Rev. Fac. Agron. (UCV) 49: 1 - 9. 2023.

La matriz H en el modelo a dos vías de clasificación bajo las restricciones usuales para modelos de rango incompleto

Rafael Román[1*], Omar Verde[2], José Aranguren[1], Benjamín Gómez[3] y Carlos Vargas†

[1]Facultad de Ciencias Veterinarias, Universidad del Zulia. Apdo. 526. Maracaibo 4001, Zulia, Venezuela
[2]Facultad de Ciencias Veterinarias, Universidad Central de Venezuela, Apdo. 4563. Maracay, 2101-A. Venezuela
[3]Tecnologico Nacional de México/IT del valle de Morelia. Morelia – Michoacán, México

RESUMEN

El análisis de datos no ortogonales, debido al desigual número de observaciones en las celdas, es frecuente en experimentos biológicos. Este hecho produce que, en el análisis a dos o más vías de clasificación, se generen efectos confundidos y, en consecuencia, los análisis no pueden ser realizados adecuadamente por procedimientos convencionales. Diferentes sumas de cuadrados se obtienen dependiendo del orden de los factores y la estrategia de estimación. En este trabajo, se obtuvo la matriz $\boldsymbol{H}$ para el análisis a dos vías de clasificación con interacción mediante la pre - multiplicación de la matriz de coeficientes de las ecuaciones normales por una inversa generalizada de esta. Así mismo, se generaron las columnas de dicha matriz colocando cada una de las doce columnas de la matriz $\boldsymbol{X}$, como variables dependientes en el modelo con base en la fórmula de cómputo de $\boldsymbol{H}$ y la solución de las ecuaciones normales para modelos de rango incompleto. Computacionalmente es más laboriosa la obtención de $\boldsymbol{H}$, asumiendo que la suma de las constantes dentro de cada efecto es cero, pero usando software especializado, las matrices pueden ser obtenidas con idénticos resultados dependiendo de la restricción usada. Se discuten los usos de esta matriz en los conceptos de estimabilidad, pruebas de hipótesis y detección del patrón de confusión, así como el cálculo de las sumas de cuadrados.

Palabras clave: Inversa Generalizada, Estimabilidad, Prueba de Hipótesis, Sumas de Cuadrados

The H matrix for the two ways classification model under the usual restrictions for the non full rank model

ABSTRACT

The analysis of non-orthogonal data due to the unequal number of observations in the cells is frequent in biological experiments, this fact produces that in the analysis of two or more classification ways the effects of the factors originate confounding and consequently the analyzes cannot be performed appropriately by conventional procedures. Different sums of squares are obtained depending on the order of the factors and the estimation strategy. In this research, the matrix $\boldsymbol{H}$ for the two-way analysis with interaction was obtained from the pre-multiplication of the coefficient matrix of the normal equations by a generalized

*Autor de correspondencia: Rafael Roman

E-mail: rafael.roman@fcv.luz.edu.ve

Recibido: 10-07-2023 Aceptado: 08-01-2024

inverse of it. Likewise, the columns of that matrix were generated by using each one of the columns of the matrix $\boldsymbol{X}$, as dependent variables; based on the computational formula of $\boldsymbol{H}$ and the solution of the normal equations for non-full rank models. Obtaining $\boldsymbol{H}$ is computationally more laborious, assuming that the sum of the constants within each effect is zero, but using specialized software, the $\boldsymbol{H}$ matrices can be obtained with identical results depending on the restriction applied. The uses of $\boldsymbol{H}$ in the concepts of estimability, hypothesis testing and detection of the confounding pattern as the calculation of the sum of squares are discussed.

Key words: Generalized Inverse, Estimability, Hypothesis Testing, Sums of Squares

INTRODUCCIÓN

Los experimentos con número desigual de observaciones a dos o más vías de clasificación tienen dificultades para los análisis, los cuales no puede realizarse apropiadamente de la forma como se analizan los experimentos ortogonales. El grado de dificultad se incrementa con el grado de desbalance, siendo máximo cuando hay celdas vacías. Algunos libros de texto, tales como: Principios y Procedimientos en Estadística con Énfasis en las Ciencias Biológicas (Steel y Torrie, 1960) cubren el análisis de datos desbalanceados; Montgomery (2001) contempla los Bloques Incompletos Balanceados, el principio de confusión y experimentos factoriales fraccionados; una presentación especializada es dada por Harvey (1960), quien puso a disposición de la comunidad científica un programa para el análisis de modelos lineales (Harvey, 1960; 1982) y Searle (1987) cubre en detalle el tema desde el punto de vista de las medias de las celdas y con el beneficio del algebra matricial; software popular usado en este tipo de análisis ha sido el Procedimiento Lineal General (PROC GLM) del Sistema de Análisis Estadístico SAS (2015) y la última versión del programa computarizado de Modelos Mixtos por Cuadrados Mínimos y Máxima Verosimilitud, versión 2 para computadores personales (LSMLMW) (Harvey, 1990).

El problema fundamental con datos desbalanceados es que cuando hay dos o más factores en el análisis, los efectos se confunden. Al realizarse los contrastes no desaparecen algunos efectos, tal como lo ilustran Littell *et al.* (2010). De esta forma, las hipótesis del tipo $\boldsymbol{H_0}$: $\boldsymbol{L' b^0 = m}$, puede que no esté probando lo que el investigador pretende, lo que evidentemente trae como consecuencia el incremento de los errores tipo I y tipo II.

Lo que es estimable puede ser determinado por el producto $\boldsymbol{l' H}$ para cualquier vector $\tilde{\boldsymbol{b}}^{\boldsymbol{0}}$. . Si la función lineal $\boldsymbol{l' H = l'}$ ' la función es estimable (Searle, 1966; 1971); $\boldsymbol{l'}$ es un vector con los coeficientes de una función lineal y $\boldsymbol{H}$ es el producto de una inversa generalizada de la matriz de coeficientes del conocido Sistema de Ecuaciones Normales por ella misma. El otro aspecto importante es que solo las funciones lineales estimables son comprobables estadísticamente en pruebas de hipótesis, dado que al ser estimables son invariantes ante la forma particular de la inversa generalizada usada (Searle, 1982). Por otro lado, cuando una función no es estimable, no se pueden probar hipótesis basados en ese tipo de contrastes, debido al hecho de que las sumas de cuadrados del error en el modelo completo y la suma de cuadrados bajo la hipótesis son idénticos y en consecuencia no es posible realizar la prueba de hipótesis, Searle (1971).

Usando la rutina para la Hipótesis Lineal General (BMD05V) del Programa Computarizado Biomedical (BMD), Johnson (1976) obtiene en forma muy laboriosa la matriz $\boldsymbol{H}$, e ilustra algunos de sus usos en estimabilidad y comprobabilidad estadística y describe el patrón de confusión para una ½ fracción de un experimento factorial 2^3.

Las matrices de asignación encontradas en experimentos son de rango incompleto, y por lo tanto, hacen que el Sistema de Ecuaciones Normales tengan infinitas soluciones, razón por la cual se recurren a las restricciones usuales para obtener una solución y completar el análisis. Con el uso de software como el LSMLMW de Harvey (1990), o el PROC GLM del SAS (2015) es posible obtener $\boldsymbol{H}$ directamente si se incluyen como variables dependientes las columnas de la matriz $\boldsymbol{X}$. En ambos casos la solución particular de cada análisis constituye una columna de $\boldsymbol{H}$.

En este trabajo se obtiene la matriz $\boldsymbol{H}$ bajo las restricciones usuales usadas con los programas disponibles: el PROC GLM y el LSMLWW y se demuestran algunas de sus aplicaciones en el análisis de experimentos desbalanceados.

MATERIALES Y MÉTODOS

En los anexos I y II, se presentan: la ecuación matricial del modelo, y las ecuaciones normales respectivamente, del ejemplo numérico presentado por Harvey (1960), posteriormente utilizado por Littell *et al.* (2010) y Román-Bravo *et al.* (2011) trabajando con datos desbalanceados para propósitos ilustrativos. La matriz $\boldsymbol{H}$ es el resultado de la premultiplicación de la matriz de coeficientes de las ecuaciones normales ($\boldsymbol{X'X}$), (usualmente referido como miembro izquierdo de las ecuaciones (MIE)) por una de las infinitas inversas Generalizadas $(\boldsymbol{X'X})^{-}$, tal como se detalla en [1]:

$$\boldsymbol{H} = \qquad \boldsymbol{H} = (\boldsymbol{X'X})^{-}\boldsymbol{X'X} \qquad [1]$$

El procedimiento para encontrarla consiste en los siguientes pasos: a) obtención de una inversa generalizada de la matriz de coeficientes de las ecuaciones normales y b) realizar el producto manualmente o con software especializado, procedimientos que pueden ser tediosos y sujeto a errores de redondo. Sin embargo, $\boldsymbol{H}$ también puede ser obtenida indirectamente usando software especializado para realizar análisis de varianza como el PROC GLM o el LSMLMW. Con base en la ecuación [1] y la fórmula de estimación de la solución en modelos con matrices de rango incompleto dada en [2], lo cual consiste en la solución particular cuando se pone en $\boldsymbol{0}$ al vector arbitrario que representa el espacio nulo de ($\boldsymbol{X'X}$), en la solución general de la solución dada por Searle (1971):

$$\tilde{\boldsymbol{b}}^{o} = (\boldsymbol{X'X})^{-}\boldsymbol{X'y} \qquad [2]$$

Supóngase que para este problema, particionamos la matriz del diseño $\boldsymbol{X}$ en sus $\boldsymbol{p}$ columnas, digamos, $\boldsymbol{X} = [\boldsymbol{x}_1 : \boldsymbol{x}_2 : \boldsymbol{x}_3 : \boldsymbol{x}_4 : \boldsymbol{x}_5 : \boldsymbol{x}_6 : \boldsymbol{x}_7 : \boldsymbol{x}_8 : \boldsymbol{x}_9 : \boldsymbol{x}_{10} : \boldsymbol{x}_{11} : \boldsymbol{x}_{12}]$. Supóngase ahora que en la solución a las ecuaciones normales dada en [2], en el lugar de vector de observaciones $\boldsymbol{y}$, colocamos cada una de las j-esimas columnas de la matriz $\boldsymbol{X}$, de esta forma el estimador del vector de parámetros esta dado por $\tilde{\boldsymbol{b}}_j^{o} = (\boldsymbol{X'X})^{-}\boldsymbol{X'x}_j$ tendríamos entonces:

$$\boldsymbol{H} = (\boldsymbol{X'X})^{-}\boldsymbol{X'}[\,\boldsymbol{x}_1 \vdots \boldsymbol{x}_2 \vdots \boldsymbol{x}_3 \vdots \boldsymbol{x}_4 \vdots \boldsymbol{x}_5 \vdots \boldsymbol{x}_6 \; \boldsymbol{x}_7 \vdots \boldsymbol{x}_8 \vdots \boldsymbol{x}_9 \vdots \boldsymbol{x}_{10} \vdots \boldsymbol{x}_{11} \vdots \boldsymbol{x}_{12}].$$

Resulta claro bajo estas condiciones que la j^{esima} columna de $\boldsymbol{H}$ es el j^{esimo} vector solución a las ecuaciones normales dado por:

$$\boldsymbol{H} = [\,\tilde{\boldsymbol{b}}_1^{o} \vdots \tilde{\boldsymbol{b}}_2^{o} \vdots \tilde{\boldsymbol{b}}_3^{o} \vdots \tilde{\boldsymbol{b}}_4^{o} \vdots \tilde{\boldsymbol{b}}_5^{o} \vdots \tilde{\boldsymbol{b}}_6^{o} \vdots \; \tilde{\boldsymbol{b}}_7^{o} \vdots \tilde{\boldsymbol{b}}_8^{o} \vdots \tilde{\boldsymbol{b}}_9^{o} \vdots \tilde{\boldsymbol{b}}_{10}^{o} \vdots \tilde{\boldsymbol{b}}_{11}^{o} \vdots \tilde{\boldsymbol{b}}_{12}^{o}] \qquad [13]$$

Se deduce de lo expuesto anteriormente que la matriz $\boldsymbol{H}$ puede ser obtenida facilmante si colocamos como variables dependientes en el estricto orden en el que hemos escrito el modelo (estos vectores los llama Johnson (1976) pseudo - data), las respectivas columnas de $\boldsymbol{X}$ en un análisis con el PROC GLM o el LSMLMW en la sentencia MODEL, obviamente, en ambos casos llegamos a diferentes matrices $\boldsymbol{H}$ dada la forma de las restricciones usadas en ambos programas. Una ventaja en este procedimiento con el PROC GLM o el LSMLMW, sobre la rutina del BMD es que en este último caso el usuario requiere introducir la matriz del diseño de dimensión mínima, en nuestro caso este problema es transparente para el usuario y las columnas de la matriz de diseño puede ser obtenidas con la sentencia IF y ELSE del SAS.

RESULTADOS Y DISCUSIÓN

1. Obtención de la matriz $\boldsymbol{H}$ asumiendo que la última constante dentro de cada efecto es cero.

La primera alternativa natural es asumir que la última constante dentro de cada efecto es cero, en consecuencia, para este problema: $r_2 = s_3 = rs_{13} = rs_{21} = rs_{22} = rs_{23} = 0$ (Searle, 1971; Littell *et al.*, 2010). Bajo esas condiciones, de pueden eliminar, en este problema, de la matriz $\boldsymbol{X}$ las columnas $\boldsymbol{x}_3$, $\boldsymbol{x}_6$, $\boldsymbol{x}_9$, $\boldsymbol{x}_{10}$, $\boldsymbol{x}_{11}$, y $\boldsymbol{x}_{12}$; las columnas restantes forman un subconjunto linealmente independiente (LIN), llamemos a la matriz resultante $\boldsymbol{X_R} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_4, \boldsymbol{x}_5, \boldsymbol{x}_7, \boldsymbol{x}_8]$, el producto $\boldsymbol{X}'_R\boldsymbol{X}_R$ es de rango completo y su inversa regular $(\boldsymbol{X}'_R\boldsymbol{X}_R)^{-1}$ puede ser encontrada por cualquier procedimiento, como por ejemplo operaciones fundamentales sobre hileras.

Para obtener con esa matriz la inversa generalizada que nos da el PROC GLM del SAS (2015), debemos seguir el procedimiento dado por Searle (1982), colocando en los lugares apropiados en la posiciones i, j esos elementos inversos y 0's en las columnas e hileras de la matriz de coeficientes eliminadas, para romper la dependencia lineal, de esta forma la inversa generalizada resultante será de dimensiones (p x p), tal como la matriz de coeficientes, la inversa generalizada obtenida se da en [4], la cual es la que lista el PROC GLM, si se lo requerimos dentro de las opciones del modelo.

La matriz $\boldsymbol{H}$ se presenta en [5] fue obtenida con MATLAB del producto $(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{X}$, un resultado idéntico es obtenido con el PROC GLM como se explicó anteriormente; esta forma de obtener $\boldsymbol{H}$, estaría disponible directamente para su interpretación por parte del investigador, solo requiere imprimir la matriz, dado que el procedimiento usado por Johnson (1976), requiere de mucha manipulación algebraica; desde luego este análisis sirve solo para eso obtener $\boldsymbol{H}$. Resulta importante destacar que el investigador debe estar seguro de la interpretación del efecto de la sentencia CLASSS del PROC GLM, en relación al orden de los parámetros en el vector solución, un ejemplo de la necesidad del conocimiento de estos principios se encuentran por ejemplo al diseñar un experimento de cuadrados latinos replicados donde sean importantes los efectos residuales, o ciertos contrastes que requieran todos los coeficientes del vector, para poder ser estimables por contrastes.

$$(X'_RX_R) = \begin{bmatrix} 18 & 8 & 4 & 8 & 2 & 5 \\ 8 & 8 & 2 & 5 & 2 & 5 \\ 4 & 2 & 4 & 0 & 2 & 0 \\ 8 & 5 & 0 & 8 & 0 & 5 \\ 2 & 2 & 2 & 0 & 2 & 0 \\ 5 & 5 & 0 & 5 & 0 & 5 \end{bmatrix}$$

$$(X'_RX_R)^{-1} = \begin{bmatrix} 0{,}2000 & -0{,}2000 & -0{,}2000 & -0{,}2000 & 0{,}2000 & 0{,}2000 \\ -0{,}2000 & 1{,}2000 & 0{,}2000 & 0{,}2000 & -1{,}2000 & -1{,}2000 \\ -0{,}2000 & 0{,}2000 & 0{,}7000 & 0{,}2000 & -0{,}7000 & -0{,}2000 \\ -0{,}2000 & 0{,}2000 & 0{,}2000 & 0{,}5333 & -0{,}2000 & -0{,}5333 \\ 0{,}2000 & -1{,}2000 & -0{,}7000 & -0{,}2000 & 2{,}2000 & 1{,}2000 \\ 0{,}2000 & -1{,}2000 & -0{,}2000 & -0{,}5333 & 1{,}2000 & 1{,}7333 \end{bmatrix}$$

$$(\boldsymbol{X}'\boldsymbol{X})^{-} = \begin{bmatrix} 0{,}2000 & -0{,}2000 & 0 & -0{,}2000 & -0{,}2000 & 0 & 0{,}2000 & 0{,}2000 & 0 & 0 & 0 & 0 \\ -0{,}2000 & 1{,}2000 & 0 & 0{,}2000 & 0{,}2000 & 0 & -1{,}2000 & -1{,}2000 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -0{,}2000 & 0{,}2000 & 0 & 0{,}7000 & 0{,}2000 & 0 & -0{,}7000 & -0{,}2000 & 0 & 0 & 0 & 0 \\ -0{,}2000 & 0{,}2000 & 0 & 0{,}2000 & 0{,}5333 & 0 & -0{,}2000 & -0{,}5333 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0{,}2000 & -1{,}2000 & 0 & -0{,}7000 & -0{,}2000 & 0 & 2{,}2000 & 1{,}2000 & 0 & 0 & 0 & 0 \\ 0{,}2000 & -1{,}2000 & 0 & -0{,}2000 & -0.5333 & 0 & 1{,}2000 & 1{,}7333 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad [4]$$

$$H = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & -1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad [5]$$

2. Obtención $\boldsymbol{H}$ asumiendo que la suma de las constantes es cero dentro de cada efecto.

La otra alternativa popular para obtener la solución a las ecuaciones normales con matrices de coeficientes de rango incompleto es asumir que la suma de las constantes dentro de cada efecto es cero, es decir, para este problema: $\sum_i r_i = \sum_j s_j = \sum_i rs_{ij} = \sum_j rs_{ij} = 0$ (Harvey, 1960; Searle, (1971); Román-Bravo *et al.*, 2011). Con las columnas de $\boldsymbol{X}$ construyamos una nueva matriz digamos $\boldsymbol{X_A}$ con seis columnas LIN donde: para la media $\boldsymbol{x}_{A.1} = \boldsymbol{x}_1$; para el efecto de ración $\boldsymbol{x}_{A.2} = \boldsymbol{x}_2 - \boldsymbol{x}_3$; para los padres $\boldsymbol{x}_{A.3} = \boldsymbol{x}_4 - \boldsymbol{x}_6$ y $\boldsymbol{x}_{A.4} = \boldsymbol{x}_5 - \boldsymbol{x}_6$; para la interacción $\boldsymbol{x}_{A.5} = \boldsymbol{x}_7 - \boldsymbol{x}_{10} - \boldsymbol{x}_9 + \boldsymbol{x}_{12}$ y $\boldsymbol{x}_{A.6} = \boldsymbol{x}_8 - \boldsymbol{x}_{11} - \boldsymbol{x}_9 + \boldsymbol{x}_{12}$. El producto $\boldsymbol{X'}_A\boldsymbol{X}_A$ es de rango completo y tendrá una inversa, la cual podemos obtener por cualquier método:

$$(\boldsymbol{X'}_A\boldsymbol{X}_A)^{-1} = \left(\frac{1}{540}\right)\begin{bmatrix} 41 & 10 & 4 & -17 & -10 & -16 \\ 10 & 41 & -10 & -16 & 4 & -17 \\ 4 & -10 & 86 & -28 & 10 & 16 \\ -17 & -16 & -28 & 65 & 16 & 4 \\ -10 & 4 & 10 & 16 & 86 & -28 \\ -16 & -17 & 16 & 4 & -28 & 65 \end{bmatrix}$$

Para encontrar la inversa correspondiente a la matriz de coeficientes original, designando la $\boldsymbol{M}$ debemos expandir las columnas e hileras de $(\boldsymbol{X'}_A\boldsymbol{X}_A)^{-1}$, encontrando las columnas e hileras eliminadas de las dependencias lineales que existen entre las columnas de $\boldsymbol{X}$ para ello una forma fácil es definir una matriz con esas relaciones:

$$\boldsymbol{T} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & -1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & -1 & 1 \end{bmatrix}$$

La matriz aumentada se puede obtener del producto $((\boldsymbol{X'}_A\boldsymbol{X}_A)^{-1}\,\boldsymbol{T})'\boldsymbol{T}$. La matriz $\boldsymbol{M}$; tiene la propiedad de ser singular, se puede comprobar fácilmente que: $\boldsymbol{M}(\boldsymbol{X'X})\boldsymbol{M} = \boldsymbol{M}$ y $(\boldsymbol{X'X})\boldsymbol{M}(\boldsymbol{X'X}) = \boldsymbol{X'X}$ en otras palabras, la inversa aumentada es una inversa generalizada de $(\boldsymbol{X'X})$, la correspondiente a asumir que la suma de las constantes es cero; ya que cumple con las condiciones *i* y *ii* dadas por Searle (1982).

$$\boldsymbol{M} = \left(\frac{1}{540}\right)\begin{bmatrix} 41 & 10 & -10 & 4 & -17 & 13 & -10 & -16 & 26 & 10 & 16 & -26 \\ 10 & 41 & -41 & -10 & -16 & 26 & 4 & -17 & 13 & -4 & 17 & -13 \\ -10 & -41 & 41 & 10 & 16 & -26 & -4 & 17 & -13 & 4 & -17 & 13 \\ 4 & -10 & 10 & 86 & -28 & -58 & 10 & 16 & -26 & -10 & -16 & 26 \\ -17 & -16 & 16 & -28 & 65 & -37 & 16 & 4 & -20 & -16 & -4 & 20 \\ 13 & 26 & -26 & -58 & -37 & 95 & -26 & -20 & 46 & 26 & 20 & -46 \\ -10 & 4 & -4 & 10 & 16 & -26 & 86 & -28 & -58 & -86 & 28 & 58 \\ -16 & -17 & 17 & 16 & 4 & -20 & -28 & 65 & -37 & 28 & -65 & 37 \\ 26 & 13 & -13 & -26 & -20 & 46 & -58 & -37 & 95 & 58 & 37 & -95 \\ 10 & -4 & 4 & -10 & -16 & 26 & -86 & 28 & 58 & 86 & -28 & -58 \\ 16 & 17 & -17 & -16 & -4 & 20 & 28 & -65 & 37 & -28 & 65 & -37 \\ -26 & -13 & 13 & 26 & 20 & -46 & 58 & 37 & -95 & -58 & -37 & 95 \end{bmatrix} \quad [6]$$

La matriz $\boldsymbol{H}$ la encontramos del producto $\boldsymbol{M}(\boldsymbol{X'X})$:

$$\boldsymbol{H} = \left(\frac{1}{6}\right)\begin{bmatrix} 6 & 3 & 3 & 2 & 2 & 2 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 3 & -3 & 0 & 0 & 0 & 1 & 1 & 1 & -1 & -1 & -1 \\ 0 & -3 & 3 & 0 & 0 & 0 & -1 & -1 & -1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 4 & -2 & -2 & 2 & -1 & -1 & 2 & -1 & -1 \\ 0 & 0 & 0 & -2 & 4 & -2 & -1 & 2 & -1 & -1 & 2 & -1 \\ 0 & 0 & 0 & -2 & -2 & 4 & -1 & -1 & 2 & -1 & -1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 & -1 & -1 & -2 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 2 & -1 & 1 & -2 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & 2 & 1 & 1 & -2 \\ 0 & 0 & 0 & 0 & 0 & 0 & -2 & 1 & 1 & 2 & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -2 & 1 & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & -2 & -1 & -1 & 2 \end{bmatrix} \quad [7]$$

Las matrices $\boldsymbol{M}$ y $\boldsymbol{H}$ dadas en [6 y 7] fueron obtenidas con MATLAB, sin embargo [7], puede ser obtenida directamente como la solución a un problema con el LSMLMW (Harvey, 1990) como en el caso anterior, colocando como variables dependientes las doce columnas de la matriz $\boldsymbol{X}$, para ello debemos ejecutar el programa PARMCARDS, editar el archivo WORK7.DAT (archivo que contiene las "*tarjetas parámetro*" del problema), en la línea correspondiente a la "*tarjeta parámetro tipo 7*", para codificar las variables dependientes, y evitar que estas sean expresadas como desviaciones de las respectivas medias, colocar en 0's en las posiciones de las medias para cada variable (columnas 11-24 para cada variable dependiente) según apéndice B del manual (Harvey, 1990) y posteriormente ejecutar el LSMLMW. Lógicamente, esta misma solución la podemos obtener por el PROC GLM usando DUMMY variables, con las seis columnas LIN y provocando que el GLM imponga la restricción de que la suma de las constantes es 0, dentro de cada efecto. Alternativamente, la matriz **M**, puede ser obtenida directamente, si usamos seis multiplicadores de Lagrange para imponer las restricciones de que la suma de las constantes es cero dentro de cada efecto. El procedimiento usado por Johnson (1976), requiere de muchos cálculos manuales, la última versión del LSMLMW, puede darnos las doce columnas de $\boldsymbol{H}$ directamente.

$$[l_1 \quad l_2 \quad l_3 \; l_4 \quad l_5 \quad l_6 l_7 \quad l_8 \quad l_9 \; l_{10} \quad l_{11} \quad l_{12}]\begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & -1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} \hat{\mu} \\ \hat{r}_1 \\ \hat{r}_2 \\ \hat{s}_1 \\ \hat{s}_2 \\ \hat{s}_3 \\ (\widehat{rs})_{11} \\ (\widehat{rs})_{12} \\ (\widehat{rs})_{13} \\ (\widehat{rs})_{21} \\ (\widehat{rs})_{22} \\ (\widehat{rs})_{23} \end{bmatrix} \quad [8]$$

$$[l_1 \;\; l_2 \;\; (l_1 - l_2) \;\; l_4 \;\; l_5 \;\; (l_1 - l_4 - l_5) \;\; l_7 \;\; l_8 \;\; (l_2 - l_7 - l_8) \;\; (l_4 - l_7)\, (l_5 - l_8) \;\; (l_1 - l_2 - l_4 - l_5 + l_7 + l_8)] \quad [9]$$

USOS

Para un mismo problema, la matriz $\boldsymbol{H}$ es dependiente de la inversa generalizada por cuanto la matriz de coeficientes es la misma. A continuación, se ilustran algunos de los usos que puede dar el investigador a la matriz $\boldsymbol{H}$.

a) La diferencia entre la forma escalonada reducida $\boldsymbol{X'X}$ y la matriz $\boldsymbol{H}$ obtenida en [5] es que las hileras de 0's no se han desplazado a la porción inferior de $\boldsymbol{H}$ esta matriz nos proporciona información valiosa sobre $\boldsymbol{X'X}$: las seis columnas con un 1 y los elementos restantes 0's son pivotes, y representan las columnas LIN. Las otras seis columnas pueden ser derivadas como combinaciones lineales de las columnas LIN y de hecho cada columna nos indica la forma de crear cada una de ellas, por ejemplo, la columna 3 es una combinación lineal de $\boldsymbol{x_1}$ y $\boldsymbol{x_2}$ de la forma $\boldsymbol{x_3 = x_1 - x_2}$.

b) La forma general de las funciones estimables la podemos obtener del producto $\boldsymbol{l'H\tilde{\beta}}_0$, en [8] se presenta la forma general de las funciones estimables que lista el SAS, tal cual se presentan en el listado y sin los parámetros se da en [9]. Esta puede ser solicitada como una opción en la sentencia LSMEANS usando opción E, también el PROC GLM, nos podría dar un listado con las funciones estimables tipo I, II, III y IV, si así lo requerimos como una opción en la sentencia MODEL con: E1, E2, E3 y E4.

c) En términos generales si $\boldsymbol{l'}_{\boldsymbol{i}}\boldsymbol{H} = \boldsymbol{l'}_i$ entonces la función lineal será estimable; a menudo tenemos algunos contrastes independientes en pruebas de hipótesis típicas del análisis de varianza y los colocamos en una matriz digamos $\boldsymbol{L}$, en este caso serán estimables si y solo si $\boldsymbol{L'H} = \boldsymbol{L'}$. Por otro lado, solo son comprobables hipótesis que sean hechas basados en funciones lineales estimables.

Ejemplos:

1) ¿Es $\bar{y}_{...}$ por si sola estimable?

$$\boldsymbol{l'}_1 = [1 \quad 0 \quad 0\,0 \quad 0 \quad 0\,0 \quad 0 \quad 0\,0 \quad 0 \quad 0]$$

$\boldsymbol{l'}_1\boldsymbol{H} = [1 \quad 0 \quad 1\,0 \quad 0 \quad 1\,0 \quad 0 \quad 0\,0 \quad 0 \quad 1] \neq \boldsymbol{l'_1}$ por lo tanto, no es estimable, lo que nos indica esta información es que la media esta confundida con otros efectos, los cuales podemos determinar multiplicando por el vector de parámetros, dando como resultado $\hat{\mu} + \hat{r}_2 + \hat{s}_3 + (\widehat{rs})_{23}$. En otras palabras, no podremos bajo estas condiciones tener una estimación de μ por si sola.

2) ¿Es estimable la media de la ración 1?

$$\boldsymbol{l'}_2 = [1 \;\; 1 \;\; 0\; 1/3 \;\; 1/3 \;\; 1/3 \;\; 1/3 \;\; 1/3 \;\; 1/3 \;\; 0\; 0\; 0]$$

$\boldsymbol{l'}_2\boldsymbol{H} = [1 \;\; 1 \;\; 0\; 1/3 \;\; 1/3 \;\; 1/3 \;\; 1/3 \;\; 1/3 \;\; 1/3 \;\; 0\; 0\; 0] = l_2$; por lo tanto, es estimable y el mejor estimador lineal insesgado (BLUE) de la media de la ración 1 es $\boldsymbol{l'}_2\widetilde{\boldsymbol{\beta}} = 4{,}3667$

3) ¿Es estimable la diferencia de las medias de las raciones 1 y 2?

$$\boldsymbol{l'}_3 = [0 \;\; 1 \;\; \text{-}1\; 0 \;\; 0 \;\; 0\; 1/3 \;\; 1/3 \;\; 1/3 \;\; \text{-}1/3 \;\; \text{-}1/3 \;\; \text{-}1/3]$$

$\boldsymbol{l'}_3\boldsymbol{H} = \boldsymbol{l'}_3$ por lo tanto, es estimable y su BLUE es $\boldsymbol{l'}_3\widetilde{\boldsymbol{\beta}} = 1{,}0444$

La hipótesis $H_0\!: (\hat{r}_1 - \hat{r}_2) + 1/3\,((\widehat{rs})_{11} + (\widehat{rs})_{12} + (\widehat{rs})_{13}) - 1/3\,((\widehat{rs})_{21} + (\widehat{rs})_{22} + (\widehat{rs})_{23}) = 0$ es por lo tanto comprobable estadísticamente.

$$Q_3 = \boldsymbol{l'}_3\widetilde{\beta}'\,(\boldsymbol{l}_3(\boldsymbol{X'X})^{-1}\boldsymbol{l'}_3)\boldsymbol{l}_3$$

$Q_3 = 3{,}5919$, la suma de cuadrados tipo III pare el efecto de ración

4) Para determinar comprobabilidad estadística en la prueba de hipótesis. Una hipótesis que no es estimable, no es comprobable estadísticamente por las razones expuestas previamente.

5) La forma general de las funciones estimables se podrá obtener por medio del producto $\boldsymbol{l'H}\widetilde{\boldsymbol{\beta}}_0$.

CONCLUSIONES

El cálculo de la matriz $\boldsymbol{H}$ usando software convencional de algebra lineal es más fácil, asumiendo que la última constante dentro de cada factor es cero, con menos posibilidades de errores de redondeo.

Con PROC GLM basta colocar como variables dependientes las columnas de $\boldsymbol{X}$, matriz que es fácil de definir en el SAS, con un grupo de sentencias IF, ELSE.

Con el LSMLMW también se deben introducir las columnas de $\boldsymbol{X}$, o las seis linealemnte independientes y generar las otras de las dependencias lineales entre las columnas de $\boldsymbol{X}$.

La matriz inversa aumentada obtenida asumiendo que la suma de las constantes dentro de cada efecto es cero, es una de las infinitas inversas generalizadas de la matriz de coeficientes de las ecuaciones normales.

Si se definen multiplicadores de Lagrange para imponer la restricción de que la suma de las constantes dentro de cada efecto es cero, se obtiene directamente la inversa aumentada por las columnas e hileras eliminadas.

REFERENCIAS BIBLIOGRÁFICAS

Harvey W. R. 1960. Least-Squares Analysis of Data with Unequal Subclass Numbers. ARS-80. Agricultural Research Service. United States Department of Agriculture, pp. 30–41.

Harvey W. R. 1982. Mixed Model Capabilities of LSML76. Journal of Animal Science, 54(6):1279–1285.

Harvey W. R. 1990. User's Guide for LSMLMW and MIXNDL PC version. Mixed Model Least-Squares and Maximum Likelihood Computer Program. 91 p.

Johnson J.E. 1976. A method of Determining Estimable Functions and Testable Hypotheses in Experimental Design. Monterey, California. Master's Thesis. Naval Postgraduate School. Monterey, California, USA. 39.

Littell R. C; Stroup W.W; Freund R.J. 2010. SAS® for Linear Models, Fourth Edition. SAS Institute Inc., SAS Campus Drive, Cary, North Carolina. 466 p.

Montgomery D. C. 2001. Design and Analysis of Experiments. Fifth Edition. John Wiley and Sons. USA. 684 p.

Román-Bravo R; Verde-Sandoval O; Aranguren-Méndez J; Yánez-Cuellar, L; Ortiz-Rodríguez R. 2011. Forzando al GLM para dar los Resultados Obtenidos con la Restricción $\sum a_i = 0$. XII Encuentro de Investigación Veterinaria y Producción Animal. Universidad Michoacana de San Nicolas de Hidalgo. Facultad de Medicina Veterinaria y Zootecnia. Morelia Michoacán, México. Pp 43-49.

SAS 2015. Institute Inc. SAS/STAT® 14.1 User's Guide. Cary, NC: SAS Institute Inc. SAS/STAT® 14.1

Searle S. R. 1966. Estimable Functions and Testable Hypotheses in Linear Models. BU-213-M. Biometrics Unit, Plant Breeding Department, Cornell University. https://ecommons.cornell.edu/bitstream/handle/1813/32309/BU-213-M.pdf?sequence=1

Searle S. R. 1971. Linear Models. John Wiley & Sons, Inc., Printed in USA. 532 p.

Searle S. R. 1982. Matrix Algebra Useful for Statistics. John Wiley & Sons, Inc., printed in USA. 438 p.

Searle S. R. 1987. Linear Models for Unbalanced Data. John Wiley and Sons. USA. 521p

Steel, R. G; Torrie J. H. 1960. Principles and Procedures of Statistics with Special Reference to the Biological Sciences. Mac Graw Hill Book Company, INC. USA. 252-272.

ANEXO I. Ecuación del Modelo a Dos Vías de Clasificación con Interacción.

$$\begin{bmatrix} 1&1&0&1&0&0&1&0&0&0&0&0\\ 1&1&0&1&0&0&1&0&0&0&0&0\\ 1&1&0&0&1&0&0&1&0&0&0&0\\ 1&1&0&0&1&0&0&1&0&0&0&0\\ 1&1&0&0&1&0&0&1&0&0&0&0\\ 1&1&0&0&1&0&0&1&0&0&0&0\\ 1&1&0&0&1&0&0&1&0&0&0&0\\ 1&1&0&0&0&1&0&0&1&0&0&0\\ 1&0&1&1&0&0&0&0&0&1&0&0\\ 1&0&1&1&0&0&0&0&0&1&0&0\\ 1&0&1&0&1&0&0&0&0&1&0&0\\ 1&0&1&0&1&0&0&0&0&0&0&0\\ 1&0&1&0&1&0&0&0&0&0&0&0\\ 1&0&1&0&0&1&0&0&0&0&0&1\\ 1&0&1&0&0&1&0&0&0&0&0&1\\ 1&0&1&0&0&1&0&0&0&0&0&1\\ 1&0&1&0&0&1&0&0&0&0&0&1\\ 1&0&1&0&0&1&0&0&0&0&0&1 \end{bmatrix} \begin{bmatrix} \hat{\mu}\\ \hat{r}_1\\ \hat{r}_2\\ \hat{s}_1\\ \hat{s}_2\\ \hat{s}_3\\ \left(\widehat{\alpha\beta}\right)_{11}\\ \left(\widehat{\alpha\beta}\right)_{12}\\ \left(\widehat{\alpha\beta}\right)_{13}\\ \left(\widehat{\alpha\beta}\right)_{21}\\ \left(\widehat{\alpha\beta}\right)_{22}\\ \left(\widehat{\alpha\beta}\right)_{23} \end{bmatrix} = \begin{bmatrix} 5\\6\\2\\3\\5\\6\\7\\3\\2\\3\\8\\8\\9\\4\\4\\6\\6\\7 \end{bmatrix}$$

ANEXO II. Ecuaciones Normales para el Modelo a Dos Vías de Clasificación con Interacción

$$\begin{bmatrix} 18&\vdots&8&10&\vdots&4&8&6&\vdots&2&5&1&2&3&5\\ \cdots&\cdots&\cdots&\cdots&\cdots&\cdots&\cdots&\cdots&\cdots&\cdots&\cdots&\cdots&\cdots&\cdots&\cdots\\ 8&\vdots&8&0&\vdots&2&5&1&\vdots&2&5&1&0&0&0\\ 10&\vdots&0&10&\vdots&2&3&5&\vdots&0&0&0&2&3&5\\ \cdots&\vdots&\cdots&\cdots&\cdots&\cdots&\cdots&\cdots&\vdots&\cdots&\cdots&\cdots&\cdots&\cdots&\cdots\\ 4&\vdots&2&2&\vdots&4&0&0&\vdots&2&0&0&2&0&0\\ 8&\vdots&5&3&\vdots&0&8&0&\vdots&0&5&0&0&3&0\\ 6&\vdots&1&5&\vdots&0&0&6&\vdots&0&0&1&0&0&5\\ \cdots&\vdots&\cdots&\cdots&\cdots&\cdots&\cdots&\cdots&\vdots&\cdots&\cdots&\cdots&\cdots&\cdots&\cdots\\ 2&\vdots&2&0&\vdots&2&0&0&\vdots&2&0&0&0&0&0\\ 5&\vdots&5&0&\vdots&0&5&0&\vdots&0&5&0&0&0&0\\ 1&\vdots&1&0&\vdots&0&0&1&\vdots&0&0&1&0&0&0\\ 2&\vdots&0&2&\vdots&2&0&0&\vdots&0&0&0&2&0&0\\ 3&\vdots&0&3&\vdots&0&3&0&\vdots&0&0&0&0&3&0\\ 5&\vdots&0&5&\vdots&0&0&5&\vdots&0&0&0&0&0&5 \end{bmatrix} \begin{bmatrix} \hat{\mu}\\ \cdots\\ \hat{r}_1\\ \hat{r}_2\\ \cdots\\ \hat{s}_1\\ \hat{s}_2\\ \hat{s}_3\\ \cdots\\ \widehat{rs}_{1,1}\\ \widehat{rs}_{1,2}\\ \widehat{rs}_{1,3}\\ \widehat{rs}_{2,1}\\ \widehat{rs}_{2,2}\\ \widehat{rs}_{2,3} \end{bmatrix} = \begin{bmatrix} 94\\ \cdots\\ 37\\ 57\\ \cdots\\ 16\\ 48\\ 30\\ \cdots\\ 11\\ 23\\ 3\\ 5\\ 25\\ 27 \end{bmatrix}$$