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:

$$\hat{\mu} = \overline{y}$$
$$\hat{s} = \overline{y}_{i..} - \overline{y}_{..}$$

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

$$\sum\sum\left(y_{ij}-\overline{y}_{..}\right)^{2}=\sum\sum\left[\left(\overline{y}_{i.}-\overline{y}_{..}\right)+\left(y_{ij}-\overline{y}_{i.}\right)\right]^{2}$$

$$=\sum\sum\left[\left(\overline{y}_{i.}-\overline{y}_{..}\right)^{2}+\left(y_{ij}-\overline{y}_{i.}\right)^{2}+2\left(\overline{y}_{i.}-\overline{y}_{..}\right)\left(y_{ij}-\overline{y}_{i.}\right)\right]$$

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

Los productos cruzados son cero, por lo tanto:

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

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

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

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)	$\sum \frac{y_{j.}^2}{k} - \frac{\left(\sum y..\right)^2}{n}$	s-1	$\frac{SC_S}{gl_s}$	$\sigma_w^2 + k_1.\sigma_s^2$
Hijos (w)	$SC_{total} - SC_s$	n-s	$\frac{SC_w}{gl_w}$	σ_w^2
Total	$\sum y_{ij}^2 - \frac{\left(\sum y..\right)^2}{n}$	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):

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

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 pérdidas 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 $b = (X'\mathrm{X})^{-} - (X'\mathrm{y})$ 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:

$$\sum \mathrm{C}_{\mathrm{ij}} \mathrm{Y}_{\mathrm{i}} = \widehat{\mathrm{c}_{\mathrm{i}}}$$

Dónde:

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

$\widehat{\mathrm{c}}_{\mathrm{i}}$ = 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:

$$\mathrm{SC_s} = \mathrm{B}' \, \mathrm{Z}^{-1}\mathrm{B}$$

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:

$$FC = \frac{\left(\sum y..\right)^2}{n}$$

Y, la suma de cuadrados del residual es:

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

Y, R (µ, sí) viene dado por:

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

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

$$\mathrm{SC_s} = SCT - SCe$$

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	$\frac{sc_s}{gl_s}$	$\sigma_w^2 + k_1.\sigma_s^2$
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):

$$CM_w = \sigma_w^2$$

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

La solución de este sistema de ecuaciones es:

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

$$\sigma_w^2 = CMw$$

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:

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

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

$$SCtotal = 687^2 + \ldots + 690^2 - \frac{27331^2}{40} = 98884$$

$$SCW = 98884 - 17197 = 81687\ (hijos)$$

Grados de libertad:

$$GLs = 5 - 1 = 4$$
$$GLtotal = 40 - 1 = 39$$
$$GLw = 40 - 5 = 35$$

Cuadrados medios:

$$CMs = \frac{17197}{4} = 4299.25$$
$$CMw = \frac{81687}{35} = 2333.9$$

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

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

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:

$$P = \begin{bmatrix} 1\,k1 \\ 1\ \ 0 \end{bmatrix}$$

Por lo tanto:

$$\begin{bmatrix} \sigma_w^2 \\ \sigma_s^2 \end{bmatrix} = \begin{bmatrix} 1\,8 \\ 1\,0 \end{bmatrix}^{-1} \begin{pmatrix} 4299.4 \\ 2333.896 \end{pmatrix} = \begin{pmatrix} 2333.896 \\ 245.688 \end{pmatrix}$$

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).