb	SCb / (b-1)
Padres- Animal	s-1	SCs	SCs / (s-1)
Interacción	(s-1) (b-1)	SCsb	SCsb / ((s-1)(b-1))
Error – Hijos	Gl total – GL resto	SCw	SCw / (bs(k-1))
Total	N-1	SCtotal	

Ejemplo 4.

En el siguiente ejemplo se usaron 2 raciones de alimento concentrado en cerdos diferentes. La cantidad de alimento es un efecto fijo y el efecto directo del animal es aleatorio. Los datos son:

Tabla 11
Datos de raciones en distintos cerdos

Ración	Verraco 1	Verraco 2	Verraco 3
1	5	2	3
1	6	3	3
1	6	5	1
1	3	6	2
1	2	7	4
2	2	8	4
2	3	8	4
2	4	9	6
2	3	7	6
2	2	8	7

Sumas de cuadrados:

$$FC = \frac{(139)^2}{30} = 644.0333$$

$$SC_{total} = 5^2 + \dots + 7^2 - 644.033 = 144.96$$

$$SC_{animal} = \frac{(36)^2 + (63)^2 + (40)^2}{10} - 644.0333 = 42.4667$$

$$SC_{racion} = \frac{(58)^2 + (81)^2}{15} - 644.0333 = 17.6333$$

$$SC_{animal \times racion} = 37.26666$$

$$SC_w = 144.96 - 42.4667 - 17.6333 - 37.2666 = 47.5934$$

Grados de libertad:

$$GL_{racion} = 2 - 1 = 1$$

$$GL_{animal} = 3 - 1 = 2$$

$$GL_{animal \times racion} = (3 - 1)(2 - 1) = 2$$

$$GL_{total} = 30 - 1 = 29$$

$$GL_w = 29 - 5 = 24$$

Cuadrados medios:

$$CM_{racion} = \frac{17.6333}{1} = 17.6333$$

$$CM_{animal} = \frac{42.4667}{2} = 21.23335$$

$$CM_{animal * racion} = \frac{37.26666}{2} = 18.6333$$

$$CM_w = \frac{47.5934}{24} = 1.983$$

Componentes de varianza:

$$\sigma_w^2 = 1.983$$

$$\sigma_{\text{animal}}^2 = \frac{21.23335 - 18.6333}{10} = 0.26$$

$$\sigma_{\text{animal} \times \text{racion}}^2 = \frac{18.6333 - 1.983}{5} = 3.33$$

Uso del SAS.

```
proc varcomp method=type1;
class racion cerdo;
model y= racion cerdo racion*cerdo/ fixed=1;
run;
quit;
```

Tabla 12
Método ANOVA, modelo mixto con interacción balanceado

Type 1 Analysis of Variance				
Source	DF	Sum of Squares	Mean Square	Expected Mean Square
racion	1	17.633333	17.633333	Var(Error)+5 Var(racion*cerdo)+Q(racion)
cerdo	2	42.466667	21.233333	Var(Error)+5 Var(racion*cerdo)+10 Var(cerdo)
racion*cerdo	2	37.266667	18.633333	Var(Error)+5 Var(racion*cerdo)
Error	24	47.600000	1.983333	Var(Error)
Corrected total	29	144.966667	.	.
Type 1 Estimates				
Variance Component			Estimate	
Var (pavo)			0.26000	
Var(racion x cerdo)			3.33000	
Var (Error)			1.98333	

CAPÍTULO 6

EL MODELO LINEAL GENERAL, MIXTO Y SUS ESTIMACIONES MÁXIMO VEROSÍMILES

Hartley y Rao (1967) propusieron el método de máxima verosimilitud (ML) para estimar parámetros en modelos lineales mixtos, este método se basa en maximizar la función de verosimilitud de los parámetros a estimar. Lo anteriormente dicho es un tanto complicado, más en el marco de estimación de componentes de varianza genéticos, pero de forma general el método ML se basa en derivar el logaritmo de la función de verosimilitud (donde se asume normalidad) en función de los parámetros y resolver el sistema de ecuaciones resultante, lo que genera un estimador de la varianza fenotípica que viene dado por:

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

Y en el caso de muchos efectos fijos (Blasco, 2021):

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

Esta definición de σ_p^2 esta sesgada, ya que está asociada a “n” grados de libertad, por lo tanto, es un estimador confiable solo con

muestras de gran tamaño. Una descripción muy buena del método se puede encontrar en Searle et al. (1992).

La función de verosimilitud es igual al producto de la función densidad conjunta de los datos, asociados a los parámetros a estimar, esto es para el modelo lineal general, igual a:

$$L = \prod f(\mu, \sigma_w^2, \sigma_s^2)$$

Asumiendo normalidad la función sigue la distribución normal, por lo tanto, todos los datos se distribuyen normalmente. Los parámetros a estimar se encuentran maximizando esta función, pero se alcanza el mismo punto, maximizando su logaritmo, por lo tanto, en la práctica se maximiza:

$$Ln[L]$$

El logaritmo de L para los parámetros del modelo (lineal general) y la distribución normal con data balanceada es:

$$\text{Log}(L) = -\frac{1}{2}N \ln(2\pi) - \frac{1}{2}s(k-1) \ln(\sigma_w^2) - \frac{1}{2}s \ln \gamma - \frac{SCw}{2\sigma_w^2} - \frac{SCs}{2\gamma} - \frac{sk(\bar{y}_{..} - \mu)^2}{2\gamma}$$

Donde:

$$\gamma = \sigma_w^2 + k_1 \cdot \sigma_s^2$$

Para encontrar el máximo de esta expresión, se deriva en función de los parámetros y se iguala a cero:

$$\begin{aligned} \frac{\partial Ln(L)}{\partial \mu} &= \frac{sk(\bar{y}_{..} - \mu)}{\gamma} = 0 \\ \frac{\partial Ln(L)}{\partial \sigma_w^2} &= \frac{s(k-1)}{2\sigma_w^2} + \frac{SCw}{2\sigma_w^4} = 0 \\ \frac{\partial Ln(L)}{\partial \gamma} &= -\frac{s}{2\gamma} + \frac{SCs}{2\gamma^2} + \frac{sk(\bar{y}_{..} - \mu)^2}{2\gamma^2} = 0 \end{aligned}$$

Este sistema de ecuaciones tiene como solución única:

$$\mu = \bar{y}$$

$$\sigma_w^2 = CMw$$

$$\gamma = \left(1 - \frac{1}{s}\right) CMs$$

Por lo tanto:

$$\sigma_s^2 = \frac{\gamma - CMe}{k} = \frac{\left(1 - \frac{1}{s}\right) CMs - CMe}{k}$$

Los estimadores máximos verosímiles, para el modelo mixto con dos factores sin interacción y data balanceada son:

$$\sigma_s^2 = \frac{\frac{SCs}{s} - \sigma_w^2}{bk}$$

$$\sigma_s^2 = \left[1 - \frac{b-1}{s(bk-1)}\right] CMw$$

Y con efecto de interacción:

$$\sigma_w^2 = CMw$$

$$\sigma_s^2 = \frac{\left(1 - \frac{1}{s}\right)(CMs - CMsb)}{bk}$$

$$\sigma_{sb}^2 = \frac{\left(1 - \frac{1}{s}\right)^* CMsb - CMw}{k}$$

Ejemplo 5.

Se usaron los datos de los ejemplos 2, 3 y 4 para mostrar el cálculo de los estimadores con el método ML.

Cálculo con el ejemplo 2.

Estimaciones máximo verosímiles:

$$\sigma_w^2 = 2333.896$$

$$\sigma_s^2 = \frac{\left(1 - \frac{1}{5}\right) 4299.25 - 2333.9}{8} = 138.2$$

Por lo tanto, la varianza fenotípica es:

$$\sigma_{p(ML)}^2 = 138.2 + 2333.9 = 2472.1$$

Lo que lleva al cálculo del índice de herencia:

$$h_{ML}^2 = \frac{4(138.2)}{2333.9 + 138.2} = 0.22$$

Uso del SAS.

```
data gallinas3;
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=ML;
class x;
model y = ;
random x;
run;
quit;
```


Tabla 13
Método ML caso balanceado

The Mixed Procedure	
Covariance Parameter Estimates	
Cov Parm	Estimate
x	138.20
Residual	2333.90

Cálculo con el ejemplo 3.

Estimaciones máximo verosímiles:

$$\sigma_{\text{residual}}^2 = \frac{\frac{2.6}{6} - 0.082}{10} = 0.035$$

$$\sigma_w^2 = \left[1 - \frac{2-1}{6(2(5)-1)} \right] 0.084 = 0.082$$

Y, el índice de constancia es:

$$R_{(ML)} = \frac{0.035}{0.082 + 0.035} = 0.29$$

Uso del SAS.

```
data pavos2;
input pavo operador y @@;
datalines;
1 1 2.3 2 1 2.2 3 1 2.1 4 1 2.3 5 1 2.1 6 1 2.0
1 2 2.2 2 2 3.1 3 2 2.6 4 2 2.4 5 2 2.0 6 2 1.9
1 1 2.2 2 1 2.7 3 1 2.7 4 1 2.1 5 1 2.2 6 1 2.6
1 2 3.0 2 2 2.2 3 2 2.0 4 2 2.6 5 2 2.0 6 2 2.5
1 1 2.4 2 1 2.5 3 1 1.9 4 1 2.0 5 1 2.0 6 1 2.3
1 2 2.8 2 2 2.6 3 2 1.9 4 2 1.7 5 2 1.8 6 2 2.0
1 1 2.6 2 1 2.9 3 1 1.8 4 1 2.4 5 1 2.0 6 1 2.5
1 2 2.5 2 2 2.3 3 2 2.3 4 2 2.1 5 2 1.7 6 2 2.2
1 1 2.2 2 1 2.2 3 1 2.1 4 1 2.5 5 1 1.8 6 1 2.0
1 2 2.3 2 2 3.0 3 2 2.2 4 2 2.7 5 2 1.7 6 2 1.7
```

```

;
proc mixed method=ML;
  class pavo operador;
  model y=operador ;
  random pavo;
run;

```

Tabla 14
Método ML, caso balanceado modelo mixto

The Mixed Procedure	
Covariance Parameter Estimates	
Cov Parm	Estimate
pavo	0.03500
Residual	0.08269

Cálculo con el ejemplo 4.

Estimaciones máximo verosímiles:

$$\sigma_w^2 = 1.98333$$

$$\sigma_{animal}^2 = \frac{\left(1 - \frac{1}{3}\right)(21.23335 - 18.6333)}{\left(1 - \frac{1}{3}\right)^{10} 18.6333 - 1.9833} = 0.17333$$

$$\sigma_{animal*racion}^2 = \frac{\left(1 - \frac{1}{3}\right)^{10} 18.6333 - 1.9833}{5} = 2.08$$

Uso del SAS.

```

proc mixed method=ML;
  class racion cerdo;
  model y = racion;
  random cerdo cerdo*racion;
run;

```

Tabla 15
Método ML, modelo mixto con interacción balanceado

Covariance Parameter Estimates	
Cov Parm	Estimate
cerdo	0,1733
racion*cerdo	2,0878
Residual	1,9833

CAPÍTULO 7

MODELO MIXTO PARA N FACTORES

ALEATORIOS Y FIJOS Y DATA DESBALANCEADA

Para el caso de n factores aleatorios y fijos, es necesario estudiar el modelo lineal mixto en su forma matricial, el cual puede expresarse como:

$$y = Xb + Zs + e$$

Dónde:

b = vector de efectos fijos asociados con registros en y por X .

S = vector de efectos aleatorios asociados con registros en y por Z .

X = Matriz de incidencia que relaciona observaciones con efectos fijos

Z = Matriz de incidencia que relaciona observaciones con efectos aleatorios

y = vector de observaciones

e = vector de residuales

Hasta este punto hemos trabajado con soluciones cerradas, que nos brinda la solución de un sistema de ecuaciones lineales ya sea mínimo cuadráticas o máximo verosímiles, por lo tanto, hemos estimado parámetros genéticos simplemente resolviendo sistema de ecuaciones lineales. Los métodos de álgebra lineal para resolver los sistemas de ecuaciones lineales expuestos en estas páginas fueron el

método de reducción, sustitución y el de la matriz inversa, sin embargo, cualquier otro método para resolver el sistema de ecuaciones puede ser usado y conllevará a resultados idénticos, debido a que se puede (por así decirlo) despejar los parámetros.

Para cualquier modelo con data desbalanceada, las soluciones máximo verosímiles para los parámetros no son de forma cerrada, es decir, que las ecuaciones son no lineales, y, por lo tanto, se debe abordar la solución de ecuaciones por métodos iterativos. La no linealidad del sistema de ecuaciones ML para datos no balanceados, puede ser desalentadora, debido a que se pueden encontrar soluciones diferentes en función del método que se use para resolver las ecuaciones no lineales.

Hoy en día, existen muchos métodos para maximizar la función de verosimilitud (o verosimilitud restringida) y una gran gama de paquetes estadísticos que los implementan, por ejemplo, el SAS usa el método de Newton-Rapson para resolver el problema no lineal para el método ML. La descripción general de estos métodos sale del alcance de estas páginas, sin embargo, una excelente descripción puede encontrarse en Searle et al. (Searle et al., 1992).

Las ecuaciones para modelos mixtos y datos no balanceados de ML pueden encontrarse en Searle et al. (Searle et al., 1992). En la práctica, en la mayoría de los casos serán no balanceados, y por lo tanto, si se desea emplear ML se necesitara aplicar iteración numérica.

Ecuaciones máximo verosímiles para los efectos fijos.

$$(X'V^{-1}X)b = (X'V^{-1}y)$$

Ecuaciones máximo verosímiles para los efectos aleatorios.

$$Trace(V^{-1}Z_I Z_I') = y' P Z_I Z_I' P y$$

Donde:

V^{-1} = Inversa de la matriz de covarianzas fenotípica

$$P = V^{-1} - V^{-1} X (X' V^{-1} X)^{-1} X' V^{-1}$$

$(X' V^{-1} X)^{-1}$ = A la inversa generalizada del matriz producto de mínimos cuadrados generalizados.

Como se mencionó anteriormente la ecuación $Trace(V^{-1} Z_i Z_i') = y' P Z_i Z_i' E y$ no tiene solución cerrada, y debe resolverse con métodos numéricos iterativos.

Ejemplo 6.

En la siguiente data se tienen pesos al destete en vacunos doble propósito. Como los animales se encontraban en diferentes fincas y eran de diferentes sexos, estos factores fueron considerados como fijos. Los datos son:

Tabla 16
Datos de pesos al destete en vacunos

Padre	Finca	Sexo	y
1	1	Macho	X
1	1	Hembra	175
1	1	Macho	200
1	2	Macho	205
2	2	Hembra	180
2	2	Hembra	185
3	2	Macho	210
3	1	Macho	208

Uso del SAS

```
data mixto1;
input animal padre madre finca sexo$ y;
missing x;
datalines;
```

```

1 1 1 1 macho X
2 1 2 1 hembra 175
3 1 1 1 macho 200
4 1 1 2 macho 205
5 2 1 2 hembra 180
6 2 2 2 hembra 185
7 3 3 2 macho 210
8 3 3 1 macho 208
;
proc mixed method=ML mmeq;
class padre madre finca sexo;
model y = finca sexo/ solution noint;
random padre;
run;
quit;

```

Y la salida es:

Tabla 17
Método ML para modelo mixto no balanceado

Covariance Parameter Estimates	
Cov Parm	Estimate
padre	5.5211
Residual	3.6897

Y, la prueba para probar la hipótesis para los efectos fijos es:

Tabla 18
Prueba de hipótesis para efectos fijos

Type 3 Tests of fixed Effects				
Effect	Núm. DF	Den DF	F Value	Pr>F
sexo	1	2	165.41	0.006
finca	1	2	4.95	0.156

De las cuales existe un efecto significativo para sexo ($F = 165.41$ con un $P = 0.006$), por lo tanto, convendría estudiar las medias corregidas por mínimos cuadrados (LSMEANS). El nuevo código SAS es:

```
proc mixed method=ML mmeq;
class padre madre finca sexo;
model y = sexo finca/ solution noint;
random padre;
lsmeans sexo;
run;
quit;
```

Y, la salida es:

Tabla 19
LSMEANS para sexo

Least Squares Means						
Effect	sexo	Estimate	Estándar Error	DF	t Value	Pr> t
sexo	hembra	179.96	1.9043	2	94.5	0,0001
sexo	macho	205.99	1.8067	2	114.01	<.0001

Por lo tanto, en promedio, los machos son superiores a las hembras.

CAPÍTULO 8

MODELOS ALEATORIOS CON EFECTOS ANIDADOS

En algunas ocasiones, interesa introducir el efecto de madres en el modelo aditivo lineal general. Gracias a la inseminación artificial, es posible obtener crías de varias madres, inseminadas de un mismo padre, y podemos repetir este proceso n veces, cuando esto ocurre el efecto de madre está dentro de padre, es decir, que las madres que sirve el semental 1 no son las mismas que sirve el semental 2, por lo tanto, los efectos están anidados.

El diseño anidado es un tipo de diseño experimental donde los factores pueden ser aleatorios o mixtos. Se trata de un diseño factorial donde solo algunos niveles de un factor estarán dentro (anidados) de otros, por lo tanto, no todos los niveles de los factores quedaran juntos; por ejemplo, para un modelo con 2 factores, el modelo aditivo lineal es:

$$Y = s + D(s) + e$$

Donde:

$D(s)$ = efecto de madre dentro de padres

Los componentes de la varianza, asumiendo solo el efecto de padre y madre dentro de padre, vienen de la solución del siguiente sistema de ecuaciones lineales:

$$\sigma_w^2 + k_2 \sigma_d^2 + k_3 \sigma_s^2 = CMs$$

$$\sigma_w^2 + k_1 \sigma_d^2 = CMd$$

$$\sigma_w^2 = CMw$$

La solución única de este sistema es:

$$\sigma_w^2 = CMw$$

$$\sigma_d^2 = \frac{CMd - CMw}{k_1}$$

$$\sigma_s^2 = \frac{CMs - CMd}{k_3}$$

Y, la heredabilidad para este modelo es:

$$h_s^2 = \frac{4\sigma_s^2}{\sigma_s^2 + \sigma_w^2 + \sigma_d^2}$$

Las estimaciones de máxima verosimilitud para este modelo anidado, y data balanceada, vienen dadas por:

$$\sigma_w^2 = CMw$$

$$\sigma_s^2 = \frac{\left(1 - \frac{1}{s}\right) CMs - CMd}{dk}$$

$$\sigma_d^2 = \frac{CMd - CMw}{k_1}$$

Ejemplo 7.

Se toman los pesos corporales (gramos) a las 8 semanas, de una bandada de gallinas White rock. Se escogieron al azar 5 padres y 15 madres para los cruces y los pesos de sus progenies fueron analizados. Se usó un diseño anidado, y se aplicó el método de ANOVA.

Tabla 20
Datos de padres y madres de gallinas White rock

	Madres	Progenie
Padre 1	1	965 813 765
	2	803 640 714
	3	644 753 705
Padre 2	4	740 798 941
	5	701 847 909 909 800
	6	853
Padre 3	7	696 807 800
	8	752 863 739
	9	686 832 796
Padre 4	10	979 798 788
	11	905 880 770
	12	797 721 765
Padre 5	13	809 756 775
	14	887 935 937
	15	872 811 925

$$SCs = \frac{6802^2 + 7498^2 + 7403^2 + 6971^2 + 7707^2}{9} - FC = 29476034 - 29412825 = 63209$$

$$FC = \frac{36381^2}{45}$$

$$SCd = \frac{2543^2 + 2157^2 + \dots + 2608^2}{3} = 29564147 - 29476034 = 88113$$

$$SCw = 29729879 - 29564147 = 165732$$

Los cuadrados medios son:

$$CMs = \frac{63209}{4} = 15802.25$$

$$CMd = \frac{88113}{10} = 8811.3$$

$$CMw = \frac{165732}{30} = 5524.4$$

El cuadrado del ANOVA es:

Tabla 21
Cuadrado del Anova

Fuente de variación	Grados de libertad	Sumas de cuadrados	Cuadrados medios	Cuadrado medio esperado
Entre gallos	4	63209	15802	$\sigma_w^2 + 3\sigma_t^2 + 9\sigma_s^2$
Entre gallinas dentro de gallos	10	88113	8811	$\sigma_w^2 + 3\sigma_t^2$
Entre hermanos completos	30	165732	5524	σ_w^2

A partir de estos datos, se estiman los componentes de varianza:

$$\sigma_s^2 = \frac{15802 - 8811}{9} = 776$$

$$\sigma_d^2 = \frac{8811 - 5524}{3} = 1095$$

$$\sigma_w^2 = 5524$$

Por lo tanto, la heredabilidad es:

$$h^2 = \frac{4(776)}{776 + 1095 + 5524} = 0.42$$

Uso del SAS.

```
data anidado;
input padre madre y;
cards;

;
proc nested;
class padre madre;
var y;
run;
```

Tabla 22
Método ANOVA modelo anidado balanceado

Nested Random Effects Analysis of Variance for variable y								
Variance Source	DF	Sum of Squares	F Value	Pr>F	Error Term	Mean Square	Variance Component	Percent of Total
Total	44	317053				7205.754545	7396.783951	100.0000
padre	4	63208	1.79	0.2068	madre	15802	776.754321	10.5012
madre	10	88113	1.59	0.1561	Error	8811.288889	1095.629630	14.8122
Error	30	165732				5524.400000	5524.400000	74.6865

Y, las estimaciones de máxima verosimilitud son:

$$\sigma_w^2 = 5524$$

$$\sigma_d^2 = \frac{8811 - 5524}{3} = 1095$$

$$\sigma_s^2 = \frac{\left(1 - \frac{1}{5}\right)15802 - 8811}{9} = 425.6$$

Uso del SAS.

```

proc mixed method=ml;
class padre madre;
model y= ;
random padre madre(padre);
run;

```

Tabla 23
Método ML modelo anidado balanceado.

The Mixed Procedure	
Covariance Parameter Estimates	
Cov Parm	Estimate
padre	425.60
madre (padre)	1095.63
Residual	5524.40

CAPÍTULO 9

EL MODELO LINEAL GENERAL, MIXTO Y SUS ESTIMACIONES MÁXIMO VEROSÍMILES RESTRINGIDAS

Puesto que el estimador ML está 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:

$$\sigma_p^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n - 1}$$

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

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

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:

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

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:

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

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:

$$\sigma_{p(REML)}^2 = 246 + 2334 = 2580$$

$$\sigma_{p(ML)}^2 = 138 + 2334 = 2472$$

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).

CAPÍTULO 10

CRITERIOS Y MÉTODOS DE SELECCIÓN

La herramienta del mejorador genético depende del objetivo de sistema de apareamiento, por ejemplo si queremos criar animales puros y el índice de herencia es alto para la característica seleccionada la respuesta a la selección será eficiente, estos sistemas de apareamientos son simples ya que trabajamos con razas puras, buscamos cruzar los mejores padres con las mejores madres de la población con el fin de tener progenies con altos rendimientos en las características seleccionadas, el problema está en saber cuáles son los mejores animales de la población.

La selección no es más que un proceso de reproducción diferencial donde se le da la oportunidad a los mejores genotipos a contribuir con sus gametos para formar la siguiente generación. Si los mejores animales son escogidos y apareados el promedio general de la siguiente generación aumentaría, eso lo medimos como la respuesta a la selección:

$$R = h^2 S$$

Dónde:

S =diferencial de selección

R =respuesta a la selección

Y sabiendo que:

$$S = \overline{y_s} - \overline{y_\mu}$$

Donde estos promedios son los del grupo seleccionado y el del rebaño.

Si buscamos maximizar la respuesta a la selección lo que se busca es aumentar el diferencial de selección, esto se logra aumentando el número de animales descartados.

Criterios de selección: Son las piezas de información que usa el mejorador genético:

- Pedigree
- Criterio individual
- Prueba de progenie
- Parientes colaterales

El criterio de selección usado depende de la actitud de la especie, tomemos como ejemplo la selección de toros en razas lecheras como la Holstein, el criterio individual quedaría completamente descartado debido a que los toros no dan leche, pero estos son transmisores de genes que contribuyen a la producción láctea, por lo tanto, un criterio por prueba de progenie o parientes colaterales (o ambos) sería apropiados.

Cuando se plantea aplicar la selección para n característica simultaneas, en un rebaño particular, es necesario conocer cuales animales (machos y hembras) tienen la capacidad de generar crías con las características deseadas. Esto se puede conocer calculando la

habilidad de transmisión predicha (PTA, para producción de leche) o la diferencia esperada de progenies (DEPs, para producción de carne) para los animales de un rebaño particular. El PTA es el mejor predictor del mérito genético de un animal, y se interpreta como una desviación de la media poblacional, es decir, representa la superioridad o inferioridad de un animal con respecto al promedio del rebaño.

En los índices de selección, los valores fenotípicos tienen que estar perfectamente ajustados por los factores ambientales que afectan al carácter (efectos fijos), de lo contrario, los estimadores podrían estar sesgados y ser menos confiables. Para la estimación de los PTA, es necesario conocer los índices de herencia, las correlaciones genéticas (si aplica) y el valor económico (si aplica) del rasgo a seleccionar. Con los índices de selección, se produce un valor óptimo para cada animal y es un valor único que está optimizado por los valores económicos, el índice de herencia y las correlaciones entre las características.

Por la teoría estadística, se conoce que el mejor predictor lineal del mérito genético de un animal es el producto del coeficiente de regresión, por el registro ajustado:

$$I_i = b_1 (y_i - \mu)$$

Donde:

I_i = Índice de selección para un rasgo particular

b_1 = Coeficiente de regresión lineal.

y_i = Valor fenotípico.

μ = Promedio del rebaño.

Si se cuenta con la información de más de un familiar:

$$I_i = \sum b_i (y_i - \mu)$$

Como se puede notar, el coeficiente de regresión, se asume como un valor poblacional, pero en realidad es un estimador del valor

real, por lo tanto, para utilizar un índice de selección los parámetros genéticos estimados, deben ser calculados a partir de muestras suficientemente grandes, para que los errores de muestreo no puedan distorsionar la eficiencia de las predicciones.

Cuando se utiliza un índice de selección para medidas repetidas de un animal (como por ejemplo producción láctea), el mejor predictor lineal del valor genético es:

$$I_i = \frac{nh^2}{1 + (n-1)R} (\bar{y}_i - \bar{y}_R) = b (\bar{y}_i - \bar{y}_R)$$

Donde:

R = Es el índice de constancia del carácter.

Para las pruebas de progenie, se utiliza una ecuación similar, pero tiene leves modificaciones, la principal es que el índice de constancia es remplazado por la heredabilidad. Para este caso el mejor predictor del mérito genético para un reproductor es:

$$I_i = \frac{\frac{1}{2}nh^2}{1 + \frac{1}{4}(n-1)h^2} (\bar{y}_i - \bar{y}_R) = b (\bar{y}_i - \bar{y}_R)$$

Donde:

h^2 = Índice de herencia de la característica en cuestión

Cuando se aplica la selección para n rasgos, el índice de selección toma en cuenta el valor de los parámetros genéticos y la importancia económica de cada rasgo (el valor económico es opcional). Este índice genera un sistema de ecuaciones que puede ser resuelto por métodos algebraicos. El vector de parámetros de regresión, se estiman resolviendo un sistema de ecuaciones que se construye de tal forma que en las filas y columnas de la matriz donde se tengan sub índices iguales se colocan varianzas, y donde sean desiguales se colocan

covarianzas. En esta estructura del lado izquierdo corresponden varianzas y covarianzas fenotípicas ponderadas por parámetros de regresión, y del lado derecho varianzas y covarianzas genéticas ponderados por valores económicos. La expresión algebraica es:

$$Pb = Ga$$

Donde:

P = matriz de covarianzas fenotípicas

b = vector de parámetros de regresión

G = matriz de covarianzas genéticas

a = vector de valores económicos

La solución de esta expresión es:

$$b = P^{-1} Ga$$

Donde:

P^{-1} = Inversa de la matriz fenotípica.

La expresión $b = P^{-1} Ga$, es equivalente (en este caso con 2 características) a resolver el siguiente sistema de ecuaciones:

$$\sigma_{py1}^2 b1 + COV_{p(y1y2)} b2 = \sigma_{ay1}^2 a1 + COV_{a(y1y2)} a2$$

$$COV_{p(y1y2)} b1 + \sigma_{py2}^2 b2 = COV_{a(y1y2)} a1 + \sigma_{ay2}^2 a2$$

Ejemplo 9

Se quiere evaluar a dos toros usando el criterio individual, con registros de 100 Kg y 200Kg, para cada uno, en un periodo específico. Para los candidatos no se conocen ascendientes de ninguno de ellos y la heredabilidad del carácter es $h^2 = 0.50$.

1.-) Modelo animal (modelo mixto).

Se construyen las ecuaciones del modelo mixto de Henderson usando la multiplicación matricial, para este caso:

Tabla 25
Datos de toros

	Promedio	Toro1	Toro2	RHS
Promedio	2	1	1	300
Toro 1	1	1	0	100
Toro 2	1	0	1	200

De color gris tenemos $\{X'X\}$, de verde $\{X'Z\}$, de Naranja $\{Z'X\}$ de celeste $\{Z'Z\}$ y de amarillo RHS.

Como no se conoce información genealógica la matriz A es igual a la matriz identidad, y, por lo tanto, su inversa:

$$A^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

El parámetro alfa es igual a $1 - \frac{h^2}{h^2}$, por lo tanto:

$$\alpha = \frac{1 - 0.5}{0.5} = 1$$

Y, si multiplicamos por la inversa de la matriz de parentesco, obtenemos para este caso la misma matriz:

$$A^{-1} * 1 = A^{-1}$$

Por lo tanto:

$$(Z'Z) + A^{-1} * \alpha = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

Entonces, las ecuaciones completas son:

$$\begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 0 \\ 1 & 0 & 2 \end{pmatrix} \begin{pmatrix} \mu \\ a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 300 \\ 100 \\ 200 \end{pmatrix}$$

Y la solución viene de:

$$\begin{pmatrix} \mu \\ a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 0 \\ 1 & 0 & 2 \end{pmatrix}^{-1} * \begin{pmatrix} 300 \\ 100 \\ 200 \end{pmatrix}$$

Resolviendo el sistema de ecuaciones tenemos:

$$\begin{pmatrix} \mu \\ a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 150 \\ -25 \\ 25 \end{pmatrix}$$

2.-) Índice de selección.

Para los índices de selección, el mejor predictor del valor genético de un animal, es el coeficiente de regresión lineal multiplicado por el registro ajustado (registro ajustado por los efectos fijos). Para este caso el efecto fijo es la media, por lo tanto, se calcula el valor estimado de la media:

$$\mu = \frac{\sum y}{n}$$

Por lo tanto:

$$\mu = \frac{100 + 200}{2} = 150$$

Notemos que el valor para el efecto fijo media, coincide perfectamente al calculado con el modelo animal. Procedemos a desviar los datos de la media:

$$y^* = 100 - 150 = -50$$

$$y^* = 200 - 150 = 50$$

Ya tenemos los desvíos de la media para los animales 1 y 2, para el criterio individual el valor del coeficiente de regresión es igual a la heredabilidad:

$$a_i = b^*(y_i - \mu) = h^2 * (y_i - \mu)$$

Por lo tanto:

$$a_1 = 0.5 (-50) = -25$$

$$a_2 = 0.5 (50) = 25$$

El cual es un resultado idéntico al encontrado con el modelo animal. Los DEPs son la mitad del valor genético, por lo tanto:

$$DEP1 = -\frac{25}{2} = -12.5kg$$

$$DEP2 = \frac{25}{2} = 12.5kg$$

Por lo tanto, se debe seleccionar al animal 2.

Se nota que con el **Modelo animal** se hizo un trabajo estadístico mayor para obtener los mismos resultados que con un índice de selección, el cual es sencillo de aplicar. Esto lleva a la reflexión, de que, si no disponemos de una buena información genealógica, las ventajas del modelo animal no son tan aparentes con respecto al índice de selección, por el contrario, si se dispone de una buena matriz A, las

estimaciones serán más precisas, y, por lo tanto, es conveniente aplicar el modelo animal.

Ejemplo 10.

Se quiere evaluar vacas por producción láctea y peso al destete, por lo tanto, se decidió estimar sus valores genéticos para estos rasgos usando un índice de selección. Los datos de los parámetros genéticos y fenotípicos son:

Y_1 = Producción de leche

Y_2 = Peso al destete

a_1 = valor económico de producción de leche

a_2 = valor económico del peso al destete

$$h_{y1}^2 = \frac{\sigma_{ay1}^2}{\sigma_{py1}^2}$$

$$h_{y2}^2 = \frac{\sigma_{ay2}^2}{\sigma_{vy2}^2}$$

Covarianza genética de y_1 , y_2

Covarianza fenotípica de y_1 , y_2

Los valores de los parámetros se colocan según el criterio mencionado:

$$100b_1 + 70.70b_2 = (45) \cdot (2) + (31.20) (1)$$

$$70.70b_1 + 200 b_2 = 31.20 (2) + (60) (1)$$

Resolviendo los productos:

$$100b_1 + 70.70b_2 = 90 + 31.20$$

$$70.70b_1 + 200 b_2 = 62.40 + 60$$

Resolviendo la aritmética:

$$100b_1 + 70.70b_2 = 121.20$$

$$70.70b_1 + 200b_2 = 122.40$$

Para encontrar los valores de los coeficientes de regresión se puede aplicar la ecuación de determinantes:

$$\begin{vmatrix} 100 & 70.70 \\ 70.70 & 200 \end{vmatrix} = 20000 - 4998.49 = 15001.51$$

$$X1 = 24240 - 8653.68 = 15586.37$$

Por lo tanto, el coeficiente de regresión para b_1 es:

$$b_1 = \frac{15586.37}{15001.51} = 1.0389$$

Usando el mismo criterio, podemos encontrar el valor para el segundo coeficiente de regresión. Para estos valores el índice de selección es:

$$I = 1.0389 (y_1 - \mu) + 0.2447 (y_2 - \mu)$$

Uso del SAS.

```
proc iml;
reset print;
P= { 100 70.70, 70.70 200};
G={45 31.20, 31.20 60};
a={ 2, 1};
b= inv(P)*G*a;
```

Tabla 26
Salida índice de selección

P	2 rows	2 cols (numeric)
	100	70.7
	70.7	200
G	2 rows	2 cols (numeric)
	45	31.2
	31.2	60
A	2 rows	1 col (numeric)
		2
		1
B	2 rows	1 col (numeric)
		1.0389834
		0.2447194

Resultados idénticos a los encontrados manualmente.

BIBLIOGRAFÍA

- Blasco, A. (2021). *Mejora genética animal*. Editorial síntesis. 142pp.
- Becker, W. (1986). *Manual de Genética Cuantitativa*. Academic Enterprises. 173 pp.
- Castejón, O. (2008). *Diseño y análisis de experimentos con Statistix*. Colección de Textos Universitarios. Ediciones del Vice Rectorado Académico.
- Gutiérrez, P. (2010). *Iniciación a la valoración Genética Animal*. Metodología adaptada al EEES. Editorial Complutense.
- Harvey. (1960) *least-squares analysis of data with Unequal Subclass Numbers Agricultural Research Service*. United states department of agriculture. 157 pp.
- Hartley, H. O. y Rao, J. N. K. (1967). Maximum Likelihood Estimation for the Mixed Analysis of Variance Model. *Biometrika*, 54, 93-108
- Harville. (1997). Matrix Algebra From a Statisticians Perspective. Springer. *Mathematical Sciences Department York-*

town Heights. NY 10598-0218

- Henderson, C.R. (1953). Estimation of variance and covariance components. *Biometrics* 9,226-252.
- Johnson, D. and Thompson, R. (1995). *Restricted Maximum Likelihood Estimation of Variance Components for Univariate Animal Models Using Sparse Matrix Techniques and Average Information*, J. Dairy Sci. 78: 449-456.
- Meyer, K. and Mark K. (2004) *Restricted maximum likelihood estimation of genetic principal components and smoothed covariance matrices*. Animal Genetics and Breeding Unit, University of New England.
- Montgomery (2005). *Diseño y análisis de experimentos. Segunda edición*. Limusa-wiley.
- Verde, O y Yañez, F. (2014). Modelos estadísticos de evaluación genética. *Logros & Desafíos de la Ganadería Doble Propósito*. GIRAZ. 107-119.
- Román, R y Aranguren, A (2014). Evaluación genética de reproductores: logros y desafíos. *Logros & Desafíos de la Ganadería Doble Propósito*. GIRAZ. 137-156.
- Patterson, H. D. y Thompson, R. (1971). Recovery of Inter-Block Information When Block Sizes are Unequal. *Biometrika*, 58, 545-554.
- Román, R y Aranguren, A. (2014). El modelo animal simple: una metodología para los genetistas. *Logros & Desafíos de la Ganadería Doble Propósito*. GIRAZ. 120-136.
- Searle, S. R. (1982). *Matrix Algebra Useful for Statistics*. John Wiley & Sons, New York. 438pp.

Searle, S. R.; Casella, G. and McCulloch, C. E. (1992). *Variance Components*. New York: Wiley.

Steel y Torrie (1985). *Bioestadística: principios y procedimientos*, segunda edición. McGRAW- HILL. 621 pp.

Fondo Editorial Universidad Rafael Urdaneta
Maracaibo- Venezuela
2024

