

Model Equivalence in General Linear Models: Set-to-Zero, Sum-to-Zero Restrictions, and Extra Sum of Squares Method

Bashiru I. I. Saeed¹, S. K. Appiah², N. N. N. Nsowah-Nuamah¹, Louis Munyakazi¹ & A. A. I. Musah³

¹ Department of Mathematics and Statistics, Kumasi Polytechnic, Kumasi, Ghana

² Department of Mathematics, College of Science, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana

³ Department of Mathematics and Statistics, Tamale Polytechnic, Tamale, Ghana

Correspondence: Bashiru I. I. Saeed, Department of Mathematics and Statistics, Kumasi Polytechnic, Kumasi, Ghana. E-mail: sirbash156@gmail.com

Received: August 26, 2014 Accepted: September 22, 2014 Online Published: October 10, 2014

doi:10.5539/ijsp.v3n4p42 URL: <http://dx.doi.org/10.5539/ijsp.v3n4p42>

Abstract

The paper is drawn from the authors' experience in teaching general and generalized linear fixed effects models at the university level. The steps followed include model specification, model estimation, and hypothesis testing in general linear model setting. Among these steps, estimation of model parameters such as the main effect least squares means and contrasts were among the most challenging for students. Since no unique solution exists, students are first exposed to the equivalence between two popular techniques that an over-parameterized model can be subjected to in order to obtain the parameter estimates. This is particularly important because existing software do not necessarily follow the same path to produce an Analysis of Variance (or Covariance) of the general, generalized linear fixed or mixed effects models. These steps are generally hidden from the users. It is therefore crucial for the students to understand the intermediary processes that ultimately produce the same results regardless of the software one uses. The equivalent techniques, the set-to-zero and sum-to-zero restrictions, used to obtain solution of the normal equations of the fixed effects model, are presented. The relationship between them is also presented and in the process, data analysis makes use of two important concepts: the generalized inverse and estimable function. The invariance property of estimable functions is also explained in details in addition to the extra sum of squares principle which is introduced to supplement the other concepts. To exemplify these ideas and put them in practice, a simple one-way treatment structure analysis of variance is performed.

Keywords: sum-to-zero restrictions, set-to-zero restrictions, general linear model, over-parameterized model, and extra-sum of squares principle

1. Introduction

The study is drawn from the authors' experience in teaching general and generalized linear fixed effects models at the university/polytechnic levels. The steps followed include model specification, model estimation, and hypothesis testing in general linear model setting. Among these steps, estimation of model parameters was the most challenging area for students. Since no unique solution exists, students are first introduced to the equivalence between two popular techniques namely, over-parameterized model so as to obtain the parameter estimates. This is particularly important because existing software do not necessarily follow the same path to produce an Analysis of Variance (or Covariance) of the general, generalized linear fixed or mixed effects models. These steps are generally latent from the users. It is therefore crucial for learners to understand the intermediary processes that ultimately produce the same results regardless of the software one applies. The two equivalent techniques, namely, the set-to-zero and sum-to-zero restrictions, for obtaining solution of the normal equations of the fixed effects model are outlined as follows:

In matrix notation, the following equation defines linear regression model:

$$y = X\beta + \varepsilon \quad (1)$$

The model is also called the *model equation* of the general linear model (Searle, 1987). The linearity of the model

equation is in terms of the parameters of β . The X is known as *model* matrix (Kempthorne, 1952). In the situation where the data come from an experiment, the matrix X is referred to as the *design* matrix (Searle, 1987) whereas the vector y is of order $N \times 1$. The matrix X which is sometimes called *incidence* matrix is a set of determinist *fixed* effects variables. The β is the unknown vector of parameters and ε are the residuals. The expected value of ε

$$E(\varepsilon) = 0 \quad (2)$$

and its variance is

$$Var(\varepsilon) = \sigma^2 \quad (3)$$

Thus, the expected value of (1) becomes:

$$E(y) = X\beta \quad (4)$$

The following formulation is quite often used for linear model

$$y \sim (X\beta, \sigma^2) \quad (5)$$

The notation in (5) means that the response y is distributed with a mean equal to $X\beta$ and a diagonal variance-covariance matrix of constant values σ^2 . In practice, the use of general linear model is motivated by the desire to determine of how and to what degree *variation in the dependent* is related to the relevant fixed effects (the random effects are not treated in this study). These effects are reflected in the right hand side of (1) and more specifically, in the set up of the X matrix implied by the treatment structure.

2. Estimation Methods

The *normal equations* and their solution require the use of the least squares method in estimating β . The least squares method minimize the sum of squares of the residuals,

$$(y - X\beta)'(y - X\beta) \quad (6)$$

leading to the normal equation:

$$X'X\beta^0 = X'y \quad (7)$$

The solution vector β^0 minimizes (6) and since the X matrix is *rank deficient*, $(X'X)^{-1}$ does not exist. This is always true for *over-parameterized* or over-determined models. There are many least squares estimates β^0 satisfying Equation (7). For non-singularity, the estimation procedure uses a pseudo-inverse or *generalized inverse* $(X'X)^{-}$, which satisfies (8):

$$X'X(X'X)^-X'X = X'X \quad (8)$$

The solution $\widehat{\beta}^0$ which is an estimator of β is given by

$$\widehat{\beta}^0 = (X'X)^-X'y \quad (9)$$

This solution vector is not unique. There exist a large number of possible solutions to the *same* normal equations. To obtain $\widehat{\beta}^0$, the concepts of *sum-to-zero* and *set-to-zero* restrictions is introduced (see Searle, 1987; Milliken & Johnson, 1994).

3. Illustrative Example

It is relatively easier to explain the concepts of *set-to-zero*, *sum-to-zero* restrictions, and *extra sum-of squares* by using matrix notation of a simple one-way treatment example. The following example is taken from McLean (1989):

A researcher is interested in determining whether or not there is any effect due to *different rations* (α_i) in the *gain weight* (y_{ij}) of a certain breed of animal over a period of six weeks. Animals were selected at random and placed in each treatment group. The model to describe this experiment is (1) or alternatively

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij} \quad (10)$$

The following set of data is obtained from the experiment:

Table 1. Effect of ration on weight gain ration

1	2	3
7	3	6
9	4	8

The design matrix X contains the intercept α_0 , and the individual identification of treatment effects α_i .

$$X = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \text{ and } y = \begin{bmatrix} 7 \\ 9 \\ 8 \\ 3 \\ 6 \\ 4 \end{bmatrix}$$

The following illustrates model (1),

$$\begin{bmatrix} 7 \\ 9 \\ 8 \\ 3 \\ 6 \\ 4 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \end{bmatrix}$$

The columns of X define respectively (i) the overall intercept, (ii) the indicator of the first ration, (iii) the indicator of the second ration type, and (iv) the indicator of the third ration type. The matrix X is a *non-full column rank* matrix since one column can be obtained by manipulating other columns. For example, the first column defining the intercept is equal to the *sum* of the other columns. Assume α_i become the levels of the i^{th} ration, then we can express α_0 the column of the overall intercept, as the sum of the others, i.e., $\alpha_0 = \sum \alpha_i$.

Earlier information leads to the resulting normal equations for not having a *unique* inverse, making it impossible to find a *unique* estimate vector β^0 of the true parameter β .

In this example:

$$X'X = \begin{bmatrix} 6 & 3 & 1 & 2 \\ 3 & 3 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 2 & 0 & 0 & 2 \end{bmatrix} \text{ and } X'y = \begin{bmatrix} 37 \\ 24 \\ 3 \\ 10 \end{bmatrix}$$

are generalized inverse (g-inverse matrix), G_1 is

$$(X'X)^- = \begin{bmatrix} 0.1145833 & -0.03125 & 0.13541667 & 0.0104166 \\ -0.031250 & 0.281250 & -0.218750 & -0.093750 \\ 0.1354166 & -0.21875 & 0.61458333 & -0.260416 \\ 0.1041667 & -0.09375 & -0.2604167 & 0.3645833 \end{bmatrix} = G_1$$

This inverse is computed using the SWEEP operation in *SAS/IML^R* (SAS, 2004).

4. Set-to-Zero Restrictions

The problem could be calculated by the generalized inverse G_2 of a *modified* $X'X$ matrix,

$$\begin{bmatrix} 6 & 3 & 1 & 0 \\ 3 & 3 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

by setting the *last* column and last *row* to zero. Searle (1987) referred as ‘*constraints on the solution*’ to obtain LIN (linear independent) functions of the solutions elements rather than *restrictions*. The reduced 3x3 matrix has

a unique inverse. Below is such an inverse in which the zeros are inserted back to the original position.

$$G_2 = \begin{bmatrix} 0.5 & -0.5 & -0.5 & 0 \\ -0.5 & 0.83333 & 0.5 & 0 \\ -0.5 & 0.5 & 1.5 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

One can also set-to-zero the column *before the last*.

$$\begin{bmatrix} 6 & 3 & 0 & 2 \\ 3 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 2 \end{bmatrix}$$

The resulting generalized inverse G_3 is

$$G_3 = \begin{bmatrix} 1 & -1 & 0 & -1 \\ -1 & 1.8 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1.5 \end{bmatrix}$$

The generalized inverse G_4 can be found to be the same way, thus, by setting the second column to zero.

$$\begin{bmatrix} 6 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 2 & 0 & 0 & 2 \end{bmatrix}$$

$$G_4 = \begin{bmatrix} 0.33333 & 0 & -0.33333 & -0.33333 \\ 0 & 0 & 0 & 0 \\ -0.33333 & 0 & 1.33333 & 0.33333 & 1 \\ -0.33333 & 0 & -.33333 & -0.83333 \end{bmatrix}$$

4.1 Solution Vectors

From each one of the above generalized inverse, a different solution vector to the normal equations is obtained. The table below presents such solutions:

Table 2. Solutions to the normal equations using each generalized (G_i)

Elements of the solution	Solution (see Equation (9))			
	1	2	3	4
$\widehat{\alpha}_0$	4	5	3	8
$\widehat{\alpha}_1$	4	3	5	0
$\widehat{\alpha}_2$	-1	-2	0	-5
$\widehat{\alpha}_3$	1	0	2	-3

4.2 Sum-of-Zero Restrictions

As for the set-to-zero restrictions, the sum-to-zero restriction converts an over-parameterized model to a model in which elements of the parameter are *estimable* and the normal equations have a single solution. These restrictions are mostly useful in balanced data but are less attractive for cases in which some subclasses are empty (Searle, 1987). If all the columns of the matrix X^* are independent of each other, then multiplying by its transpose X^{*t} generates a non-singular matrix that has a *unique* inverse $(X^{*t}X^*)^{-1}$. This may be achieved by letting one of the columns of α_i be the negative sum of the others, hence the name sum to zero restrictions.

$$X^* = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \end{bmatrix}$$

The created columns under this restriction have the values of 0, 1, or -1 (Cook & Weisberg, 1987; SAS, 1999 and other standard software). In general for α levels, there are $\alpha - 1$ independent terms ($\alpha_1 \dots \alpha_{\alpha-1}$). They are

$$\alpha_j = \begin{cases} 1, & \text{if } S = j \\ -1, & \text{if } S = a \\ 0, & \text{otherwise} \end{cases} \tag{11}$$

In our example, we have three treatment levels (rations). The sum-to-zero restriction implies that one of the level in α_j , say α_3 is equal to $-\alpha_1 - \alpha_2$. In other words α_4 is equal to $-\sum \alpha_i$ ($i = 1, 2$). In this case, the usual ordinary least square (OLS) solution is uniquely defined by:

$$b = (X^* ' X^*)^{-1} X^* ' y \tag{12}$$

which is the solution vector of β and contains $[\alpha_0, \alpha_1, \alpha_2]$ where α_0 is the *deviation* from the overall mean for ration 1, α_1 is the mean *deviation* for ration 1, and α_2 is the mean *deviation* for ration 2. The restriction allows *direct* estimation of the model parameters. This method is also referred to as the *Least Squares Dummy Variables* or LSDV3 in Hun (2009). Now from the above we have:

$$X'X = \begin{bmatrix} 6 & 1 & -1 \\ 1 & 5 & 2 \\ -1 & 2 & 3 \end{bmatrix} \text{ and } X'y = \begin{bmatrix} 37 \\ 14 \\ -7 \end{bmatrix}$$

where the g-inverse matrix (G_5) is

$$(X'X)^- = \begin{bmatrix} 0.2037037 & -0.092593 & 0.1296296 \\ -0.092593 & 0.3148148 & -0.240741 \\ 0.1296296 & -0.240741 & 0.5370370 \end{bmatrix} = G_5$$

The solution b is

$$b_5 = \begin{bmatrix} 5.33333 \\ 2.66667 \\ -2.33333 \end{bmatrix}$$

Alternatively the matrix can be set differently with the restriction $\alpha_2 = -\alpha_1 - \alpha_3$

$$X^o = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & -1 & -1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \text{ and } X^o'y = \begin{bmatrix} 37 \\ 21 \\ 7 \end{bmatrix}$$

Consequently,

$$X^{o'}X^o = \begin{bmatrix} 6 & 2 & 1 \\ 2 & 4 & 1 \\ 1 & 1 & 3 \end{bmatrix}$$

The generalized inverse is

$$G_6 = \begin{bmatrix} 0.2037037 & -0.092593 & -0.037037 \\ -0.092593 & 0.3148148 & -0.074074 \\ -0.037037 & -0.074074 & 0.5370370 \end{bmatrix}$$

and the solution is

$$b_6 = \begin{bmatrix} 5.33333 \\ 2.66667 \\ -3.33333 \end{bmatrix}$$

The last restriction is on α_1 setting it equal to $-\alpha_2 - \alpha_3$. As before, the important matrices are

$$X^a = \begin{bmatrix} 1 & -1 & -1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}, \quad X^{a'}X^a = \begin{bmatrix} 6 & -2 & -1 \\ -2 & 4 & 3 \\ -1 & 3 & 5 \end{bmatrix} \text{ and } X^{a'}y = \begin{bmatrix} 37 \\ -21 \\ -14 \end{bmatrix}$$

The generalized inverse matrix (G_7) is

$$(X^a X^a)^- = \begin{bmatrix} 0.2037037 & 0.1296296 & -0.037037 \\ 0.1296296 & 0.537037 & -0.296296 \\ -0.037037 & -0.2296296 & 0.3703704 \end{bmatrix} = G_7$$

The solution b_7 is

$$\begin{bmatrix} 5.33333 \\ -2.33333 \\ -0.33333 \end{bmatrix}$$

4.3 Estimable Functions

It is obvious by now that there are many solutions to the same question of interest. From the simple example of Table 1, only seven (7) of all the possible results are shown. These estimators should *never* be considered *valid estimates of the treatment effects* since they are all *biased*. This property (though not desired) must be overcome. The use of estimable functions is a very important step to do that. Estimable functions lead to estimates that are best linear unbiased estimates (BLUE). The estimates such as the means or differences between means are *linear* functions of the observations. They are *unbiased*, thus, their expected values are equal to the true parameters. Such estimates have, among other possible linear functions, the *smallest* variances (Scheffe, 1959; Searle, 1987).

We then apply the estimable functions to obtain BLUEs of various descriptive statistics of interest. For this introduction in the treatment of the general linear model (1), we limit ourselves to the estimation of the following:

- Treatment means (or means of the three rations),
- Various contrasts between means and respective variances
- Sum of squares between treatments (i.e., rations)
- Error sum of squares and the corresponding mean squares
- Hypotheses testing of the type $H_0: K_i' \widehat{\beta} = 0$ vs. $H_1: K_i' \widehat{\beta} \neq 0$, where K_i is an estimable vector.

In particular we exploit the *invariance* property of the estimable functions. Table 3 presents different vectors, denoted by K that, when multiplied by the various solution vectors (from both set-to-zero and sum-to-zero restrictions). They yield identical results regardless of the solution vector used. Each estimate is defined by (12):

$$K' \widehat{\beta} \tag{13}$$

where $\widehat{\beta}$ defines any of the seven (7) solution vectors. The check of 'estimability' of $K' \widehat{\beta}$ is done by verifying whether the product the $KHK' = K$ is true, where the matrix $H = X(X'X')^-X'$.

Table 3. Estimable vectors K for the treatment means and contrasts

(*)	Ration			Contrasts	
	1	2	3	1 vs. 2	(1+3) vs. 2
$\widehat{\alpha}_0$	1	1	1	0	0
$\widehat{\alpha}_1$	1	0	0	1	0.5
$\widehat{\alpha}_2$	0	1	0	-1	-1
$\widehat{\alpha}_3$	0	0	1	0	0.5

For example, to compute the mean of ration 1 using solution 1 of Table 4, we have

$$K_1' \widehat{\beta} = 1 * 4 + 1 * 4 + 0 * (-1) + 0 * 1 = 8$$

To compute the same mean using solution 3 of Table 4, we have $K_3' \widehat{\beta} = 1 * 3 + 1 * 5 + 0 * 0 + 0 * 2 = 8$.

The invariance property of estimable functions is shown in the Table 4 below.

Table 4. Estimates of $K'\widehat{\beta}$ from set-to-zero constraints

Elements of the solution	Solution (see Equation (9))			
	1	2	3	4
$\widehat{\alpha}_0$	4	5	3	8
$\widehat{\alpha}_1$	4	3	5	0
$\widehat{\alpha}_2$	-1	-2	0	-5
$\widehat{\alpha}_3$	1	0	2	-3

Means

$K'_1\widehat{\beta}$	8	8	8	8
$K'_2\widehat{\beta}$	3	3	3	3
$K'_3\widehat{\beta}$	5	5	5	5

Contrasts

$K'_4\widehat{\beta}$	5=(4-(-1))	5=(3-(-2))	=(5 - 0)	=0-(-5)
$K'_5\widehat{\beta}$	-1 - {1/2*4 + (1/2*1)} -3.5	3.5	3.5	3.5

Sum of Squares

a. SS Ration = $\widehat{\beta}'X'y - N\bar{y}^2$ are virtually identical for all the $\widehat{\beta}'$ s.

22.833333	22.833314	22.833314	22.833603
-----------	-----------	-----------	-----------

b. SSEerror = $\mathbf{e}'\mathbf{e} = (y - X\widehat{\beta})'(y - X\widehat{\beta})$ are identical for all $\widehat{\beta}'$ s.

4	4	4	4
---	---	---	---

Mean Square

c. Ration = *SS Ration*

22.833333/2	11.4167	11.4167	11.4168
-------------	---------	---------	---------

d. Error = SSEerror/df_{error} = $\widehat{\sigma}^2$

4/3	1.3333	1.3333	1.3333
-----	--------	--------	--------

F_obs

$\frac{MS_{Ration}}{MS_{Error}}$	8.56	8.56	8.56	8.56
----------------------------------	------	------	------	------

Tests of Hypothesis

e. $H_0: K'_4\widehat{\beta} = 0$ vs. $H_1: K'_4\widehat{\beta} \neq 0$. This is obtained by applying the following formula to get the appropriate standard error:

$$[K'_4(X'X)^{-1}K'_4\widehat{\sigma}^2]^{1/2}$$

1.33333	1.33333	1.33333	1.33333
---------	---------	---------	---------

Similarly the hypothesis $H_0: K'_5\widehat{\beta} = 0$ vs. $H_1: K'_5\widehat{\beta} \neq 0$ is computed the same way:

$$[K'_5(X'X)^{-1}K'_5\widehat{\sigma}^2]^{1/2}$$

2.538591	2.538591	2.538591	2.538591
----------	----------	----------	----------

Note that the elements of the solution are the means of the ration 1 i.e., $K_1'\widehat{\beta}$, ration 2 i.e., $K_2'\widehat{\beta}$, and ration 3 i.e., $K_3'\widehat{\beta}$ respectively (see Table 4). Therefore all the contrasts and the combination of estimable functions are also estimable (Searle 1971, 1987; Milliken, 1971).

4.6 Special Cases of the General Linear Model

Two special cases deserve to be mentioned. The first is the simple linear model where the predictor is made of one continuous variable x_i measured on each individual $i = 1, \dots, I$ leading to the model:

$$y_i = \beta_0 + \beta x_i + \varepsilon_i \tag{14}$$

The estimation of the parameters β_0 and β are straightforward since the columns of the model matrix are LIN. The second special case is simply a reduction of the number of treatment to two. In the simple case, the two-sample t-test assumes the response $y_{ij} \sim N(\mu_i, \sigma^2)$. The model can use dummy variables to indicate group membership (Kiebel & Holmes, n.d.). This parameterization of the model without an intercept term is similar to the one explained above.

4.7 Principle of Conditional Error or Extra-Sum of Squares

The principle for conditional error was first introduced by Bose (1949) in the context of estimates of combinations of the parameters of a linear model (Milliken, 1971). The basic idea is to compute sum of squares corresponding to the hypothesis by first obtaining the sum of squares due to error for the restricted or reduced model minus the sum of squares due to error from the unrestricted model. The degrees of freedom (df) are computed by the difference in degrees-of-freedom of the corresponding sum of squares. Milliken and Johnson (1992) used this procedure in a strategy to determine the final form of an analysis of covariance model. By other authors, the procedure is known as the model comparison method (Draper & Smith, 1981), or the extra sum of squares principle (Draper and Smith, 1998). The principle is quite general (McDonald & Milliken, 1974) and find extensive use in econometrics (Greene, 2003; Baltagi, 2008; Wooldridge, 2003), in linear models (Searle, 1971, 1987; Milliken & Johnson, 1984; Gujarati, (1970) as well as in non-linear models (Milliken & Debruin, 1978). The ratio of the difference between the SS due to error of the restricted model (RSS_0) and the SS error due to the unrestricted model (RSS_1) to the error SS of the full model is:

$$F_{obs} = \frac{(RSS_0 - RSS_1)/(p_0 - p_1)}{\frac{RSS_1}{J-p_1-1}} \sim F_{p_0 - p_1; J-p_1-1} \tag{15}$$

where p_1 is the rank of the full matrix X , p_0 is the rank of X_2 , the matrix of the reduced model, and J is the total number of observations. The numerator is distributed as a non-central chi-square random variable; and the denominator is distributed as a central chi-square. Both chi-square variables are independent (Milliken, 1971). The full and reduced models can be explained in matrix form by partitioning in two parts as follow:

$$y = [X_1 : X_2] \begin{bmatrix} \alpha_1 \\ \dots \\ \alpha_2 \end{bmatrix} + \varepsilon \tag{16}$$

Equation (15) defines the full model where the complete X matrix and consequently the β vector were partitioned into two components respectively. The interest is to test the hypothesis

$$H_0 : \alpha_2 = 0 \tag{17}$$

In the above example, let the matrix X be subdivided into two matrices: one (X_1) to represent the column of intercept and the other matrix (X_2) is for the three levels of ration (treatment). In this special case, the F_{obs} in (14) is exactly the standard F-statistics of the usual one-way analysis of variance (ANOVA).

Table 5. Application of extra-sum of squares in one-way ANOVA

Model Name	Formulation	df	Residuals	
			SS	F_{obs}
Reduced	$y_{ij} = \mu$	5	26.8333	$\frac{(26.833-4.0)/(5-3)}{4.0/3} = 8.56$
Full	$y_{ij} = \mu + \alpha_i$	3	4.0000	

The above $F_{obs} = 8.56$ is exactly the same as the usually reported test from a one-way analysis of variance (see also its derivation above). The extra sum of squares principle is easily extended to an analysis that involves more than one factor. The statistical analysis system (SAS, 2004) produces a partial type of sum of squares known as Type III SS in which the effects in the full model may each be considered a product of the principle of conditional error.

4.8 Alternative Formulations to Extra Sum of Squares

The above tests can be re-expressed in terms of R^2

$$\frac{(R_1^2 - R_0^2)/p}{(1 - R_1^2)/(T - p - 1)} \tag{18}$$

R_1^2 and R_0^2 refer to the coefficients of determination for the unrestricted and the restricted model respectively. The tests can also be articulated in terms of the residuals e for the unrestricted, e^* for the restricted model respectively.

$$\frac{(e^*{}' e^* - e' e)/p}{e' e/(T - r_k)} \tag{19}$$

A third way of testing the linear restriction hypothesis is by way of matrices presented earlier

$$\frac{K' \beta^o (K' G K)^{-1} K \beta^{o' '}}{\sigma^2} \tag{20}$$

If one replaces σ^2 by its sample estimate $r_k * S^2$ adjusted for the df , then (19) is the F-statistic

$$\frac{K' \beta^o (K' G K)^{-1} K \beta^{o' '}}{r_k S^2} \tag{21}$$

Asymptotically all three tests described are equivalent. However, they can give different or conflicting results in a finite sample. This fact was raised by Greene (2003) and Geweke, 1984). It is therefore advised that, for a given sample, and whenever possible, one should run them all and pay particular consideration to the choice of significance level. Bozivich, Bancroft and Hartley (1956) proposed to use an $\alpha > 0.05$ in order to control for Type II error. Hendrix et al. (1982) used 0.25.

Table 6. Equivalence of the alternative formulations

Alternative Formulas to (14)		
$\frac{(R_1^2 - R_0^2)/p}{(1 - R_1^2)/(T - p - 1)}$	$\frac{(e^*{}' e^* - e' e)/p}{e' e/(T - r_k)}$	$\frac{K' \beta^o (K' G K)^{-1} K \beta^{o' '}}{r_k S^2}$
$F_{obs} = \frac{(0.850932 - 0.0/2)}{(1 - 0.850932)/(6 - 2 - 1)} = \frac{(26.833 - 4.0)/2}{\frac{4}{3}} = \frac{[3 \ -2] \begin{bmatrix} 1.5 & -0.5 \\ -0.5 & 0.8333 \end{bmatrix} \begin{bmatrix} 3 \\ -2 \end{bmatrix}}{2 * (\frac{4}{3})} = 8.56$		

The result from $K' \beta^o (K' G K)^{-1} K \beta^{o' '}$ is not affected by use of any of the G and β^o presented earlier.

4.9 Goodness-of-Fit Using Likelihood Ratio Test

A likelihood ratio test is often used in linear and nonlinear (fixed/mixed) models to compare the goodness of fit of two competing models, one nested within the other. Both the full and the reduced nested models fit the data by the method of maximum likelihood and the values of the respective log-likelihoods are retained. The likelihood ratio test (LRT) is

$$\tau = \frac{l(\beta_0)}{l(\beta)} \tag{22}$$

Under the null,

$$LRT = -2 \ln(\tau) = -2 \ln\{l(\beta) - l(\beta_0)\} \tag{23}$$

Equivalently one can compute $-2 \ln(\tau^*)$ where τ^* is

$$\tau^* = \left\{ 1 + \frac{Q}{SSR_1} \right\}^{-\frac{1}{2} T} \tag{24}$$

where Q is the quadratic given by

$$K'\beta^o\{K'GK\}^{-1}K\beta^o'/r_k \quad (25)$$

To put in the LRT the frame of the principle conditional error, Equations (22) and (24) are reformulated as

$$-2\ln\left\{\frac{RS S_1}{SSR_0}\right\}^{\frac{1}{2}T} \sim \text{Chi - Square} \quad (26)$$

The test-statistic is *chi square* distributed with the number of degrees of freedom (df) equal the number of restrictions of the so-called parameter space (that is, the number of restrictions imposed by the null hypothesis). In our particular case, $df = 2$ since the interest lies on whether the inclusion of the *rations* is above the model containing the *intercept* only.

Table 7. Some choices of LRT formulas in linear models

Likelihood Ratio Tests		
$-2\ln\{l(\beta) - l(\beta_0)\}$	$-2\ln\left\{1 + \frac{Q}{SSR_1}\right\}^{-\frac{1}{2}T}$	$-2\ln\left\{\frac{RSS_1}{SSR_0}\right\}^{\frac{1}{2}T}$
$-2\ln(-7.297 + 13.007)$	$-2\ln\left\{1 + \frac{11.41333}{4.00}\right\}^{-\frac{1}{2}6}$	$-2\ln\left\{\frac{4}{26.833}\right\}^{\frac{1}{2}6}$

4.10 Additional Resources

Searle R. S. et al. (1981) presented additional computational and model equivalence in analyses of variance of unequal-subclass-numbers data. More on these issues is found in Speed et al. (1978), Manzumdar S. et al. (1980), Milliken and Johnson (1984), McClean (1989), Kiebel and Holmes (1986), Healy (1986), Searle (1971, 1987), Mould (1989), Hocking (1986), Draper and Smith (1999) among others.

5. Conclusion

In this study, we have shown four ways of how a general linear unbalanced model can be analyzed. These different methods of analysis lead to exactly the same estimates regardless of how the re-parameterization of the linear model was done and what generalized inverse used. Importantly, we showed that the results are not software dependent. This was made possible through the use of estimable functions. The study also outlined the concepts of *least squares means* and the *partition* of sum-of squares in more complex design of experiments and again extended the property of estimable functions to derive *interaction* terms in a model that does not specify such interaction in its original formulation. All these concepts have been explained and illustrated using a practical examples.

References

- Cook, D., & Weisberg, S. (1987). *Applied Regression Including Computing and Graphics*. Wiley Series in Probability and Statistics.
- Draper, & Smith. (1999). *Applied Linear Regression Analysis*. NY: John Wiley & Sons.
- Healy, M. (1986). *Matrix for Statistics*. Oxford: Oxford University Press.
- Hocking, R. R. (1985). *The Analysis of Linear Models*. Monterey, California:Brooks-Cole.
- Hun, M. P. (2009). *Linear Regression Models for Panel Data Using SAS, Stata, LIMDEP, and SPSS*. University Information Techn. Sciences Indiana Univ.
- Kempthorne, O. (1952). *The Design and Analysis of Experiments*. New York: John Wiley and Sons.
- Kiebel, S. J., & Holmes, A. P. (n.d.). *The General Linear Model*.
- Manzumdar, S. et al. (1980). Correspondence between a Linear Restriction and a Generalized Inverse in Linear Model Analysis. *The American Statistician*, 34(2), 103-105.
- McClean, R. A. (1989). An introduction to general linear models. In *Applications of Mixed Models in Agriculture and Related Disciplines*, Southern Cooperative Series Bulletin No. 343, pp. 9-31. Louisiana Agriculture Experiment Station. Baton Rouge.
- Milliken, G. A. (1971). Criteria for Estimability for Linear Models. *The Annals of Mathematical Statistics*, 42(5), 1588-1594.
- Milliken, G. A., & Johnson. (1984). *Analysis of Messy Data* (Vol. II). CRC Press.

- Mould, R. (1989). *Introductory Medical Statistics* (2nd ed.). London: Institute of Physics Publishing.
- Neter, J. W., Wasserman, & Kutner, M. H. (1985). *Applied Linear Statistical Models*. Richard Irwin, Inc.
- Scheffe, H. (1959). *The Analysis of Variance*. New York: Wiley.
- Searle, R. (1971). *Linear Models*. New York, NY:Wiley and Sons.
- Searle, R. (1987). *Linear Models for Unbalanced Data*. John Wiley & Sons Inc.
- Searle, R. S. et al. (1981). Some Computational and Model Equivalence in Analyses of Variance of Unequal-subclass-numbers Data. *The American Statistician*, 35, 16-33.
- Speed et al. (1978). Methods of analysis of linear models with unbalanced data. *Journ. Stat. Assoc*, 73, 105-112.
<http://dx.doi.org/10.1080/01621459.1978.10480012>

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).