ESTIMABILITY IN THE CELL MEANS GENERAL LINEAR MODEL

by

BU-730-M   S. R. Searle[1] and F. M. Speed[2]   January, 1981

Abstract

The cell means representation of the general linear model usually involves no restrictions on the cell means, in which case the mean for each cell that contains data is estimable. So is any linear combination of those means. Including restrictions on the cell means as part of the model does not alter estimability of means of cells containing data, but it does affect the estimability of means of empty cells and linear combinations involving them. A theorem identifies which combinations are estimable.

1. Introduction

The cell means representation of the general linear model does not specifically identify factors and levels of factors in the traditional design of experiments sense. For example, if y_{ijk} is the k'th observation in the i'th level of factor A and j'th level of factor B, the traditional model equation for y_{ijk} is of the form

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

where μ is a general mean, α_i is the effect of the i'th level of A, β_j the effect of the j'th level of B and γ_{ij} the interaction effect, with E representing expectation over repeated sampling. In contrast, the cell means representation of this

1/ Biometrics Unit, Cornell University, Ithaca, New York.

2/ Speedman Oil Inc., Corpus Christi, Texas.

is

$$E(y_{ijk}) = \mu_{ij} \; . \tag{2}$$

This has also been called the μ_{ij}-model.

Motivation for the cell means model has, over the last decade, come from the widening realization that models like (1), with their problems of over-parameterization, are not satisfactory for data in which some cells defined by the factors contain no data (empty cells). Cell means models, on the other hand, cope very adequately with the difficulties brought on by the occurrence of empty cells. Speed (1969), Searle (1971), Urquhart et al. (1973), Speed et al. (1978), and Urquhart and Weeks (1978) are some of the places where the merits of the cell means model are expounded.

In its simplest and yet most general form, the cell means model can be represented as

$$E(\underset{\sim}{y}) = \underset{\sim}{X}\underset{\sim}{\mu} \tag{3}$$

where $\underset{\sim}{y}$ is the vector of data, with all observations on each cell occurring consecutively in $\underset{\sim}{y}$. Then $\underset{\sim}{X} = \bigoplus_{t=1}^{k} \underset{\sim}{1}_{n_t \times 1}$, a direct sum of $\underset{\sim}{1}$-vectors (vectors having every element unity), where n_t is the number of observations in the t'th cell containing data (filled cell), there being k such cells; and $\underset{\sim}{\mu}$ is the vector of the population cell means of these k cells. Here, and throughout the paper, the dispersion matrix of $\underset{\sim}{y}$ is taken to be $\sigma^2 \underset{\sim}{I}_N$ where N is the total number of observations, $N = n_{.} = \sum_{t=1}^{k} n_t$.

Direct application of the principle of least squares to (3) shows, very easily, that the best, linear, unbiased, estimator (b.l.u.e.) of $\underset{\sim}{\mu}$ is $\bar{\underset{\sim}{y}}$, the vector of observed cell means. Thus each cell mean μ_t is estimated by its corresponding observed cell mean, and the sampling variance of that estimator is σ^2/n_t.

Example 1

Consider a 2-way crossed classification of two rows and two columns. Grid 1 shows the cell means when all cells are filled.

Grid 1

μ_{11}	μ_{12}
μ_{21}	μ_{22}

.

The model (3) is

$$E(\underset{\sim}{y}) = \begin{bmatrix} \underset{\sim}{1}_{n_{11}} & \cdot & \cdot & \cdot \\ \cdot & \underset{\sim}{1}_{n_{12}} & \cdot & \cdot \\ \cdot & \cdot & \underset{\sim}{1}_{n_{21}} & \cdot \\ \cdot & \cdot & \cdot & \underset{\sim}{1}_{n_{22}} \end{bmatrix} \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{21} \\ \mu_{22} \end{bmatrix} = \underset{\sim}{X}\underset{\sim}{\mu} \tag{4}$$

where n_{ij} is the number of observations in the cell defined by row i and column j; and a dot in a matrix represents zeros. The normal equations, which come from minimizing $(\underset{\sim}{y} - \underset{\sim}{X}\underset{\sim}{\mu})'(\underset{\sim}{y} - \underset{\sim}{X}\underset{\sim}{\mu})$ with respect to each element of $\underset{\sim}{\mu}$ are, as is well known,

$$\underset{\sim}{X}'\underset{\sim}{X}\hat{\underset{\sim}{\mu}} = \underset{\sim}{X}'\underset{\sim}{y}, \text{ i.e., } \begin{bmatrix} n_{11} & \cdot & \cdot & \cdot \\ \cdot & n_{12} & \cdot & \cdot \\ \cdot & \cdot & n_{21} & \cdot \\ \cdot & \cdot & \cdot & n_{22} \end{bmatrix} \begin{bmatrix} \hat{\mu}_{11} \\ \hat{\mu}_{12} \\ \hat{\mu}_{21} \\ \hat{\mu}_{22} \end{bmatrix} = \begin{bmatrix} y_{11\cdot} \\ y_{12\cdot} \\ y_{21\cdot} \\ y_{22\cdot} \end{bmatrix} \tag{5}$$

where $y_{ij\cdot}$ is the total of the observations in the (i,j) cell. Clearly, from (5)

$$\hat{\mu}_{ij} = \bar{y}_{ij\cdot} \ .$$

This is the most usual application of the cell means model, and is patently straightforward: the population mean of a filled cell is estimated by the corresponding observed cell mean, and all linear combinations of such populations cell means are estimable. Also, under normality assumptions, tests of hypotheses can be made about them. For example, the F-statistic for testing $H: \underset{\sim}{T}'\underset{\sim}{\mu} = \underset{\sim}{m}$ is

$$F = (\underset{\sim}{T}'\bar{\underset{\sim}{y}} - \underset{\sim}{m})'(\underset{\sim}{T}'\underset{\sim}{D}\underset{\sim}{T})^{-1}(\underset{\sim}{T}'\underset{\sim}{y} - \underset{\sim}{m})/s\hat{\sigma}^2 \tag{6}$$

where $\underset{\sim}{T}'$ has full row rank s, $\underset{\sim}{D}$ is the diagonal matrix of diagonal elements $1/n_t$, and $\hat{\sigma}^2$ is the pooled within-cell mean square. When $\underset{\sim}{m} \equiv \underset{\sim}{0}$, then (6) simplifies to the more familiar

$$F = \bar{\underset{\sim}{y}}'\underset{\sim}{T}(\underset{\sim}{T}'\underset{\sim}{D}\underset{\sim}{T})^{-1}\underset{\sim}{T}'\bar{\underset{\sim}{y}}/s\hat{\sigma}^2 \tag{7}$$

2. A More General Cell Means Model

The 2-way cross-classification [for which (1) is the traditional over-parameterized model] has, by definition, ab cells when there are a levels of one factor and b of the other. When some cells are empty, the cell mean $\mu + \alpha_i + \beta_j + \gamma_{ij}$ is estimable only for filled cells. But if the over-parameterized model <u>without interaction</u> is used, $E(y_{ijk}) = \mu + \alpha_i + \beta_j$ and the cell mean $\mu + \alpha_i + \beta_j$ is estimable for every cell, whether filled or empty. The cell means model (3) is therefore unsuitable in this without-interaction case because it utilizes (and estimates means for) only the filled cells. Consequently a more general model is needed. It must involve <u>all</u> the cell means and, in a no-interaction case like that just considered, it must take account of the absence of interactions.

At first thought, involving all the cell means seems easy. $\underset{\sim}{\mu}$ is defined as being the vector of all cell means that could exist in the data, even if some are empty; e.g., for the 2-way crossed classification, $\underset{\sim}{\mu}$ contains ab elements. And then in writing the model equation as

$$E(\underset{\sim}{y}) = \underset{\sim}{W}\underset{\sim}{\mu} \tag{8}$$

$\underset{\sim}{W}$ would be a direct sum of $\underset{\sim}{1}$-vectors, as is $\underset{\sim}{X}$ of (3), <u>except</u> that corresponding to each cell mean for the empty cells there is a column of zeros in $\underset{\sim}{W}$.

Example 2

Suppose in Grid 1 there are no data in cell (2,2), indicated in Grid 2 by having μ_{22} in parentheses:

Grid 2

μ_{11}	μ_{12}
μ_{21}	(μ_{22})

.

Then the model equation here is

$$E(\underset{\sim}{y}) = \begin{bmatrix} \underset{\sim}{1}_{n_{11}} & \cdot & \cdot & 0 \\ \cdot & \underset{\sim}{1}_{n_{12}} & \cdot & 0 \\ \cdot & \cdot & \underset{\sim}{1}_{n_{21}} & 0 \end{bmatrix} \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{21} \\ \mu_{22} \end{bmatrix} = \underset{\sim}{W}\underset{\sim}{\mu} \tag{9}$$

where the column of 0's (scalar zeros) corresponds to μ_{22}, the cell mean of the empty cell.

<u>2.1. Unrestricted models</u>

The estimation of $\underset{\sim}{\mu}$ from (8) by least squares leads, quite formally, to normal equations

$$\underset{\sim}{W}'\underset{\sim}{W}\hat{\underset{\sim}{\mu}} = \underset{\sim}{W}'\underset{\sim}{y} \; . \tag{10}$$

Because $\underset{\sim}{W}$ has null columns, as illustrated in (9), $\underset{\sim}{W}'\underset{\sim}{W}$ has null rows and $\underset{\sim}{W}'\underset{\sim}{y}$ has zero elements, corresponding to empty cells. A solution of (10) is therefore

$$\hat{\mu} = (W'W)^{-}W'y \tag{11}$$

where the superscript minus indicates a generalized inverse matrix. Since $W'W$ is diagonal with diagonal elements n_{ij} (including $n_{ij} = 0$ for each empty cell), a straightforward value for $(W'W)^{-}$ is that it be diagonal with diagonal elements $[1 - \delta_{0,n_{ij}}]/[n_{ij} + \delta_{0,n_{ij}}]$, where $\delta_{0,n_{ij}}$ is a Kronecker delta; i.e., $(W'W)^{-}$ has diagonal elements $1/n_{ij}$ for $n_{ij} \neq 0$ and 0 for $n_{ij} = 0$. Then (11) gives

$$\hat{\mu}_{ij} = \bar{y}_{ij} \quad \text{for each filled cell}$$

and

$$\hat{\mu}_{ij} = 0 \quad \text{for each empty cell.} \tag{12}$$

There is a weakness in the estimation procedure of (12): without any data being available in the empty cells, the means of those cells are being estimated (as zero). This seems to contradict intuition, that without having data in those cells their means cannot be estimated and so are non-estimable. There is also an inconsistency in considering (10) as the normal equations for the model (8). They are, but only pro forma. Least squares is a procedure for estimating parameters of a model for a set of data. But (8) has zero columns corresponding to μ_{ij}'s for empty cells, as illustrated in (9), so that those μ_{ij}'s do not appear explicitly in (8) at all. Therefore the least squares procedure of differentiating $(y - W\mu)'(y - W\mu)$ with respect to μ_{ij}'s can be applied only for the μ_{ij}'s of the filled cells. To execute this, re-sequence elements of μ so that those for filled cells, to be denoted as μ_f, come first, followed by those for empty cells, μ_e. Then

$$\mu = \begin{bmatrix} \mu_f \\ \mu_e \end{bmatrix} . \tag{13}$$

Define X as in (3): a direct sum of $\mathbf{1}$-vectors corresponding to filled cells.

Then (8) is

$$E(\underset{\sim}{y}) = \underset{\sim}{X}\underset{\sim}{\mu}_f + \underset{\sim}{0}\underset{\sim}{\mu}_e \ . \tag{14}$$

Least squares applied to (14) now clearly involves differentiating $(\underset{\sim}{y} - \underset{\sim}{X}\underset{\sim}{\mu}_f)'(\underset{\sim}{y} - \underset{\sim}{X}\underset{\sim}{\mu}_f)$ with respect to elements of $\underset{\sim}{\mu}_f$ so giving

$$\hat{\underset{\sim}{\mu}}_f = (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{y} = \bar{\underset{\sim}{y}} \ , \tag{15}$$

as the estimator of $\underset{\sim}{\mu}_f$. ($\bar{\underset{\sim}{y}}$ represents the vector of observed cell means.)

The estimates in (15) are, of course, precisely the same for filled cells as those in (12). But the situation is not the same for empty cells. In (15), nothing is said about their means. And neither anything should be said about them because those cells have no data and their means are not estimable in the model formulated as (8).

2.2. Restricted models

The last sentence of the preceding section may seem to imply that nothing can ever be said about the estimation of cell means of empty cells. This is clearly not so because, for example, we know with the no-interaction over-parameterized model $E(y_{ijk}) = \mu + \alpha_i + \beta_j$ for data from something like Grid 2, that the cell mean $\mu + \alpha_2 + \beta_2$ for a missing cell can be estimated. This situation is accommodated in the cell means model as follows.

First, for data like those of Grid 1, with no empty cells, the no-interaction feature is embodied in the cell means model by imposing restrictions on the μ_{ij}'s such as

$$\mu_{11} - \mu_{12} - \mu_{21} + \mu_{22} = 0 \ . \tag{16}$$

In general, let restrictions, homogeneous in the elements of $\underset{\sim}{\mu}$, be represented as

$$\underset{\sim}{G}\underset{\sim}{\mu} = \underset{\sim}{0} \ , \tag{17}$$

where $\underset{\sim}{\mu}$, as in (13), involves all the cell means, both for filled cells and empty cells. Indeed, partition $\underset{\sim}{G}$ as $\underset{\sim}{G} = [\underset{\sim}{G}_f \ \ \underset{\sim}{G}_e]$ conformably with (13), so that (17) is

$$\underset{\sim}{G}_f \underset{\sim}{\mu}_f + \underset{\sim}{G}_e \underset{\sim}{\mu}_e = \underset{\sim}{0} \ . \tag{18}$$

The general, restricted, cell means model is now (14) and (18) combined:

$$E(\underset{\sim}{y}) = \underset{\sim}{X} \underset{\sim}{\mu}_f + \underset{\sim}{0} \underset{\sim}{\mu}_e \tag{14}$$

and

$$\underset{\sim}{G}_f \underset{\sim}{\mu}_f + \underset{\sim}{G}_e \underset{\sim}{\mu}_e = \underset{\sim}{0} \ . \tag{18}$$

Insofar as applying least squares to this is concerned it contains a logical difficulty: the data, (14), involve only $\underset{\sim}{\mu}_f$, whereas the restrictions, (18), also involve $\underset{\sim}{\mu}_e$. But do they? In terms of the model for the data, (18) can be viewed as simply defining some other parameters $\underset{\sim}{\mu}_e$. And, from (18) these have the form (using standard results for solving linear equations, e.g., Theorem 4 in Searle, 1971, page 12)

$$\underset{\sim}{\mu}_e = -\underset{\sim}{G}_e^- \underset{\sim}{G}_f \underset{\sim}{\mu}_f + (\underset{\sim}{I} - \underset{\sim}{G}_e^- \underset{\sim}{G}_e)\underset{\sim}{\tau} \tag{19}$$

where $\underset{\sim}{G}_e^-$ is a generalized inverse of $\underset{\sim}{G}_e$ and $\underset{\sim}{\tau}$ is any arbitrary vector of appropriate order. Then, using (19) in (18) the restrictions on $\underset{\sim}{\mu}_f$ are

$$(\underset{\sim}{I} - \underset{\sim}{G}_e \underset{\sim}{G}_e^-)\underset{\sim}{G}_f \underset{\sim}{\mu}_f = \underset{\sim}{0} \ . \tag{20}$$

Since $\underset{\sim}{I} - \underset{\sim}{G}_e \underset{\sim}{G}_e^-$ is idempotent, (20) means that

$$\underset{\sim}{G}_f \underset{\sim}{\mu}_f = \underset{\sim}{0} + \underset{\sim}{G}_e \underset{\sim}{G}_e^- \underset{\sim}{\omega} \tag{21}$$

for any arbitrary vector $\underset{\sim}{\omega}$. We therefore take the restrictions as

$$\underset{\sim}{G}_f \underset{\sim}{\mu}_f = \underset{\sim}{0} \ . \tag{22}$$

The model on which we carry out least squares is now (14) and (22),

$$E(\underset{\sim}{y}) = \underset{\sim}{X}\underset{\sim}{\mu}_f + \underset{\sim}{0} \quad \text{and} \quad \underset{\sim}{G}_f\underset{\sim}{\mu}_f = \underset{\sim}{0}$$

or, equivalently

$$E(\underset{\sim}{y}) = \underset{\sim}{X}\underset{\sim}{\mu}_f \quad \text{and} \quad \underset{\sim}{G}_f\underset{\sim}{\mu}_f = \underset{\sim}{0} \,. \tag{23}$$

The resulting normal equations are

$$\begin{bmatrix} \underset{\sim}{X}'\underset{\sim}{X} & \underset{\sim}{G}_f' \\ \underset{\sim}{G}_f & \underset{\sim}{0} \end{bmatrix} \begin{bmatrix} \hat{\underset{\sim}{\mu}}_f \\ \underset{\sim}{\lambda} \end{bmatrix} = \begin{bmatrix} \underset{\sim}{X}'\underset{\sim}{y} \\ \underset{\sim}{0} \end{bmatrix} \tag{24}$$

where $\underset{\sim}{\lambda}$ is a vector of Lagrange multipliers. Solutions to (24) are

$$\hat{\underset{\sim}{\mu}}_f = (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{y} - (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{G}_f'[\underset{\sim}{G}_f(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{G}_f']^{-}\underset{\sim}{G}_f(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{y}$$

and (25)

$$\underset{\sim}{\lambda} = [\underset{\sim}{G}_f(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{G}_f']^{-}\underset{\sim}{G}_f(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{y} \,.$$

That these satisfy (24) is readily verified, and in particular we note that

$$\underset{\sim}{G}_f\hat{\underset{\sim}{\mu}}_f = \underset{\sim}{0} \,. \tag{26}$$

Furthermore, using (23)

$$E(\hat{\underset{\sim}{\mu}}_f) = \underset{\sim}{\mu}_f \,. \tag{27}$$

And also, because $\underset{\sim}{X}$ is a direct sum of $\underset{\sim}{1}$-vectors,

$$(\underset{\sim}{X}'\underset{\sim}{X})^{-1} = \text{diag}\{1/n_{ij}\} \equiv \underset{\sim}{D} \tag{28}$$

for n_{ij}'s of the filled cells, and $(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{y} = \bar{\underset{\sim}{y}}$. Hence we have the slightly simpler form for $\hat{\underset{\sim}{\mu}}_f$ of (25) as

$$\hat{\underset{\sim}{\mu}}_f = \bar{\underset{\sim}{y}} - \underset{\sim}{D}\underset{\sim}{G}_f'(\underset{\sim}{G}_f\underset{\sim}{D}\underset{\sim}{G}_f')^{-}\underset{\sim}{G}_f\bar{\underset{\sim}{y}} \,. \tag{29}$$

Of course, without restrictions there is no $\underset{\sim}{G}_f$ and so then

$$\hat{\underset{\sim}{\mu}}_f = \bar{\underset{\sim}{y}} \quad \text{when there are no restrictions.} \tag{30}$$

Also, on replacing $\underset{\sim}{\mu}_f$ in (19) by $\hat{\underset{\sim}{\mu}}_f$ we have

$$\hat{\underset{\sim}{\mu}}_e = -\underset{\sim}{G}_e^-\underset{\sim}{G}_f\hat{\underset{\sim}{\mu}}_f + (\underset{\sim}{I} - \underset{\sim}{G}_e^-\underset{\sim}{G}_e)\underset{\sim}{\tau}$$

which, from (26) is

$$\hat{\underset{\sim}{\mu}}_e = (\underset{\sim}{I} - \underset{\sim}{G}_e^-\underset{\sim}{G}_e)\underset{\sim}{\tau} \tag{31}$$

for arbitrary $\underset{\sim}{\tau}$.

Note, too, that when there are no empty cells there is no $\underset{\sim}{\mu}_e$ nor $\underset{\sim}{G}_e$, $\underset{\sim}{\mu}_f \equiv \underset{\sim}{\mu}$ and $\underset{\sim}{G}_f \equiv \underset{\sim}{G}$ and the normal equations (24) are

$$\begin{bmatrix} \underset{\sim}{X}'\underset{\sim}{X} & \underset{\sim}{G}' \\ \underset{\sim}{G} & \underset{\sim}{0} \end{bmatrix} \begin{bmatrix} \hat{\underset{\sim}{\mu}} \\ \underset{\sim}{\lambda} \end{bmatrix} = \begin{bmatrix} \underset{\sim}{X}'\underset{\sim}{y} \\ \underset{\sim}{0} \end{bmatrix} . \tag{32}$$

3. Estimability

We consider the estimability of

$$\underset{\sim}{\delta} = \underset{\sim}{H}\underset{\sim}{\mu} = \underset{\sim}{H}_f\underset{\sim}{\mu}_f + \underset{\sim}{H}_e\underset{\sim}{\mu}_e \tag{33}$$

for matrices $\underset{\sim}{H}$ partitioned as $\underset{\sim}{H} = [\underset{\sim}{H}_f \;\; \underset{\sim}{H}_e]$. Clearly, when there are no empty cells, as in (31), there is no $\underset{\sim}{H}_e$ and $\underset{\sim}{H}_f \equiv \underset{\sim}{H}$, and

$$\widehat{\underset{\sim}{H}\underset{\sim}{\mu}} = \underset{\sim}{H}\hat{\underset{\sim}{\mu}} \quad \text{for any } \underset{\sim}{H}, \text{ for all cells filled.} \tag{34}$$

Also, when there are empty cells but $\underset{\sim}{H}_e \equiv \underset{\sim}{0}$

$$\widehat{\underset{\sim}{H}\underset{\sim}{\mu}} = \widehat{\underset{\sim}{H}_f\underset{\sim}{\mu}_f} = \underset{\sim}{H}_f\hat{\underset{\sim}{\mu}}_f \quad \text{for } \underset{\sim}{H}_e \equiv \underset{\sim}{0} . \tag{35}$$

Both (34) and (35) are valid whether or not there are restrictions $\mathbf{G}\boldsymbol{\mu} = \mathbf{0}$; and when there are not, $\hat{\boldsymbol{\mu}} = \bar{\mathbf{y}}$ in (34) and $\hat{\boldsymbol{\mu}}_f = \bar{\mathbf{y}}$ in (35). These simple cases constitute part (i) of the theorem that follows, which also deals with more general situations, namely the estimability of functions $\mathbf{H}\boldsymbol{\mu}$ that involve the cell means of empty cells.

Theorem. When $\boldsymbol{\mu}' = [\boldsymbol{\mu}_f' \;\; \boldsymbol{\mu}_e']$ is the vector of cell means in a linear model, with $\boldsymbol{\mu}_f$ corresponding to filled cells and $\boldsymbol{\mu}_e$ to empty cells, and when $\mathbf{G}\boldsymbol{\mu} \equiv \mathbf{G}_f\boldsymbol{\mu}_f + \mathbf{G}_e\boldsymbol{\mu}_e = \mathbf{0}$ are the restrictions on $\boldsymbol{\mu}$, then estimable functions are either $\mathbf{H}\boldsymbol{\mu}$ with best linear unbiased estimator (b.l.u.e.) $\widehat{\mathbf{H}\boldsymbol{\mu}} = \mathbf{H}_f\hat{\boldsymbol{\mu}}_f$, or $\mathbf{LH}\boldsymbol{\mu}$ with b.l.u.e. $\widehat{\mathbf{LH}\boldsymbol{\mu}} = \mathbf{LH}_f\hat{\boldsymbol{\mu}}_f$, under the following conditions.

(i) When all cells are filled ($\mathbf{H}_e$ non-existent), $\widehat{\mathbf{H}\boldsymbol{\mu}} = \mathbf{H}\hat{\boldsymbol{\mu}} \equiv \mathbf{H}_f\hat{\boldsymbol{\mu}}_f$; or when some cells are empty, if $\mathbf{H}_e = \mathbf{0}$ then $\widehat{\mathbf{H}\boldsymbol{\mu}} = \mathbf{H}_f\hat{\boldsymbol{\mu}}_f$.

(ii) If $\mathbf{G}_e$ has full column rank, $\widehat{\mathbf{H}\boldsymbol{\mu}} = \mathbf{H}_f\hat{\boldsymbol{\mu}}_f$.

(iii) If $\mathbf{H}_e = \mathbf{MG}_e$ for some $\mathbf{M}$, then $\widehat{\mathbf{H}\boldsymbol{\mu}} = \mathbf{H}_f\hat{\boldsymbol{\mu}}_f$.

(iv) If $\mathbf{LH}_e = \mathbf{MG}_e$ for some $\mathbf{L}$ and $\mathbf{M}$, then $\widehat{\mathbf{LH}\boldsymbol{\mu}} = \mathbf{LH}_f\hat{\boldsymbol{\mu}}_f$.

Proof. From partitioning $\mathbf{H}$ and $\boldsymbol{\mu}$

$$\mathbf{H}\boldsymbol{\mu} = \mathbf{H}_f\boldsymbol{\mu}_f + \mathbf{H}_e\boldsymbol{\mu}_e$$

which, by (19), is

$$\mathbf{H}\boldsymbol{\mu} = \mathbf{H}_f\boldsymbol{\mu}_f - \mathbf{H}_e\mathbf{G}_e^-\mathbf{G}_f\boldsymbol{\mu}_f + (\mathbf{H}_e - \mathbf{H}_e\mathbf{G}_e^-\mathbf{G}_e)\boldsymbol{\tau}$$

so that

$$\widehat{\mathbf{H}\boldsymbol{\mu}} = \mathbf{H}_f\hat{\boldsymbol{\mu}}_f - \mathbf{H}_e\mathbf{G}_e^-\mathbf{G}_f\hat{\boldsymbol{\mu}}_f + (\mathbf{H}_e - \mathbf{H}_e\mathbf{G}_e^-\mathbf{G}_e)\boldsymbol{\tau}$$

which by (26) is

$$\widehat{\mathbf{H}\boldsymbol{\mu}} = \mathbf{H}_f\hat{\boldsymbol{\mu}}_f + (\mathbf{H}_e - \mathbf{H}_e\mathbf{G}_e^-\mathbf{G}_e)\boldsymbol{\tau} \qquad (36)$$

and this contains no arbitrary vector τ whenever

$$H_e = H_e G_e^- G_e \ ; \tag{37}$$

i.e., (37) is the condition under which $H\mu$ is estimable with b.l.u.e.

$$\widehat{H\mu} = H_f \hat{\mu}_f \ . \tag{38}$$

It is clear that at least the following situations satisfy (37):

(i) H_e not existing, or $H_e = 0$,

(ii) G_e of full column rank, in which case $G_e^- = (G_e' G_e)^{-1} G_e'$ and so $G_e^- G_e = I$,

and

(iii) $H_e = MG_e$, because then $H_e G_e^- G_e = MG_e G_e^- G_e = MG_e = H_e$.

These correspond to parts (i) - (iii) of the theorem. And part (iv) holds because

(iv) if $LH_e = MG_e$ then $LH_e = LH_e G_e^- G_e$ is valid because $LH_e G_e^- G_e = MG_e G_e^- G_e = MG_e = LH_e$. Q.E.D.

Comments

(a) When all cells are filled only part (i) of the theorem applies.

(b) When some cells are empty but $H_e = 0$, parts (iii) and (iv) of the theorem are satisfied and only part (i) applies, regardless of the form of G.

(c) If there are no restrictions $G\mu = 0$, only part (i) applies, with $\hat{\mu}_f = \bar{y}$.

(d) If there are restrictions on μ, then $H\mu$ is estimable only when $H_e = H_e G_e^- G_e$.

(e) In parts (i) - (iii) of the theorem, the hypothesis $H : H\mu = m$ can be tested using

$$F = (H_f \hat{\mu}_f - m)'(H_f V^{-1} H_f)^{-1}(H_f \hat{\mu}_f - m)/s\hat{\sigma}^2$$

where

$$\underset{\sim}{V} = \text{var}(\hat{\underset{\sim}{\mu}}_f)/\sigma^2 = (\underset{\sim}{X}'\underset{\sim}{X})^{-1} - (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{G}_f'[\underset{\sim}{G}_f(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{G}_f']^{-}\underset{\sim}{G}\underset{\sim}{X}'\underset{\sim}{X}'$$

s = full row rank of $\underset{\sim}{H}$

$\hat{\sigma}^2$ = residual mean square

$$= \underset{\sim}{y}'[\underset{\sim}{I} - \underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-}\underset{\sim}{X}']\underset{\sim}{y}/[N - r(\underset{\sim}{X}) - r(\underset{\sim}{G}_f)]$$

where $r(\underset{\sim}{X})$ and $r(\underset{\sim}{G}_f)$ are the ranks of $\underset{\sim}{X}$ and $\underset{\sim}{G}_f$, respectively, with $r(\underset{\sim}{X})$ being the number of filled cells.

4. Examples

4.1. Example 1 (continued)

There are no empty cells and so part (i) of the theorem applies: $\underset{\sim}{H}\underset{\sim}{\mu}$ is estimable, whether there are restrictions on $\underset{\sim}{\mu}$ or not. With no restrictions, $\hat{\underset{\sim}{\mu}} = \bar{\underset{\sim}{y}}$, i.e., $\hat{\mu}_{ij} = \bar{y}_{ij\cdot}$. With the restriction

$$\mu_{11} - \mu_{12} - \mu_{21} + \mu_{22} = 0 \tag{39}$$

the normal equations (32) are

$$\begin{bmatrix} n_{11} & \cdot & \cdot & \cdot & 1 \\ \cdot & n_{12} & \cdot & \cdot & -1 \\ \cdot & \cdot & n_{21} & \cdot & -1 \\ \cdot & \cdot & \cdot & n_{22} & 1 \\ 1 & -1 & -1 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mu_{11}^o \\ \mu_{12}^o \\ \mu_{21}^o \\ \mu_{22}^o \\ \lambda \end{bmatrix} = \begin{bmatrix} y_{11\cdot} \\ y_{12\cdot} \\ y_{21\cdot} \\ y_{22\cdot} \\ 0 \end{bmatrix} \tag{40}$$

with solution

$$\mu_{ij}^o = \bar{y}_{ij\cdot} - (-1)^{i+j}\lambda/n_{ij}$$

for (41)

$$\lambda = (\bar{y}_{11\cdot} - \bar{y}_{12\cdot} - \bar{y}_{21\cdot} + \bar{y}_{22\cdot})/\theta \quad \text{and} \quad \theta = 1/n_{11} + 1/n_{12} + 1/n_{21} + 1/n_{22}\ .$$

4.2. Example 2 (continued)

Here, in the unrestricted model, $\hat{\mu}_{ij} = \bar{y}_{ij\cdot}$, except for $\hat{\mu}_{22}$ which does not exist; μ_{22} is not estimable. But in keeping with part (i) of the theorem, linear combinations of μ_{11}, μ_{12} and μ_{21} are estimable, i.e., any $\underset{\sim}{H}_f\underset{\sim}{\mu}_f$ for $\underset{\sim}{\mu}_f' = [\mu_{11}\ \mu_{12}\ \mu_{21}]$ is estimable.

With the restriction $\underset{\sim}{G}\underset{\sim}{\mu} = \mu_{11} - \mu_{12} - \mu_{21} + \mu_{22} = 0$, the submatrix $\underset{\sim}{G}_f$ of $\underset{\sim}{G}$ is $\underset{\sim}{G}_f = [1\ \ -1\ \ -1]$ and the normal equations (24) are

$$\begin{bmatrix} n_{11} & \cdot & \cdot & 1 \\ \cdot & n_{12} & \cdot & -1 \\ \cdot & \cdot & n_{21} & -1 \\ 1 & -1 & -1 & 0 \end{bmatrix} \begin{bmatrix} \hat{\mu}_{11} \\ \hat{\mu}_{12} \\ \hat{\mu}_{21} \\ \lambda \end{bmatrix} = \begin{bmatrix} \bar{y}_{11\cdot} \\ \bar{y}_{12\cdot} \\ \bar{y}_{21\cdot} \\ 0 \end{bmatrix}$$

with solution for $\hat{\underset{\sim}{\mu}}_f$

$$\hat{\underset{\sim}{\mu}}_f = \begin{bmatrix} \bar{y}_{11\cdot} - \lambda/n_{11} \\ \bar{y}_{12\cdot} + \lambda/n_{12} \\ \bar{y}_{21\cdot} + \lambda/n_{21} \end{bmatrix} \quad \text{with} \quad \lambda = \frac{\bar{y}_{11\cdot} - \bar{y}_{12\cdot} - \bar{y}_{21\cdot}}{\dfrac{1}{n_{11}} + \dfrac{1}{n_{12}} + \dfrac{1}{n_{21}}} .$$

And from (31)

$$\begin{aligned} \hat{\underset{\sim}{\mu}}_e &= -\hat{\mu}_{11} + \hat{\mu}_{12} + \hat{\mu}_{21} + \tau \\ &= -\bar{y}_{11\cdot} + \bar{y}_{12\cdot} + \bar{y}_{21\cdot} + \lambda(1/n_{11} + 1/n_{12} + 1/n_{21}) + \tau \\ &= \tau , \quad \text{for arbitrary } \tau . \end{aligned}$$

4.3. Example 3

Consider a 3-way crossed classification with two levels of factor A, denoted by A1 and A2, two levels B1 and B2 of factor B, and three levels of factor C, namely C1, C2 and C3. Two different cases of empty cells are dealt with, denoted by e_1 and e_2. The first is e_1: cells 111 and 121 are assumed empty, with (μ_{111}) and (μ_{121}) appearing in Grid 3. The second is e_2: cells 111 and 211 assumed empty. To show both sets of empty cells in the same grid we use $[(\mu_{111})]$ and $[\mu_{211}]$.

Grid 3

	B1			B2		
	C1	C2	C3	C1	C2	C3
A1	$[(\mu_{111})]$	μ_{112}	μ_{113}	(μ_{121})	μ_{122}	μ_{123}
A2	$[\mu_{211}]$	μ_{212}	μ_{213}	μ_{221}	μ_{222}	μ_{223}

The vector of cell means is

$$\underset{\sim}{\mu}' = [\underset{\substack{\uparrow \\ (e_1) \\ [e_2]}}{\mu_{111}}\ \mu_{112}\ \mu_{113}\ \underset{\substack{\uparrow \\ (e_1)}}{\mu_{121}}\ \mu_{122}\ \mu_{123}\ \underset{\substack{\uparrow \\ [e_2]}}{\mu_{211}}\ \mu_{212}\ \mu_{213}\ \mu_{221}\ \mu_{222}\ \mu_{223}] ,$$

underneath which the e_1's and e_2's indicate the two cases of empty cells. We now assume there are no 3-factor interactions, an assumption which can be stated as restrictions $\underset{\sim}{G}\underset{\sim}{\mu} = \underset{\sim}{0}$ for

$$\underset{\sim}{G} = \begin{bmatrix} \underset{}{1} & -1 & 0 & -1 & 1 & 0 & -1 & 1 & 0 & 1 & -1 & 0 \\ \underset{\substack{\uparrow \\ (e_1) \\ [e_2]}}{0} & 1 & -1 & \underset{\substack{\uparrow \\ (e_1)}}{0} & -1 & 1 & \underset{\substack{\uparrow \\ [e_2]}}{0} & -1 & 1 & 0 & 1 & -1 \end{bmatrix}$$

Therefore the $\underset{\sim}{G}_e$'s for the two cases of empty cells are

$$G_{e_1} = \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad G_{e_2} = \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix} .$$

Neither G_{e_1} nor G_{e_2} (which happen to be equal) have full column rank. Part (ii) of the theorem therefore does not apply.

Now consider linear combinations of μ_{ijk}'s corresponding to traditional contrasts among levels of main effects and 2-factor interactions. For example, for factor A,

$$\bar{\mu}_{1..} - \bar{\mu}_{2..} \propto \delta_A = \mu_{1..} - \mu_{2..} .$$

We then have

$$\delta_A = [1 \ \ 1 \ \ 1 \ \ 1 \ \ 1 \ \ 1 \ \ -1 \ \ -1 \ \ -1 \ \ -1 \ \ -1 \ \ -1]\mu .$$

The matrix (vector in this case) multiplying μ is, in terms of the theorem, an H-matrix. This and the H-matrices for the other contrasts, δ_B, δ_C, δ_{AB}, δ_{AC} and δ_{BC} are shown in Table 1.

To ascertain if a δ is estimable, its H_e must be identified in H for each of the 2 cases e_1 and e_2 of missing cells, and then parts (iii) and (iv) of the theorem must be applied to H_{e_1} and G_{e_1}, and to H_{e_2} and G_{e_2} to ascertain the estimability of δ in e_1 and e_2. Identification of H_{e_1} and H_{e_2} is done in Table 2, aided by the indicators e_1 and e_2 under the cell headings in Table 1.

Table 2 shows only four different H_e-matrices. In Table 3 the relationship of each of these four to

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

(the value of G_{e_1} and G_{e_2}) is shown, in terms of parts (iii) and (iv) of the theorem.

The resulting estimability of the corresponding $\underset{\sim}{\delta}$ (or $\underset{\sim\sim}{L\delta}$) is also shown in Table 3, the results of which are then applied to the δ's in Table 2, for each of the two cases e_1 and e_2 of empty cells. It is clear that even when $\underset{\sim}{G}_e$ is the same for both cases, estimability of contrasts differs from one case of empty cells to another.

4.4. Example 4

Consider example 3 with the further assumption of no AC interactions. Then from $\underset{\sim}{G}$ of example 3 and $\underset{\sim}{H}$ for δ_{AC} in Table 1, the $\underset{\sim}{G}$-matrix is now

$$\underset{\sim}{G} = \begin{bmatrix} 1 & -1 & 0 & -1 & 1 & 0 & -1 & 1 & 0 & 1 & -1 & 0 \\ 0 & 1 & -1 & 0 & -1 & 1 & 0 & -1 & 1 & 0 & 1 & -1 \\ 1 & -1 & 0 & 1 & -1 & 0 & -1 & 1 & 0 & -1 & 1 & 0 \\ 1 & 0 & -1 & 1 & 0 & -1 & -1 & 0 & 1 & -1 & 0 & 1 \end{bmatrix},$$

$$\begin{matrix} \uparrow & & & \uparrow & & & \uparrow \\ e_1 & & & e_1 & & & \\ e_2 & & & & & & e_2 \end{matrix}$$

so that

$$\underset{\sim}{G}_{e_1} = \begin{bmatrix} 1 & -1 \\ 0 & 0 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} \quad \text{and} \quad \underset{\sim}{G}_{e_2} = \begin{bmatrix} 1 & -1 \\ 0 & 0 \\ 1 & -1 \\ 1 & -1 \end{bmatrix}.$$

In this case $\underset{\sim}{G}_{e_2}$ has full column rank so that part (ii) of the theorem applies and $\widehat{\underset{\sim\sim}{H\mu}} = \underset{\sim}{H}_f\hat{\underset{\sim}{\mu}}_f$ is the b.l.u.e. of the estimable $\underset{\sim\sim}{H\mu}$ for any $\underset{\sim}{H} = [\underset{\sim}{H}_{f_1} \;\; \underset{\sim}{H}_{e_1}]$. Hence all $\underset{\sim}{\delta}$'s in Table 1 (save δ_{AC}) are estimable in the empty cell case e_1. This differs from the penultimate column of Table 3, thus illustrating that different restrictions on $\underset{\sim}{\mu}$, even with the same pattern of empty cells (case e_1), leads to different estimability conclusions.

$\underset{\sim}{G}_{e_2}$ does not have full column rank, and conclusions about estimability are the same as those in the last column of Table 3.

5. Conclusions

Estimability of $\underset{\sim}{\delta} = \underset{\sim}{H}\underset{\sim}{\mu}$ in cell means models clearly depends on $\underset{\sim}{\mu}$, and hence on the model. It also depends, just as clearly, on $\underset{\sim}{H}$. There is also dependence on both the pattern of empty cells and the restrictions (if any) on the elements of $\underset{\sim}{\mu}$. Different patterns of empty cells with the same restrictions on $\underset{\sim}{\mu}$ can lead to different conclusions regarding estimability (as in Example 3); and different restrictions on $\underset{\sim}{\mu}$ with the same pattern of empty cells can yield different estimability conclusions. (Examples 3 and 4 for the empty cell case e_1 illustrate this.)

References

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

Speed, F. M. (1969). A new approach to the analysis of linear models. Technical Memorandum, TMX-58030, National Aeronautics and Space Administration, Houston, Texas.

Speed, F. M., Hocking, R. R. and Hackney, O. P. (1978). Methods of analysis of linear models with unbalanced data. J. Amer. Stat. Assoc. 73, 105-112.

Urquhart, N. S. and Weeks, D. L. (1978). Linear models in messy data: some problems and alternatives. Biometrics 34, 696-705.

Urquhart, N. S., Weeks, D. L. and Henderson, C. R. (1973). Estimation associated with linear models: a revisitation. Communications in Statistics 1, 303-330.

<u>Table 1</u>

H-matrices for main effect and 2-factor contrasts, for Grid 2

Contrast	H-matrices											
	Cell											
	μ_{111}	μ_{112}	μ_{113}	μ_{121}	μ_{122}	μ_{123}	μ_{211}	μ_{212}	μ_{213}	μ_{221}	μ_{222}	μ_{223}
	↑ e_1 e_2			↑ e_1			↑ e_2					
δ_A	[1	1	1	1	1	1	-1	-1	-1	-1	-1	-1]
δ_B	[1	1	1	-1	-1	-1	1	1	1	-1	-1	-1]
$\underset{\sim}{\delta}_C$	[1	-1	0	1	-1	0	1	-1	0	1	-1	0
	1	0	-1	1	0	-1	1	0	-1	1	0	-1]
δ_{AB}	[1	1	1	-1	-1	-1	-1	-1	-1	1	1	1]
$\underset{\sim}{\delta}_{AC}$	[1	-1	0	1	-1	0	-1	1	0	-1	1	0
	1	0	-1	1	0	-1	-1	0	1	-1	0	1]
$\underset{\sim}{\delta}_{BC}$	[1	-1	0	-1	1	0	1	-1	0	-1	1	0
	1	0	-1	-1	0	1	1	0	-1	-1	0	1]

Table 2

$\underset{\sim}{H}_e$-matrices for contrasts in Table 1 and estimability of those contrasts

Contrast	$\underset{\sim}{H}_e$-matrices		Type of $\underset{\sim}{H}_e$ (for Table 3)		Estimability of δ (see Table 3)	
	$\underset{\sim}{H}_{e_1}$	$\underset{\sim}{H}_{e_2}$	$\underset{\sim}{H}_{e_1}$	$\underset{\sim}{H}_{e_2}$	Case e_1	Case e_2
δ_A	$\begin{bmatrix} 1 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & -1 \end{bmatrix}$	I	II	NE 1/	Est. 2/
δ_B	$\begin{bmatrix} 1 & -1 \end{bmatrix}$	$\begin{bmatrix} 1 & 1 \end{bmatrix}$	II	I	Est.	NE
$\underset{\sim}{\delta}_C$	$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$	III	III	$\underset{\sim}{L}\delta$ Est. 3/	$\underset{\sim}{L}\delta$ Est.
δ_{AB}	$\begin{bmatrix} 1 & -1 \end{bmatrix}$	$\begin{bmatrix} 1 & -1 \end{bmatrix}$	II	II	Est.	Est.
$\underset{\sim}{\delta}_{AC}$	$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix}$	III	IV	$\underset{\sim}{L}\delta$ Est.	Est.
$\underset{\sim}{\delta}_{BC}$	$\begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix}$	$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$	IV	III	Est.	$\underset{\sim}{L}\delta$ Est.

1/ NE = not estimable

2/ Est. = estimable

3/ $\underset{\sim}{L}\delta$ Est. = $[1 \;\; -1]\delta$ is estimable

Table 3

The four values of $\mathbf{H}_e$ in Table 2 and the resulting estimability conclusions

The four values of $\mathbf{H}_e$		Relationship of $\mathbf{H}_e$ with $\mathbf{G}_e = \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix}$	Application of theorem	Estimability
Type	$\mathbf{H}_e$			
I	$\begin{bmatrix} 1 & 1 \end{bmatrix}$	$\mathbf{LH}_e \neq \mathbf{MG}_e$ for all $\mathbf{L}$ and $\mathbf{M}$	No part applies	Neither $\boldsymbol{\delta}$ nor $\mathbf{L}\boldsymbol{\delta}$ is estimable
II	$\begin{bmatrix} 1 & -1 \end{bmatrix}$	$\mathbf{H}_e = [1 \;\; 1]\mathbf{G}_e$	(iii)	$\boldsymbol{\delta}$ is estimable
III	$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$	$[1 \;\; -1]\mathbf{H}_e = [0 \;\; 1]\mathbf{G}_e$	(iv)	$[1 \;\; -1]\boldsymbol{\delta}$ is estimable
III	$\begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix}$	$\mathbf{H}_e = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\mathbf{G}_e$	(iii)	$\boldsymbol{\delta}$ is estimable