FAULTS IN AN ALGORITHM FOR REPARAMETERIZING LINEAR MODELS

by

S. R. Searle and Harold V. Henderson*
Biometrics Unit, Cornell University, Ithaca, New York

BU-710-M July, 1980

Abstract

An algorithm for reparameterizing linear models so that effects due to the levels of each factor add to zero is shown to fail for certain interaction models for data that have empty cells.

1. Introduction

Linear models that are not of full rank are often reparameterized to be of full rank by imposing restrictions on the parameters of the model. One popular form of such restrictions is that which is coming to be known as (e.g., Searle, et al. 1981) the Σ-restrictions. These define the effects for each factor so that they add to zero; for example, if α_i for $i = 1, \cdots, a$ represent the effects due to levels of a factor A, then the Σ-restrictions define the α_i's such that $\sum_{i=1}^{a} \alpha_i = 0$.

The Σ-restrictions have a long history in linear model theory for the analysis of data from well designed and well executed experiments, data that are usually called orthogonal, balanced or equal-subclass-numbers data. This paper describes and illustrates a popular algorithm for applying these restrictions, and demonstrates its faults for certain kinds of non-orthogonal, unbalanced or unequal-subclass-numbers data. This is done in terms of a two-factor model, the 2-way crossed classification, but it clearly extends to situations of more than two factors.

---

* Biometrician, Ruakura Agricultural Research Centre, Hamilton, New Zealand.
Paper No. BU-710-M in the Biometrics Unit.

2. Data With All Cells Filled

The model equation for the familiar two-way crossed classification over-parameterized model, with a rows, b blocks, and n observations in each cell can be represented as

$$E(y_{ijk}) = \mu + \alpha_i + \beta_j + \gamma_{ij} , \tag{1}$$

where y_{ijk} is the k'th observation on the i'th row in the j'th column, for $k = 1, \cdots, n$, and where μ is a general mean, α_i is the effect due to the i'th row for $i = 1, \cdots, a$, β_j is the effect due to the j'th column for $j = 1, \cdots, b$, and γ_{ij} is the interaction between the i'th row and j'th column; and E represents expectation over repeated sampling. An equivalent model for (1) is

$$E(\underset{\sim}{y}) = \underset{\sim\sim}{X}\underset{\sim}{b} , \tag{2}$$

where $\underset{\sim}{y}$ is the vector of observations arrayed in lexicon order and $\underset{\sim}{b}$ is the vector of parameters in the model, $\underset{\sim}{b}' = [\mu \ \ \underset{\sim}{\alpha}' \ \ \underset{\sim}{\beta}' \ \ \underset{\sim}{\gamma}']$ for $\underset{\sim}{\alpha} = \{\alpha_i\}$, $\underset{\sim}{\beta} = \{\beta_j\}$ and $\underset{\sim}{\gamma} = \{\gamma_{ij}\}$ for $i = 1, \cdots, a$ and $j = 1, \cdots, b$; and $\underset{\sim}{X}$ is the incidence matrix.

Example

For data where $a = 2$, $b = 3$ and $n = 2$, which we represent by Grid 1,

Grid 1

2	2	2
2	2	2

$$\underset{\sim}{b}' = [\mu \ \ \alpha_1 \ \ \alpha_2 \ \ \beta_1 \ \ \beta_2 \ \ \gamma_{11} \ \ \gamma_{12} \ \ \gamma_{13} \ \ \gamma_{21} \ \ \gamma_{22} \ \ \gamma_{23}]$$

and

$$\underset{\sim}{X} = \begin{bmatrix} 1 & 1 & . & 1 & . & . & 1 & . & . & . & . & . \\ 1 & 1 & . & 1 & . & . & 1 & . & . & . & . & . \\ 1 & 1 & . & . & 1 & . & . & 1 & . & . & . & . \\ 1 & 1 & . & . & 1 & . & . & 1 & . & . & . & . \\ 1 & 1 & . & . & . & 1 & . & . & 1 & . & . & . \\ 1 & 1 & . & . & . & 1 & . & . & 1 & . & . & . \\ 1 & . & 1 & 1 & . & . & . & . & . & 1 & . & . \\ 1 & . & 1 & 1 & . & . & . & . & . & 1 & . & . \\ 1 & . & 1 & . & 1 & . & . & . & . & . & 1 & . \\ 1 & . & 1 & . & 1 & . & . & . & . & . & 1 & . \\ 1 & . & 1 & . & . & 1 & . & . & . & . & . & 1 \\ 1 & . & 1 & . & . & 1 & . & . & . & . & . & 1 \end{bmatrix} \qquad (3)$$

where a dot represents zero.

An abbreviated form of $\underset{\sim}{X}$, based on noting that for all observations in the same cell the rows of $\underset{\sim}{X}$ are the same, is shown in Table 1, where n_{ij} represents the number of observations in row i and column j of the data, in this case $n_{ij} = 2$ for all cells, as in Grid 1.

Table 1: Rows of the $\underset{\sim}{X}$-matrix for the over-parameterized model for data of Grid 1

No. of rows in $\underset{\sim}{X}$	Column in $\underset{\sim}{X}$											
	μ	α_1	α_2	β_1	β_2	β_3	γ_{11}	γ_{12}	γ_{13}	γ_{21}	γ_{22}	γ_{23}
$n_{11} = 2$	1	1	.	1	.	.	1	.	.	.	.	.
$n_{12} = 2$	1	1	.	.	1	.	.	1	.	.	.	.
$n_{13} = 2$	1	1	.	.	.	1	.	.	1	.	.	.
$n_{21} = 2$	1	.	1	1	.	.	.	.	.	1	.	.
$n_{22} = 2$	1	.	1	.	1	.	.	.	.	.	1	.
$n_{23} = 2$	1	.	1	.	.	1	.	.	.	.	.	1

For data having unequal numbers of observations in the cells, but with every cell having some observations, Grid 2 applies:

Grid 2

n_{11}	n_{12}	n_{13}
n_{21}	n_{22}	n_{23}

and so does Table 1 with deletion of the 2's in the first column.

Notice a characteristic of each γ_{ij}-column in Table 1 which is, of course, also a characteristic of the corresponding column of $\underset{\sim}{X}$ in (3): each γ_{ij}-column is the product, element by element, of the elements in the associated α_i-column and β_j-column. For example, in Table 1 the

$$\gamma_{12}\text{-column is } \begin{bmatrix} . \\ 1 \\ . \\ . \\ . \\ . \end{bmatrix} = \begin{bmatrix} 1 \times 0 \\ 1 \times 1 \\ 1 \times 0 \\ 0 \times 0 \\ 0 \times 1 \\ 0 \times 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} * \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} = (\alpha_1\text{-column}) * (\beta_2\text{-column}). \quad (4)$$

This element-by-element product, represented here by the symbol *, is a well established matrix operation known as the Hadamard or Schur product. Its general definition for any two matrices $\underset{\sim}{A} = \{a_{ij}\}$ and $\underset{\sim}{B} = \{b_{ij}\}$ of the same order is $\underset{\sim}{A} * \underset{\sim}{B} = \{a_{ij}b_{ij}\}$. Styan (1973) gives a history of this product and discusses its properties. Its use here is the special case of the matrices being columns, and as illustrated in (4) it applies to each of the γ_{ij}-columns in Table 1.

Consider reparameterizing the model (1) to make it a model of full rank, by means of the Σ-restrictions. In doing so, distinguish the elements of the reparameterized model from those of (1) by using the same symbols but with a dot above them; e.g., $\dot{\alpha}_i$ in contrast to α_i. And make the reparameterization, as is often done, by using each Σ-restriction to replace the effect having the largest

subscript (i.e., the "last" effect) by minus all the others. For data from Grid 1 or 2 the Σ-restrictions and replacements of the nature described are as follows.

Σ-restriction	Replacement
$\dot{\alpha}_1 + \dot{\alpha}_2 = 0$	$\dot{\alpha}_2 = -\dot{\alpha}_1$
$\dot{\beta}_1 + \dot{\beta}_2 + \dot{\beta}_3 = 0$	$\dot{\beta}_3 = -\dot{\beta}_1 - \dot{\beta}_2$
$\dot{\gamma}_{11} + \dot{\gamma}_{12} + \dot{\gamma}_{13} = 0$	$\dot{\gamma}_{11} = \dot{\gamma}_{11}$
$\dot{\gamma}_{21} + \dot{\gamma}_{22} + \dot{\gamma}_{23} = 0$	$\dot{\gamma}_{12} = \dot{\gamma}_{12}$
$\dot{\gamma}_{11} + \dot{\gamma}_{21} = 0$	$\dot{\gamma}_{13} = -\dot{\gamma}_{11} - \dot{\gamma}_{12}$
$\dot{\gamma}_{12} + \dot{\gamma}_{22} = 0$	$\dot{\gamma}_{21} = -\dot{\gamma}_{11}$, $\dot{\gamma}_{22} = -\dot{\gamma}_{12}$
$\dot{\gamma}_{13} + \dot{\gamma}_{23} = 0$	$\dot{\gamma}_{23} = \dot{\gamma}_{11} + \dot{\gamma}_{12}$.

(5)

The obvious equations $\dot{\gamma}_{11} = \dot{\gamma}_{11}$ and $\dot{\gamma}_{12} = \dot{\gamma}_{12}$ are included for the sake of completeness. Their inclusion emphasizes that all the $\dot{\gamma}$'s are represented by just 2 = (a-1)(b-1) different γ's. After using $\dot{\alpha}_i$, $\dot{\beta}_j$ and $\dot{\gamma}_{ij}$ in place of α_i, β_j and γ_{ij} in (1), and then using the replacements of (5), the reparameterized model shall be written as

$$E(\underset{\sim}{y}) = \underset{\sim\sim}{\dot{X}}\underset{\sim}{b} ,$$

whereupon $\underset{\sim}{\dot{X}}$ has rows as shown in Table 2.

Table 2: Rows of the $\dot{X}$-matrix for data from Grids 1 or 2, after using the Σ-restrictions

No. of rows in $\underset{\sim}{\dot{X}}$	Column in $\underset{\sim}{\dot{X}}$					
	$\dot{\mu}$	$\dot{\alpha}_1$	$\dot{\beta}_1$	$\dot{\beta}_2$	$\dot{\gamma}_{11}$	$\dot{\gamma}_{12}$
n_{11}	1	1	1	.	1	.
n_{12}	1	1	.	1	.	1
n_{13}	1	1	-1	-1	-1	-1
n_{21}	1	-1	1	.	-1	.
n_{22}	1	-1	.	1	.	-1
n_{23}	1	-1	-1	-1	1	1

Again we see the Hadamard product (H-p) algorithm operating: each of the $(a-1)(b-1) = 2$ $\dot{\gamma}$-columns is the H-p of the corresponding $\dot{\alpha}$-column and $\dot{\beta}$-column. And this is so whenever all cells have data in them, even for situations with more than two factors. For example, the column in an $\dot{\underset{\sim}{X}}$-matrix corresponding to a 3-factor interaction is the H-p of the three columns corresponding to the appropriate levels of the main effects. But, in contrast, the crux of this paper is that in certain cases of data grids with empty cells this H-p algorithm does not operate. Furthermore, this seems to be why certain statistical computer packages are unable to analyze such data.

3. Data With Empty Cells

3.1. At least one row and one column with all cells filled

Consider Grid 3, the same as Grid 2 but with cell 1,2 being empty.

Grid 3

n_{11}	-	n_{13}
n_{21}	n_{22}	n_{23}

The corresponding $\dot{\underset{\sim}{X}}$-matrix is shown in Table 3.

Table 3: Rows of the $\dot{\underset{\sim}{X}}$-matrix for data from Grid 3, after using the Σ-restrictions

No. of rows in $\dot{\underset{\sim}{X}}$	Column in $\dot{\underset{\sim}{X}}$				
	$\dot{\mu}$	$\dot{\alpha}_1$	$\dot{\beta}_1$	$\dot{\beta}_2$	$\dot{\gamma}_{11}$
n_{11}	1	1	1	.	1
n_{13}	1	1	-1	-1	-1
n_{21}	1	-1	1	.	-1
n_{22}	1	-1	.	1	.
n_{23}	1	-1	-1	-1	1

Once again the H-p algorithm is seen to be operating: the $\dot{\gamma}_{11}$-column is the Hadamard product of the $\dot{\alpha}_1$- and $\dot{\beta}_1$-columns.

Note, though, that not every possible Hadamard product occurs: e.g., that of the $\dot{\alpha}_1$- and $\dot{\beta}_2$-columns does not occur, because the $(\dot{\alpha}_1, \dot{\beta}_2)$ combination corresponds to an empty cell. Indeed, when all cells are filled, as in Grids 1 and 2, there are $(a-1)(b-1)$ $\dot{\gamma}$'s, and all possible $(a-1)(b-1)$ Hadamard products of an $\dot{\alpha}_i$-column and a $\dot{\beta}_j$-column exist, for $i = 1, \cdots, a-1$ and $j = 1, \cdots, b-1$. But when N_0 cells have no data, and $s = ab - N_0$ contain data, then only

$$(a-1)(b-1) - N_0 = s - a - b + 1 \qquad (6)$$

Hadamard products are required. It would seem that this is the reason that the computer routine BMDP2V is unable to handle interaction models with empty cells, for which it instead prints out the message "The number of parameters to be estimated exceeds the total number of degrees of freedom. This is usually caused by missing cells." Apparently it assumes $(a-1)(b-1)$ to be the number of parameters and $s - a - b + 1$ to be the degrees of freedom, so that the output message is indeed a consequence of (6).

Now consider Grid 4:

Grid 4

n_{11}	n_{12}	-
n_{21}	n_{22}	n_{23}

It is essentially the same as Grid 3, the only difference being that in Grid 4 it is the 1,3 cell that is empty rather than the 1,2 cell. Indeed, interchanging column labels 2 and 3 in Grid 3 yields Grid 4. And Table 4 shows the corresponding $\dot{\underset{\sim}{X}}$-matrix.

Table 4: Rows of the $\dot{\underset{\sim}{X}}$-matrix for data from Grid 4, after using the Σ-restrictions

No. of rows in $\dot{\underset{\sim}{X}}$	Column in $\dot{\underset{\sim}{X}}$				
	$\dot{\mu}$	$\dot{\alpha}_1$	$\dot{\beta}_1$	$\dot{\beta}_2$	$\dot{\gamma}_{11}$
n_{11}	1	1	1	.	1
n_{12}	1	1	.	1	-1
n_{21}	1	-1	1	.	-1
n_{22}	1	-1	.	1	1
n_{23}	1	-1	-1	-1	.

And now the H-p algorithm is <u>not</u> operating: the $\dot{\gamma}_{11}$-column in Table 4 is not the H-p of the $\dot{\alpha}_1$- and $\dot{\beta}_1$-columns. (If it were operating, the [1 -1 -1 1 .]' of the $\dot{\gamma}_{11}$-column would be [1 . -1 . 1]'.) The inconsistency of this with Table 3 (wherein the H-p algorithm is operating), arises solely from the location of the empty cell in Grid 4 compared to its location in Grid 3. The importance of its location in Grid 4 is that the last column in Grid 4 does not have all cells filled – whereas it does in Grid 3. The reason that this otherwise apparently inconsequential difference of Grid 4 from Grid 3 leads to the breakdown of the

H-p algorithm is as follows.

Customary usage of the H-p algorithm is based, as here, on replacing the last effect in each Σ-restriction by minus all the others therein, as illustrated in equations (5); e.g., from $\dot{\beta}_1 + \dot{\beta}_2 + \dot{\beta}_3 = 0$, $\dot{\beta}_3$ is replaced by $-\dot{\beta}_1 - \dot{\beta}_2$. The H-p algorithm is always a correct representation of these replacements when all cells are filled, as in Table 2. And it is also correct when some cells are empty, providing the last row and column of the Grid have all cells filled – as in Grid 3 and Table 3. Then, so far as $\dot{\gamma}$'s are concerned, having empty cells is equivalent to having all cells filled but with simply deleting the $\dot{\gamma}$'s corresponding to empty cells. Thus Table 3 is simply Table 2 with the $\dot{\gamma}_{12}$-column and the n_{12}-row deleted.

But the H-p algorithm is not a correct representation of the "replace the last effect" usage of the Σ-restrictions for data Grid 4. This is because the last row and the last column of the data grid do not have all cells filled. The H-p algorithm works only when all cells in the last row and column of the data grid have data in them. This is so because having these cells filled ensures that in each Σ-restriction for the $\dot{\gamma}$'s, there is in that last row and column a $\dot{\gamma}$ (in the model for the data) that can be replaced by other $\dot{\gamma}$'s. For example, in Grid 3 the Σ-restriction for $\dot{\gamma}$'s in row 1 is $\dot{\gamma}_{11} + \dot{\gamma}_{13} = 0$, from which the $\dot{\gamma}_{13}$ of the last column can be replaced by $-\dot{\gamma}_{11}$. But in Grid 4 that Σ-restriction is $\dot{\gamma}_{11} + \dot{\gamma}_{12} = 0$ which contains no $\dot{\gamma}_{13}$ corresponding to the last column. Of course, an appropriate replacement is readily ascertained in easy cases like this one, but a general algorithm using this procedure is usually tied to making all replacements from one row and one column, and so requires that row and column to have all cells filled. This is why the SAS HARVEY procedure requires that the data have (or be resequenced to have) the last row and column having all cells filled. And Option 9 in the SPSS ANOVA procedure has the same requirement except

there it must be the first row and column that have all cells filled.

3.2. No row and/or no column having all cells filled

Data Grid 5 is for the example of Table 7.6 in Searle (1971).

Grid 5

3	-	1	2
2	2	-	-
-	2	2	4

It is also data set 5 of the Annotated Computer Outputs (see Searle, 1979). The $\dot{\underset{\sim}{X}}$-matrix is indicated in Table 5.

Table 5: Rows of the $\dot{\underset{\sim}{X}}$-matrix for data from Grid 5, after using the Σ-restrictions

No. of rows in $\dot{\underset{\sim}{X}}$	Column of $\dot{\underset{\sim}{X}}$							
	$\dot{\mu}$	$\dot{\alpha}_1$	$\dot{\alpha}_2$	$\dot{\beta}_1$	$\dot{\beta}_2$	$\dot{\beta}_3$	$\dot{\gamma}_{11}$	$\dot{\gamma}_{13}$
$n_{11} = 3$	1	1	.	1	.	.	1	.
$n_{13} = 1$	1	1	.	.	.	1	.	1
$n_{14} = 2$	1	1	.	-1	-1	-1	-1	-1
$n_{21} = 2$	1	.	1	1	.	.	-1	.
$n_{22} = 2$	1	.	1	.	1	.	1	.
$n_{32} = 2$	1	-1	-1	.	1	.	-1	.
$n_{33} = 2$	1	-1	-1	.	.	1	.	-1
$n_{34} = 4$	1	-1	-1	-1	-1	-1	1	1

The H-p algorithm is not operating here: the $\dot{\gamma}_{11}$-column does not equal the H-p of the $\dot{\alpha}_1$- and $\dot{\beta}_1$-columns. If it did, the elements -1, 1 and -1 in the rows of Table 5 corresponding to n_{21}, n_{22} and n_{32} would be 0, 0 and 0. This is what they are in the associated rows of the $\underset{\sim}{X}$-matrix of the SAS HARVEY procedure. Presumably this error is the basis for that procedure's not being able to analyze such data

and yielding the error message "1 error(s) force abnormal termination of procedure Harvey". Certain it is that Grid 5 has neither a row nor a column with all cells filled and so does not satisfy the requirement of SAS HARVEY in this regard.

The pattern of $\dot{\gamma}$'s occurring in the data of Grid 5, corresponding to Table 5, is shown in Table 6. It is clear that the Σ-restrictions for $\dot{\gamma}$'s are satisfied.

Table 6: Occurrence of $\dot{\gamma}$'s in data from Grid 5, corresponding to Table 5

$\dot{\gamma}_{11}$	-	$\dot{\gamma}_{13}$	$-\dot{\gamma}_{11} - \dot{\gamma}_{13}$
$-\dot{\gamma}_{11}$	$\dot{\gamma}_{11}$	-	-
-	$-\dot{\gamma}_{11}$	$-\dot{\gamma}_{13}$	$\dot{\gamma}_{11} + \dot{\gamma}_{13}$

In contrast the pattern of $\dot{\gamma}$'s corresponding to what Table 5 is for SAS HARVEY is shown in Table 7.

Table 7: Occurrence of $\dot{\gamma}$'s in data from Grid 5, corresponding to the form of Table 5 coming from SAS HARVEY (the same as Table 5 as shown, except for 0's in the $\dot{\gamma}_{11}$-column for the n_{21}, n_{22} and n_{32} rows)

$\dot{\gamma}_{11}$	-	$\dot{\gamma}_{13}$	$-\dot{\gamma}_{11} - \dot{\gamma}_{13}$
		-	-
-		$-\dot{\gamma}_{13}$	$\dot{\gamma}_{11} + \dot{\gamma}_{13}$

Clearly, the Σ-restrictions are not satisfied in Table 7.

4. Conclusion

The Σ-restrictions to reparameterize a linear model must be used with care in the presence of empty cells, because the Hadamard-product representation of them fails in certain cases of empty cells in the presence of interactions. This appears to be the reason for the SAS HARVEY and SPSS ANOVA Option 9 statistical

computing package routines being unable to analyze certain cases of such data. BMDP2V avoids the issue altogether by not analyzing missing-cell data with interaction models at all. SAS GLM uses a generalized inverse technique that does not get explicitly involved with reparameterization.

References

Searle, S. R. (1971). Linear Models. Wiley, New York.

Searle, S. R. (1979). Annotated Computer Output for analysis of variance of unequal-subclass-numbers data. The American Statistician 33, 222-223.

Searle, S. R., Speed, F. M. and Henderson, H. V. (1981). Some computational and model equivalences in analyses of variance of unequal-subclass-numbers data. The American Statistician. (in press)

Styan, G. P. H. (1973). Hadamard products and multivariate statistical analysis. Linear Algebra and its Applications 6, 217-240.