Agrociencia

APPLICATION OF LAGRANGE MULTIPLIERS IN TWO-WAY CLASSIFICATION ANALYSIS OF VARIANCE WITH INTERACTION

Rafael María **Román-Bravo**[1,4], Benjamín **Gómez-Ramos***[1,2], Omar Gerardo **Verde-Sandoval**[3], José Atilio **Aranguren-Méndez**[4], Isaías de Jesús **Díaz-Maldonado**[2], Bricia Elena **Corona-Pacheco**[2]

[1]Universidad Michoacana de San Nicolás de Hidalgo, Faculty of Veterinary Medicine and Animal Husbandry, Avenida Acueducto, Morelia, Michoacán, Mexico. Postal Code 58000.
[2]National Technological Institute of Mexico, Valle de Morelia Institute of Technology, Morelia–Salamanca Highway km 6.5, Los Ángeles Neighborhood, Morelia, Michoacán, Mexico. Postal Code 58100**.**
[3]Central University of Venezuela. Faculty of Veterinary Sciences. Extensión 19 de abril Avenue and Casanova Godoy Avenue, 2103, Maracay, Aragua, Venezuela. Postal Code 1050.
[4]University of Zulia. Faculty of Veterinary Sciences. Goajira Avenue, University City, Maracaibo, Zulia, Venezuela. Postal Code 4005.
* Author for correspondence: benjamin.gr@vmorelia.tecnm.mx

Citation: Román-Bravo RM, Gómez-Ramos B, Verde-Sandoval OG, Aranguren-Méndez JA, Díaz-Maldonado IJ, Corona-Pacheco BE. 2025. Application of Lagrange multipliers in two-way classification analysis of variance with interaction. **Agrociencia.** https://doi.org/ 10.47163/agrociencia.v59i7.3482

Editor in Chief:
Dr. Fernando C. Gómez Merino

Received: April 29, 2025.
Approved: October 02, 2025.
Published in Agrociencia: October 28, 2025.

ABSTRACT

The objective of this study was to use Lagrange multipliers to obtain the solution to the normal equations, followed by the sums of squares of the analysis of variance, for a two-way classification model with interaction. The model included the main effects of rations and sires at two and three levels, respectively, as well as their interaction. The least squares method was used in an unbalanced data set with all cells filled. The modified normal equations were formed with the matrix $(\boldsymbol{X'X})_{(12,12)}$ in the upper part, together with the matrix $\boldsymbol{P}_{(12x6)}$ in the lower part, its transpose $\boldsymbol{P}'_{(6x12)}$ appears along with the zero matrix $\mathbf{0}_{(6x6)}$, which together constitutes the left-hand side (LHS) of the modified normal equations. Both the parameter vector and the right-hand side (RHS) were augmented with six θ's and six 0's, respectively. In the inverse of the LHS of the modified equations, within the section corresponding to $(\boldsymbol{X'X})$, the inverse of the coefficient matrix of the incomplete rank matrix was obtained, which resulted in the same property. It was verified, however, that this inverse is the generalized inverse corresponding to the system of equations when the sum of the constants is assumed to be zero for each effect. The direct procedure was used for the decomposition of the models' sum of squares, where both the vector of parameter estimators and the inverse elements of the submatrix for each effect were reduced according to their corresponding degrees of freedom. The results were identical to those previously obtained with conventional procedures, with the advantage that all inverse elements were available for calculating standard errors, reducing the possibility of errors.

Keywords: Confounding effect, orthogonality, least squares, incomplete rank.

INTRODUCTION

Lack of orthogonality in biological experiments can occur from deficiencies in the experimental material that prevent all treatments to be assigned to their respective blocks. This forces the researcher to use designs such as balanced incomplete blocks or partially balanced incomplete blocks. Such imbalance is referred to as planned unbalance (Searle, 1987), in which treatments are compared with equal variances, or treatment groups with similar variances, respectively (Martin, 1995). However, lack of orthogonality often occurs as a result of accidents during the experimental phase, such as deaths of experimental units or measurement errors that require deleting observations from the experiment. From a statistical perspective, these issues lead to severe problems in the computation of sums of squares. This situation is known as unplanned lack of orthogonality and results in consequences such as fewer degrees of freedom than expected for the source of variation or, in extreme cases, a sum of squares equal to zero for that factor in the analysis of variance (Searle, 1987; Martin, 1995).

One of the consequences of imbalance is that, in experiments with two or more factors, confounding effects occur because a correlation is established between the effects. This leads to difficulties when performing hypothesis tests, since linear functions of the parameter estimates are contaminated by fractions of other factors, requiring the assumption that these confounding effects are zero in order for the tested hypothesis to be valid (Steel *et al.,* 1997; Littell *et al.,* 2010).

There are several alternatives to obtain the solution to a system of linear equations. The simplest is to augment the coefficient matrix with the right-hand side of the normal equations and to perform row operations until reaching the reduced row echelon form. However, the inverse elements of the coefficient matrix will still be required as part of the statistical inference process. To solve the normal equations, the most common alternatives are either to set some elements to zero and to use a generalized inverse or to assume that the sum of the constants for each factor is zero (Harvey, 1960; Damond and Harvey, 1987). Given the invariance property of certain quantities typical of the analysis of variance, either of these two procedures is sufficient (Searle, 1971).

The use of Lagrange multipliers is described in detail from a theoretical standpoint and illustrated with a numerical example for the one-way classification analysis (Searle, 1971; 1987). Henderson (1978) the undesirable properties of regressed least squares for predicting animal genetic merit using linear mixed models. However, in this case, to address the linear dependence problem on the left-hand side of the normal equations (LHM), the author simplified the calculations by dropping the mean equation, thereby reducing the problem to a single linear dependence in the system of equations, which is a strategy consistently used by Searle (1971) in solving least squares estimation problems. The literature on the use of Lagrange multipliers in the context of linear models is not very abundant, it was later addressed by Zhu and Li (2007) and more recently in large-scale problems by Scott J, and Tůma M (2022).

Due to the above, the present research has the objective to implement Lagrange multipliers to solve the problem of linear dependence in the LHM of the coefficient matrix of the normal equations in a two-way classification model with interaction. In addition, it aims to obtain all the inverse elements in order to avoid errors that may arise when calculating them from the dependency relationships among the effects in the model.

MATERIALS AND METHODS

For this research, data from an example published by Harvey (1960) are used to analyze a two-way classification experiment with two factors: rations at two levels and sires at three levels, including their interaction, with all cells filled but an unequal number of observations in each (Table 1).

Table 1. Data used to illustrate a two-way classification model with interaction between rations and sires. Adapted from Harvey (1960).

		Sire (*j*)					
Ration (*i*)		1	2	3	$y_{i...}$	$n_{i...}$	
	y_{1jk}	5–6	2-3-5-6-7	3	37	8	
1	$y_{1j.}$	11	23	3			
	$n_{1j.}$	2	5	1			
	y_{2jk}	2–3	8–8-9	4–4–6–6-7	61	10	
2	$y_{2j.}$	5	25	31			
	$n_{2j.}$	2	3	5			
	$y_{.j.}$	16	48	34			
	$n_{.j.}$	4	8	6		$y_{...}$	94
						$n_{...}$	18

i, *j*, *k* = subscript for ration, sire and individual observation identification respectively; y_{1jk} = individual observations for ration 1 and sire j; $y_{1j.}$ = sum of observations in ration 1 for sires 1, 2 and 3; $n_{1j.}$ = number of observations in ration 1 for sires 1, 2 and 3; y_{2jk} = individual observations in ration 2 and sire j; $y_{2j.}$ = sum of observations on ration2 for sires 1, 2 and 3; $n_{2j.}$ = number of observations for sires 1,2 and 3 in ration 2; $y_{.j.}$ = sum of observations for sire *j*; $n_{.j.}$ = number of observations for sire *j*; $y_{i...}$ = sum of observation for ration *i*; $n_{i...}$ = number of observation for sire *i*; $y_{...}$ = grand total and $n_{...}$ = total observations in the data set.

The model, expressed using matrix algebra, can be written as:

$$\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{e} \tag{1}$$

where y is the observation vector, and $\boldsymbol{X}$ is an incidence matrix consisting of zeros and ones, which relates elements of the observation vector to the effects of the factors included in the model. In this case, the twelve columns of are represented as follows: one for the mean, μ; two for the ration effects, α_i, i=1, … ,2; three for the sire effects, β_j, j=1, … 3; and six for the interaction between the them, $(\alpha\beta)_{i,j}$; the number of observations within each cell, varied from k=1, … , $n_{..k}$. The vector $\boldsymbol{b}$ contains the parameters to be estimated by least squares, assuming that the factors included in the model are fixed, with distributional properties as follows: mean, $E(\boldsymbol{y}) = \boldsymbol{Xb}$ and variance $V(\boldsymbol{y}) = \boldsymbol{I}\sigma^2$; finally, $\boldsymbol{e}$ represents the errors with distributional properties $E(\boldsymbol{e}) = \boldsymbol{0}$ and $V(\boldsymbol{e}) = \boldsymbol{I}\sigma^2$, which are assumed to be independent of each other, so that $Cov\ (e_{ijk}, e_{ijk'}) = 0$. The matrix representation of model (1) is shown in terms of the data used in Table 2.

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

Table 2. Equation of the two-way classification model with interaction model.

Searle (1971) proposed implementing restrictions on the model, indicating that the least squares solution leads to solving the next system of equations:

$$\begin{bmatrix} \boldsymbol{X}'\boldsymbol{X} & \boldsymbol{P} \\ \boldsymbol{P}' & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta^o} \\ \boldsymbol{\theta} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}'\boldsymbol{y} \\ \boldsymbol{0} \end{bmatrix} \qquad (2)$$

In an ordinary least squares problem, $\boldsymbol{X'X}$ represents the coefficient matrix for the normal equations. In this case, it is an incomplete rank matrix whose weakness is to be addressed by applying this methodology. $\boldsymbol{P}$ is a matrix required to be of full column rank in order to solve the linear dependence problem present in $\boldsymbol{X'X}$. $\boldsymbol{P}'$ is the transpose of $\boldsymbol{P}$; and $\boldsymbol{0}$ represents a square matrix of zeros, with dimensions sufficient to maintain the symmetry of the modified coefficient matrix used in the methodology. Together, these matrices form the LHM. The parameter vector must be augmented with a column subvector $\boldsymbol{\theta}$ containing as many elements as there are dependencies. On the right-hand side of the normal equations (RHM), the original $\boldsymbol{X'y}$ vector is arranged with the totals and subtotals corresponding to the mean and the remaining factor levels, augmented with as many zeros as the number of dependencies in the problem.

The sums of squares for each source of variation are obtained using the direct procedure by multiplying the transpose of the reduced segment for each factor of the solution vector by the corresponding inverse of the reduced segment of the inverse by the reduced segment of the vector (Harvey, 1960) All calculations were performed using MATLAB version 8.5.0 (R2015a) (The MathWorks Inc., 2015) on an ACER Aspire 3 computer, model N20C5, equipped with 20 GB of random-access memory.

RESULTS AND DISCUSSION

In general terms, for this specific problem, the solution using the least squares method combined with the principle of Lagrange multipliers is straightforward. The procedure involved computing the matrix product $\boldsymbol{X'X}$, which in this case is a 12 × 12 matrix of rank six, and identifying the six dependency relationships in the coefficient matrix. A matrix $\boldsymbol{P}$ is defined to impose restrictions ensuring that the sum of the effects within each factor is zero, with its transpose $\boldsymbol{P'}$ used to maintain symmetry. A 6 × 6 matrix of zeros is included in the modified normal equations, and the parameter vector and RHM are augmented with six θ's and six $0'$s, respectively. The rest of the analysis follows the standard least squares method using matrix algebra. This approach is consistent with the procedures outlined by Searle (1971) and Henderson (1978).

The normal equations and the inverse of the coefficient matrix

The normal equations applying the principle of Lagrange multipliers for this problem according to (2), are expressed as follows:

$$
\left[\begin{array}{cccccccccccc|cccccc}
18 & 8 & 10 & 4 & 8 & 6 & 2 & 5 & 1 & 2 & 3 & 5 & 0 & 0 & 0 & 0 & 0 & 0 \\
8 & 8 & 0 & 2 & 5 & 1 & 2 & 5 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
10 & 0 & 10 & 2 & 3 & 5 & 0 & 0 & 0 & 2 & 3 & 5 & 1 & 0 & 0 & 0 & 0 & 0 \\
4 & 2 & 2 & 4 & 0 & 0 & 2 & 0 & 0 & 2 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
8 & 5 & 3 & 0 & 8 & 0 & 0 & 5 & 0 & 0 & 3 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
6 & 1 & 5 & 0 & 0 & 6 & 0 & 0 & 1 & 0 & 0 & 5 & 0 & 1 & 0 & 0 & 0 & 0 \\
2 & 2 & 0 & 2 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\
5 & 5 & 0 & 0 & 5 & 0 & 0 & 5 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\
1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
2 & 0 & 2 & 2 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\
3 & 0 & 3 & 0 & 3 & 0 & 0 & 0 & 0 & 0 & 3 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\
5 & 0 & 5 & 0 & 0 & 5 & 0 & 0 & 0 & 0 & 0 & 5 & 0 & 0 & 0 & 1 & 0 & 0 \\
\hline
0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{array}\right]
\begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ (\alpha\beta)_{11} \\ (\alpha\beta)_{12} \\ (\alpha\beta)_{13} \\ (\alpha\beta)_{21} \\ (\alpha\beta)_{22} \\ (\alpha\beta)_{23} \\ \cdots \\ \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \\ \theta_5 \\ \theta_6 \end{bmatrix}
=
\begin{bmatrix} 94 \\ 37 \\ 57 \\ 16 \\ 48 \\ 30 \\ 11 \\ 23 \\ 3 \\ 5 \\ 25 \\ 27 \\ \cdots \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}
\tag{3}
$$

The LHM is labeled as the coefficient matrix of the system $\boldsymbol{M}_{18x18}$, represented in a partitioned form according to Searle (1971). The elements of the original incomplete rank matrix are assigned to $\boldsymbol{M}_{(1:12,\ 1:12)}$ as the $(\boldsymbol{X}'\boldsymbol{X})_{12x12}$. The submatrix $\boldsymbol{M}_{(1:12,\ 18:18)}$ corresponds to the full column rank matrix $P_{(12x6)}$, with its transpose $\boldsymbol{P}'_{(6x12)}$ in the submatrix $\boldsymbol{M}_{(18:18,\ 1:12)}$. A zero matrix $\mathbf{0}_{6x6}$ is defined in positions $\boldsymbol{M}_{(13:19,\ 13:18)}$. The parameter vector is augmented with a subvector of six θ_1 restrictions corresponding to the dependencies, and the RHS vector of the normal equations is similarly augmented with six zeros.

Following the matrix of the normal equations in (3) and $\boldsymbol{P}'_{(6x12)}$, the restrictions imposed on the normal equations to bring $\boldsymbol{M}_{18x18}$ to full rank can be deduced. The first $\boldsymbol{\theta}_1' = [0|1\ 1|0]_{(1,12)}$, allows to assume that the sum of both rations is zero, $\alpha_1 + \alpha_2 = 0$. The second, $\boldsymbol{\theta}_2' = [0|1\ 1\ 1|0]_{(1,12)}$, ensures that the sire effects sum to zero. For the interaction, four θ_i restrictions are required: $\boldsymbol{\theta}_3$ sets the sire subcells to sum to zero at the first level of rations; $\boldsymbol{\theta}_4$ does the same at the second level of rations; $\boldsymbol{\theta}_5$ sets the ration subcells to sum to zero at the first sire level; and $\boldsymbol{\theta}_6$ applies the same procedure at the second sire level. By structuring the coefficient matrix in this manner, the matrix $\boldsymbol{M}_{18x18}$ attains linearly independent columns, ensuring the existence of an inverse and a unique solution to the system of equations (Table 3).

−0.0759	0.0185	−0.0185	0.0074	−0.0315	0.0241	−0.0185	−0.0296	0.0481	0.0185	0.0296	−0.0481	−0.5	−0.3333	−0.1667	−0.1667	0.	0.
0.0185	0.0759	−0.0759	−0.0185	−0.0296	0.0481	0.0074	−0.0315	0.0241	−0.0074	0.0315	−0.0241	0.5	0.	−0.1667	0.1667	0.	0.
−0.0185	−0.0759	0.0759	0.0185	0.0296	−0.0481	−0.0074	0.0315	−0.0241	0.0074	−0.0315	0.0241	0.5	0.	0.1667	−0.1667	0.	0.
0.0074	−0.0185	0.0185	0.1593	−0.0519	−0.1074	0.0185	0.0296	−0.0481	−0.0185	−0.0296	0.0481	0.	0.3333	0.1667	0.1667	−0.5	0.
−0.0315	−0.0296	0.0296	−0.0519	0.1204	−0.0685	0.0296	0.0074	−0.0370	−0.0296	−0.0074	0.0370	0.	0.3333	0.1667	0.1667	0.	−0.5
0.0241	0.0481	−0.0481	−0.1074	−0.0685	0.1759	−0.0481	−0.0370	0.0852	0.0481	0.0370	−0.0852	0.	0.3333	−0.3333	−0.3333	0.5	0.5
−0.0185	0.0074	−0.0074	0.0185	0.0296	−0.0481	0.1593	−0.0519	−0.1074	−0.1593	0.0519	0.1074	0.	0.	0.1667	−0.1667	0.5	0.
−0.0296	−0.0315	0.0315	0.0296	0.0074	−0.0370	−0.0519	0.1204	−0.0685	0.0519	−0.1204	0.0685	0.	0.	0.1667	−0.1667	0.	0.5
0.0481	0.0241	−0.0241	−0.0481	−0.0370	0.0852	−0.1074	−0.0685	0.1759	0.1054	0.0685	−0.1759	0.	0.	0.6667	0.3333	−0.5	−0.5
0.0185	−0.0074	0.0074	−0.0181	−0.0296	0.0481	−0.1593	0.0519	0.1074	0.1595	−0.0519	−0.1074	0.	0.	−0.1667	0.1667	0.5	0.
0.0296	0.0315	−0.0315	−0.0296	−0.0074	0.0370	0.0519	−0.1204	0.0685	−0.0519	0.1204	−0.0685	0.	0.	−0.1667	0.1667	0.	0.5
−0.0481	−0.0241	0.0241	0.0481	0.0370	−0.0852	0.1074	0.0685	−0.1759	−0.1074	−0.0685	0.1759	0.	0.	0.3333	0.6667	−0.5	−0.5
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
−0.5	0.5	0.5	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
−0.3333	0.	0.	0.3333	0.3333	0.3333	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
−0.1667	−0.1667	0.1667	0.1667	0.1667	−0.3333	0.1667	0.1667	0.6667	−0.1667	−0.1687	0.3333	0.	0.	0.	0.	0.	0.
−0.1667	0.1667	−0.1667	0.1667	0.1667	−0.3333	−0.1667	−0.1667	0.3333	0.1667	0.1687	0.6667	0.	0.	0.	0.	0.	0.
0.	0.	0.	−0.5000	0.	0.5	0.5	0.	−0.5	0.5	0.	−0.5	0.	0.	0.	0.	0.	0.
0.	0.	0.	0.	−0.5000	0.5	0.	0.5	−0.5	0.	0.5	−0.5	0.	0.	0.	0.	0.	0.

Table 3. Inverse of the coefficient matrix M of the normal equations, augmented according to the Lagrange multipliers methodology and partitioned according to the elements of the original matrix.

The matrix $\boldsymbol{G}_{12x12}$, which contains all the inverse elements corresponding to the incomplete rank matrix $\boldsymbol{X}'\boldsymbol{X}$, has been partitioned by rows and columns according to the model terms for clarity. These partitions correspond to the mean, the rations, the sire, and their interaction effects:

$$\boldsymbol{G} = \left[\begin{array}{rcrrcrrrcrrrrrr}
0.0759 & \vdots & 0.0185 & -0.0185 & \vdots & 0.0074 & -0.0315 & 0.0241 & \vdots & -0.0185 & -0.0296 & 0.0481 & 0.0185 & 0.0296 & -0.0481 \\
\cdots & & \cdots & \cdots & & \cdots & \cdots & \cdots & & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
0.0185 & \vdots & 0.0759 & -0.0759 & \vdots & -0.0185 & -0.0296 & 0.0481 & \vdots & 0.0074 & -0.0315 & 0.0241 & -0.0074 & 0.0315 & -0.0241 \\
-0.0185 & \vdots & -0.0759 & 0.0759 & \vdots & 0.0185 & 0.0296 & -0.0481 & \vdots & -0.0074 & 0.0315 & -0.0241 & 0.0074 & -0.0315 & 0.0241 \\
\cdots & & \cdots & \cdots & & \cdots & \cdots & \cdots & & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
0.0074 & \vdots & -0.0185 & 0.0185 & \vdots & 0.1593 & -0.0519 & -0.1074 & \vdots & 0.0185 & 0.0296 & -0.0481 & -0.0185 & -0.0296 & 0.0481 \\
-0.0315 & \vdots & -0.0296 & 0.0296 & \vdots & -0.0519 & 0.1204 & -0.0685 & \vdots & 0.0296 & 0.0074 & -0.0370 & -0.0296 & -0.0074 & 0.0370 \\
0.0241 & \vdots & 0.0481 & -0.0481 & \vdots & -0.1074 & -0.0685 & 0.1759 & \vdots & -0.0481 & -0.0370 & 0.0852 & 0.0481 & 0.0370 & -0.0852 \\
\cdots & & \cdots & \cdots & & \cdots & \cdots & \cdots & & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
-0.0185 & \vdots & 0.0074 & -0.0074 & \vdots & 0.0185 & 0.0296 & -0.0481 & \vdots & 0.1593 & -0.0519 & -0.1074 & -0.1593 & 0.0519 & 0.1074 \\
-0.0296 & \vdots & -0.0315 & 0.0315 & \vdots & 0.0296 & 0.0074 & -0.0370 & \vdots & -0.0519 & 0.1204 & -0.0685 & 0.0519 & -0.1204 & 0.0685 \\
0.0481 & \vdots & 0.0241 & -0.0241 & \vdots & -0.0481 & -0.0370 & 0.0852 & \vdots & -0.1074 & -0.0685 & 0.1759 & 0.1074 & 0.0685 & -0.1759 \\
0.0185 & \vdots & -0.0074 & 0.0074 & \vdots & -0.0185 & -0.0296 & 0.0481 & \vdots & -0.1593 & 0.0519 & 0.1074 & 0.1593 & -0.0519 & -0.1074 \\
0.0296 & \vdots & 0.0315 & -0.0315 & \vdots & -0.0296 & -0.0074 & 0.0370 & \vdots & 0.0519 & -0.1204 & 0.0685 & -0.0519 & 0.1204 & -0.0685 \\
-0.0481 & \vdots & -0.0241 & 0.0241 & \vdots & 0.0481 & 0.0370 & -0.0852 & \vdots & 0.1074 & 0.0685 & -0.1759 & -0.1074 & -0.0685 & 0.1759
\end{array}\right] \qquad (4)$$

The submatrix $\boldsymbol{G}_{2,3:\,2,3}$ represents the inverse the augmented ration effect; the element $\boldsymbol{G}_{(2,3)}$ is the negative of $\boldsymbol{G}_{(2,2)}$, so that the rows and columns sum to zero, thus satisfying the imposed restriction. The same can be observed for the sire effects in the submatrix $\boldsymbol{G}_{(4,6:\,4,6)}$: the negative of the sum of (0.1593 - 0.0519) is -0.1074, and the negative of the sum of (-0.0519 + 0.1204) is -0.0685, which also occurs by columns within this submatrix. The same can be demonstrated in the submatrix $\boldsymbol{G}_{(7,12:\,7,12)}$, corresponding to the interaction effects. The inverse elements corresponding to this submatrix have the same magnitude as those published by Harvey (1960), with differences in the positions of the elements. This author did not show the values of the inverse elements for the rows and columns eliminated to break the linear dependence in the coefficient matrix, although he illustrated the calculations.

With ($\boldsymbol{X'X}$) being the incomplete-rank coefficient matrix of the original normal equations, it can be verified using linear algebra software that G is a generalized inverse, since it satisfies the properties ($\boldsymbol{X'X}$) $\boldsymbol{G}$ ($\boldsymbol{X'X}$) = ($\boldsymbol{X'X}$) and $\boldsymbol{G}$ ($\boldsymbol{X'X}$) $\boldsymbol{G} = \boldsymbol{G}$. This generalized inverse of $X'X$ corresponds to the generalized inverse when imposed the constrain that the sum of the constants within each effect add up to zero. However, it fails to comply with the symmetry properties required by a Moore-Penrose inverse, as pointed out by Searle (1982), specifically in points (iii) and (iv) corresponding to the products ($\boldsymbol{X'X}$) $\boldsymbol{G}$ and $\boldsymbol{G}$ ($\boldsymbol{X'X}$), respectively.

$$\begin{aligned}\boldsymbol{\beta}^{0\prime} = [&4.8889 - 0.5222\ 0.5222 - 0.8889\ 1.5778 - 0.6889\ 2.0222 \\ &- 1.3444 - 0.6778 - 2.0222 \\ &- 1.3444\ 0.67780 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0]\end{aligned}$$

The solution vector to the normal equations is:

The elements corresponding to the estimators of the model parameters are identical to those previously published, with differences only in their arrangement (Harvey, 1960), as that author presented the results placing the factors in the order of sires, rations, and sires × rations.

Computing the sum of squares for the analysis of variance

If $\boldsymbol{y}$, $\mathbf{1'}_{12}$, $\boldsymbol{\beta'}_0$, and $\boldsymbol{X'y}$ denote the observation vector, the summing vector, the parameter estimator vector, and the RHM, respectively, the sums of squares are obtained in the usual manner. The reduction in the sum of squares due to the mean, often called the correction factor, is: $SC_\mu = [(\mathbf{1'}_{12y})\ (\mathbf{1'}_{12y})]\ /\ (\mathbf{1'}_{12}\mathbf{1'}_{12}) = 490.8889$. The total sum squares corrected for the mean is $SC_{Total\,|\mu} = \boldsymbol{y'y} - SC_\mu = 77.1111$. The reduction in the sum squares due to the full model is $SS_{R(\mu,\,\alpha_i,\,\beta_j,\,(\alpha\beta)_{ij})} = \boldsymbol{\beta'}_0\,\boldsymbol{X'y} = 541.9333$, and finally, the error sum of squares is estimated as $SC_{error} = \boldsymbol{y'y} - \boldsymbol{\beta'}_0\,\boldsymbol{X'y} = 26.0667$.

Decomposition of the sum of squares for the full model

There are several alternatives for obtaining the sums of squares for the main effects and the interaction. In this study, however, the direct procedure described by Harvey (1960) is applied. It should be noted that, in this case, both the matrix $\boldsymbol{G}$ and the submatrices corresponding to the factors included in the model are of incomplete rank, due to the restrictions imposed. To obtain the sums of squares using the direct procedure, each submatrix had to be reduced to full rank by eliminating the linearly dependent rows and columns. The same reduction is applied to the vector of parameter estimators, with the corresponding subvectors adjusted accordingly. The sums of squares for the effects are computed as follows:

Sum of squares for the ration effects

$$SC_{ration} = \hat{\alpha}_1'\,(\boldsymbol{G}_R'\boldsymbol{G}_R)^{-1}\hat{\alpha}_1$$
$$= (-0.5222)(0.0759)^{-1}(-0.5222)$$

$SC_{ration} = 3.5919$

Sum of squares for the sire effects

$$SC_{sire} = [\hat{\beta}_1 \quad \hat{\beta}_2]'\,(\boldsymbol{G}_P'\boldsymbol{G}_P)^{-1}\begin{bmatrix}\hat{\beta}_1\\ \hat{\beta}_2\end{bmatrix}$$
$$= [-0.8889 \quad 1.5778]'\begin{bmatrix}0.1593 & -0.0519\\ -0.0519 & 0.1204\end{bmatrix}^{-1}\begin{bmatrix}-0.8889\\ 1.5778\end{bmatrix}$$

$SC_{sire} = 21.0009$

Sum of squares for the sire× ration interaction effects

$$SC_{ration*sire} = \left[\left(\widehat{\alpha\beta}\right)_{11} \quad \left(\widehat{\alpha\beta}\right)_{12}\right]'\,(\boldsymbol{G}_{rs}'\boldsymbol{G}_{rs})^{-1}\begin{bmatrix}\left(\widehat{\alpha\beta}\right)_{11}\\ \left(\widehat{\alpha\beta}\right)_{12}\end{bmatrix}$$
$$= [2.0222 \quad -1.3444]'\begin{bmatrix}0.1593 & -0.0519\\ -0.0519 & 0.1204\end{bmatrix}^{-1}\begin{bmatrix}2.0222\\ -1.3444\end{bmatrix}$$

$SC_{ration*sire} = 30.2255$

The theoretical difficulties encountered when obtaining sums of squares by differences of submodels, which attempt to obtain Type III sums of squares (Searle, 1972) and are implemented in the Statistical Analysis System (SAS), are resolved by using the direct method in conjunction with the Lagrange multiplier approach. The sums of squares obtained coincide with those reported by Harvey (1960), with discrepancies only due to rounding errors, and correspond to Yates' weighted squares of means (Harvey, 1960; Searle, 1971; Herr, 1986; Steel *et al*., 1997).

With the inverse submatrix $\boldsymbol{G}$ partitioned as in (4), the inverse elements of the reduced submatrices corresponding to each factor can be identified for computing the sums of squares using the direct procedure. For more complex factorial arrays, it is sufficient to define an identity matrix $\boldsymbol{W}$ with the same dimension as $\boldsymbol{G}$ and set to zero the elements of the columns known to be linearly dependent. By computing the matrix product $\boldsymbol{WGW}$, the non-zero elements of the resulting matrix correspond to the required inverse elements, in this case those associated with the mean and the five degrees of freedom of the model. The advantage of this procedure is that all the inverse elements of the matrix $\boldsymbol{X'X}$ are immediately available for calculating the standard errors of the parameter estimators or any linear function of them, thus avoiding computing errors.

CONCLUSIONS

It is feasible and relatively simple to perform analysis of variance in a two-way classification experiment using Lagrange multipliers. The algorithm can be easily implemented in software, reducing the risk of rounding errors. The square segment of the inverse of the coefficient matrix corresponding to the incomplete rank matrix serves as a generalized inverse of that matrix, representing the case in which the imposed restriction is that the sum of the constants within each effect equals zero. This procedure allows all constants and inverse elements required for calculating linear functions to be obtained directly from the model parameter estimators, avoiding semantic and rounding errors associated with sum-and-difference operations between inverse elements in more complex models. It is recommended that the behavior of this procedure be examined in cases involving empty cells.

ACKNOWLEDGMENTS

The authors wish to express their gratitude to the authorities of the Faculty of Veterinary Medicine and Animal Husbandry (FMVZ-UMSNH), the National Institute of Technology of Mexico/Technological Institute of the Valley of Morelia (TecNM-ITVM), the Faculty of Veterinary Sciences of the Central University of Venezuela (UCV), and the Faculty of Veterinary Sciences of the University of Zulia (LUZ) for the institutional support provided to complete this research.

REFERENCES

Damond RA, Harvey WR. 1987. Experimental design, ANOVA and Regression. Harper and Row Publishers: New York, NY, USA. 508 p.

Harvey WR. 1960. Least-squares analysis of data with unequal subclass numbers. United States Department of Agriculture: Washington, DC, USA. 157 p.

Henderson CR. 1978. Undesirable properties of regressed least squares prediction of breeding values. Journal of Dairy Science 61 (1): 114–120. https://doi.org/10.3168/jds.S0022-0302(78)83559-0

Herr DG. 1986. On the history of ANOVA in unbalanced, factorial designs: The first 30 years. The American Statistician 40 (4): 265–270.

Littell RC, Stroup WW, Freund RJ. 2010. SAS for linear models (Fourth edition). SAS Institute Inc.: Cary, NC, USA. 466 p.

Martin FG. 1995. Statistical design and analysis. Notes for STA 6209. Design and analysis of experiments. University of Florida: Gainesville, FL, USA. 316 p.

Scott J, Tůma M. 2022. Solving large linear least squares problems with linear equality constraints. BIT Numerical Mathematics. 62:1765-1782. https://doi.org/10.1007/s10543-022-00930-2

Searle SR. 1971. Linear models. Wiley series in probability and mathematical statistics. John Wiley and Sons: New York, NY, USA. 532 p.

Searle SR. 1972. Using the R() notation for reductions in sum of squares when fitting linear models. Cornell University: Ithaca, NY, USA. 24 p.

Searle SR. 1982. Matrix algebra useful for statistics. Wiley series in probability and mathematical statistics. John Wiley and Sons: New York, NY, USA. 438 p.

Searle SR. 1987. Linear model for unbalanced data. Wiley series in probability and mathematical statistics. John Wiley and Sons: New York, NY, USA. 536 p.

Steel RG, Torrie JH, Dickey DA. 1997. Principles and procedures of statistics: A biometrical approach (Third edition). *In* McGraw-Hill Series in Probability and Statistics. McGraw-Hill: New York, NY, USA, pp: 459–476.

The MathWorks Inc. 2015. MATLAB version: 8.5.0 (R2015a). Natick, MA, USA. https://mathworks.com

Zhu Y, Li XR. 2007. Recursive least squares with linear constraints. Communications in information and system. 7(3): 287-312.