MEANS MODEL WITH ADDITIONAL INFORMATION: ESTIMATION AND ANALYSIS

Oscar O. Melo MARTÍNEZ [1]
Luis A. López PÉREZ[1]
Sandra E. Melo MARTÍNEZ[2]

▪ *ABSTRACT: In this paper, original theoretical developments have been carried out to estimate parameters and residual variance as functions of initial estimates, for models of cell means of N-ways of classification with fixed effects when we have additional information into the cells. We have proven the properties of those new estimators such as unbiasedness and minimum variance. In the adjustment of these new parameters we do not require knowledge of initial information. Only for upgrading must we use the initial estimates of the parameters and the residual variance. This situation is usually present in industrial and biological experimentation. Finally, we work on the application of a factorial design in which we apply the obtained results.*

▪ *KEYWORDS: Unbalanced designs; models of cell means; additional information; connectedness.*

1 Introduction

There is a problem in studies on experimental design with unbalanced structure data in an N-way model of classification. Some of the common mistakes that researchers make are those related to the use of theory for a balanced design, which means that a hypothesis system is tested with an inappropriate system, leading to a wrong analysis of the information. Then, it is important to know what kinds of hypotheses are testable when the problem exists and determine if they are relevant to the study.

[1]Department of Statistics, Universidad Nacional de Colombia, Bogotá, D.C., Colombia. E-mail: *oomelom@unal.edu.co / lalopezp@unal.edu.co*
[2]Department of Agronomy, Universidad Nacional de Colombia, Bogotá, D.C., Colombia. E-mail: *semelom@unal.edu.co*

To find a solution in the presence of empty cells or missing data, statisticians have proposed several solutions, some of them are: to repeat the experiment under the same initial conditions in order to obtain estimations for missing data, minimizing the sum of squared errors, naming auxiliary variables or using iterative methods of estimation and analyzing the data in its original state, which implies making an analysis with empty cells.

Also, we can start with a balanced design, but the proper conditions of the experiment require that additional information be considered into all combinations of cells getting an unbalancing in the design, or at the beginning of the experiment, it may occur that information about some points for the design are missing. Besides, when the replications are made under similar conditions, information for those cells can be obtained or we can have new information in cells that were originally empty. The inclusion of this information implies the estimation of new parameter sets. As a consequence, an appropriate analysis is required.

Unbalancing can occur given the additional information into the cells, therefore, recent study carried out by López (1992), Iemma (1993), Mondardo (1994), Iemma et al. (1999), López (1999), Melo et al. (1999, 2000) is useful, since it shows the trouble that researchers find when the design has an unbalanced structure of classification and when there are empty cells. The care that must be taken in the identification of the true testing hypotheses such, as in the characterization of the estimable functions when that problem occurs, is shown.

This paper has the theoretical support of linear models, mainly the cell means model (Hocking 1996, 2002) and the modified cell means model (Murray and Smith, 1985). In the next section some basic ideas about cell means models are presented. In the third section the developments that allow the upgrade in the estimation of parameters such as residual variance are performed. The fourth section includes a numeric example to illustrate the theoretical developments presented in section three. And finally some conclusions are presented.

2 Preliminary concepts

Linear models are used for the theoretical development of a problem. Specifically, the cell means models and modified cell mean models are presented in the next subsections.

2.1 Cell means model

The cell means model is defined as the linear model (Speed, et al. 1978)

$$Y_1 = W_1\mu_1 + e_1 \tag{1}$$

where, Y_1 is a vector of random variables with size $k \times 1$, W_1 is a block matrix with size $k \times q$ where the i-th diagonal block corresponds to a column vector composed 1's with size $n_{ij...s}$, where $n_{ij...s}$ is the number of observations of the $ij...s$-th cell,

μ_1 is a vector of means with size $q \times 1$ and e_1 is a vector of non-observable random variables of size $k \times 1$ such that $e_1 \sim N(0, \sigma_1^2 I_1)$.

When there are no missing cells, W_1 is a full column rank; otherwise there are missing cells, where W_1 has a zero column for each missing cell and, therefore, its rank is not full; μ_1 is taken as if every cell would has been observed.

In this context the system has the following solution: $\hat{\mu}_1 = ({W^t}_1 W_1)^- {W^t}_1 Y_1$, which is the same as the least square estimation for μ_1. Thus, the best linear unbiased estimator (BLUE) for $\mu_{ij...s}$ is the mean of cell $ij...s$, this means that BLUE $(\mu_{ij...s}) = \bar{y}_{ij...s}$.

If the following restriction to the model (1) is imposed

$$G\mu_1 = g \tag{2}$$

Which is known as the cell means model with restriction, where G is a matrix composed of unknown contrasts with size $s \times q$ and rank s.

2.2 The modified cell means model

If $g = 0$ in (2), then the cell means model is characterized by Hocking (1996, 2002), Speed, et al. (1978) as the model (1) with the restriction

$$G\mu_1 = 0 \tag{3}$$

where G was specified above and represents the known linear relationships over the cell means. Generally it specifies contrasts of no-interaction in the cell means model as we can see in Hocking (1996, 2002), Searle (1987) and Murray and Smith (1985).

The constraint matrix of G can be reorganized and G can be reordered and partitioned into two submatrices, which means $G = [G_1|G_2]$, where G_2 is $s \times s$ of rank s and G is $s \times (q-s)$. The partition of μ_1 is independent of the data obtained; in particular it is independent of the number of missing cells (for example f) and their location. The partition depends on how the experiment was planned and not how it was done. Also, it depends on G and the relations among the cell means. Hance, the columns that compose μ_1 can be reorganized in agreement with the partitioning of G, giving

$$\mu_1^t = \begin{pmatrix} m_1^t & \vdots & m_2^t \end{pmatrix} \tag{4}$$

and hence (2) can be written as

$$G\mu_1 = G_1 m_1 + G_2 m_2 = g \tag{5}$$

The choice of submatrix G_2 is arbitrary and when it is nonsingular, the only solution for m_2 in terms of m_1 exists and is given by

$$m_2 = G_2^{-1}(g - G_1 m_1) \tag{6}$$

This partitioning is now used to rewrite the cell means model (1). The columns of incidence matrix w must be reorganized in agreement with the reordering of G and μ_1, giving

$$Y_1 = \begin{pmatrix} \omega_1 & \vdots & \omega_2 \end{pmatrix} \begin{pmatrix} m_1 \\ m_2 \end{pmatrix} + e_1 \tag{7}$$

and substituting (6) in (7), the model (1) can be written as

$$Y_1 - \omega_2 G_2^{-1} g = (\omega_1 - \omega_2 G_2^{-1} G_1) m_1 + e_1 \tag{8}$$

Making $Y_1^* = Y_1 - \omega_2 G_2^{-1} g$ and $V = (\omega_1 - \omega_2 G_2^{-1} G_1)$ of size $k \times (q-s)$, we get the modified unconstrained model by Murray and Smith (1985)

$$Y_1^* = V m_1 + e_1 \tag{9}$$

If the rank of V is $(q-s)$, then $V^t V$ is non singular and we can apply the usual method of least squares for unconstrained full rank models, obtaining the BLUE for m_1 as

$$\hat{m}_1 = (V^t V)^{-1} V^t Y_1^* \tag{10}$$

Substituting (10) in (6) we find the BLUE for m_2,

$$\hat{m}_2 = G_2^{-1}[g - G_1 (V^t V)^{-1} V^t Y_1^*] \tag{11}$$

If there are non empty cells, then the rank of V is $(q-s)$. If there are empty cells, V may still have rank $(q-s)$, in which case (10) and (11) are the only BLUE's for m_1 and m_2. If the rank of V is less than $(q-s)$, then we are in the same problem as that in the overparametric model (see Searle 1971).

Based on the model (9), when we impose non-interaction restrictions, we establish the possibility to connect empty cells with information from non-empty cells getting the base for the estimable functions in an easy way. Due the importance of these two concepts a synthesis of estimability and connectedness is presented in section 2.2.1.

2.2.1 Estimability and connectedness

In this section, we relate the estimability of cell means to the concept of connectedness to develop a simple test for estimability of μ_1 based on the modified μ_1-model (9).

The term connectedness initially referred to the physical location of the filled cells in the two-way array associated with the randomized block design (John, 1974). In Searle (1987), the simple idea is that the design is connected, if possible, to join cells in which we alternately step in the same row (treatment) or to the same column (block). Thus, treatments i and r in the cells (i, j) and (r, t) may be connected by the path.

$$(i,j) \rightarrow (i,v) \rightarrow (u,v) \rightarrow (u,t) \rightarrow (r,t)$$

Hocking (1996, 2002) and Searle (1971, 1987) give us a definition of connectedness for two-way classification models without interaction, as generalized by Weeks and Williams (1964), who presented a proposal for studying connectedness in linear models for N-way crossed classification without interaction.

Recognizing that the goal of connectedness is estimability and that this includes both the design and model leads us to the following very general definition for the cell means model:

Definition 1. An experiment conformed by a data set and associated with a cell means model is connected if μ_1 is linearly estimable in a unique way.

A particular data set can be connected to a model and not to another. For example, in a model of classification with three factors, several levels of connectedness are presented, and this depends on the necessary restriction for getting estimability of μ_1. The model without restriction requires that $n_{ijk} \geq 0$. The non interaction condition can be enough to allow the estimation of all μ_{ijk} in the case of missing cells. If this is not enough, then the imposition of one or more restrictions of non interaction with a smaller dimension than the factors can lead to the estimability.

The search for estimability of the parameters does not justify the supposition in the restrictions. This assumption should be made before collecting the information, based on a-priori information regarding the relationships among the cell means. That is, we are interested in making the analysis as it was initially conceived and not an analysis based on some conveniences dictated by the data.

Based on the modified cell means model and on definition 1, Murray and Smith (1985) presented an approach for the study of connectedness, which is summarized in the following theorem:

Theorem 1. For model (1) and (2), the experiment is connected if and only if V has full column rank, which means that the rank of V is $q - s \leq k$.

Proof. You can see a proof in Murray and Smith (1985).

If V is a full column rank, then the experiment is connected, μ_1 is lineally estimable in its entirety and the original analysis can be carried out as planned. If V does not have full column rank, then the experiment is not connected, μ_1 is not linealy estimable and the original analysis cannot be carried out. In this case, the cell means model (1) and (2) are not full rank.

In many situations, some of those non observed means can be estimated using restriction (3) and the observed means. In other situations, only linear combinations of the non observed means can be estimated.

3 Upgrade of parameters and residual variance

Some experimental designs can be balanced designs, but due to the characteristic conditions of the experiment, additional information is included in

some or all of the combinations of cells, bearing this last to an unbalancing in the design.

In the same way, it may happen that at the beginning of the experiment information that is not available in some points of the design can be completed when those additional replications are carried out under similar conditions but at different times.

The inclusion of this additional information in the design gives the estimate a new set of parameters and, in consequence, the problem should be studied by keeping in mind this additional information.

In this section, it is assumed that additional information arrives to any of the cells, that is to say, it may happen that there is additional information in cells where there was already information or in other cells where information was not reported initially. As the objective of this paper is the upgrade of parameters estimation and the residual variance, it is important to consider the following partition of the model,

$$\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} = \begin{pmatrix} W_1 \\ W_2 \\ W_3 \end{pmatrix} \mu + \begin{pmatrix} e_1 \\ e_2 \\ e_3 \end{pmatrix} \tag{12}$$

where

$$\begin{pmatrix} e_1 \\ e_2 \\ e_3 \end{pmatrix} \sim N\left(\begin{pmatrix} 0_1 \\ 0_2 \\ 0_3 \end{pmatrix}; \begin{pmatrix} \sigma_1^2 I_1 & 0 & 0 \\ 0 & \sigma_2^2 I_2 & 0 \\ 0 & 0 & \sigma_2^2 I_3 \end{pmatrix} \right)$$

and Y_1 of size $k \times 1$ is a vector of random variables corresponding to the initial information, Y_2 of size $t \times 1$ is a vector of random variables with respect to the additional information reaching the same cells where information already existed, Y_3 of size $(n-k-t)\times 1$ is a vector of random variables with respect to the additional information that arrives to the cells that initially did not have any information. $\sigma_1^2 I_1$ is a matrix of size $k \times k$ containing variances and covariances of the random variables corresponding to the initial information, $\sigma_2^2 I_2$ of size $t \times t$ is a matrix of variances and covariances of the variables corresponding to the additional information that arrives to the same cells where information already existed, $\sigma_2^2 I_3$ of size $(n-k-t)\times(n-k-t)$ is a matrix of variances and covariances of the random variables corresponding to the additional information that reaches cells that initially had not information, W_1, W_2, W_3 are matrices of blocks with the i-th diagonal block corresponding to a vector column of 1's of size $n_{ij...s}$, where $n_{ij...s}$ is the number of observations of the $ij...s$-th population for the initial information as well as for the additional information, and μ is a vector of population means.

3.1 Upgrade of the parameters estimate with additional information

In (12) it is important to carry out the estimate of the parameters and the residual variance. Therefore, we begin with the logarithm of the likelihood function, that is,

$$
\begin{aligned}
-2\ln\big[L(\mu,\sigma_1^2,\sigma_2^2/Y)\big] = \quad & n\ln(2\pi)+\ln\big[(\sigma_1^2)^k(\sigma_2^2)^{n-k}\big]+\tfrac{1}{\sigma_1^2}(Y_1^t-\mu^tW_1^t)\\
& (Y_1-\mu W_1)+\tfrac{1}{\sigma_2^2}\Big[(Y_2^t-\mu^tW_2^t)(Y_2-\mu W_2)\\
& +(Y_3^t-\mu^tW_3^t)(Y_3-\mu W_3)\Big]
\end{aligned}
\tag{13}
$$

Giving a solution for (13) and being consistent with the proposal, we suppose that the variance of the initial information is the same as the variance for the additional information, so we have the normal equations.

$$
(W_1^tW_1+W_2^tW_2+W_3^tW_3)\mu = W_1^tY_1+W_2^tY_2+W_3^tY_3 \tag{14}
$$

or equivalently, the model can be written as

$$
E\begin{pmatrix} Y_1\\ Y_2\\ Y_3\end{pmatrix} = \begin{pmatrix} W_1\\ W_2\\ W_3\end{pmatrix}\mu \tag{15}
$$

In a similar form to the developments made in section 2.2, we have

$$
\mu^t = (m_1^{*t}\vdots m_2^{*t}) \tag{16}
$$

and in this way (2) can be rewritten as

$$
G\mu = G_1m_1^*+G_2m_2^* = g \tag{17}
$$

An only solution for m_2^* in terms of m_1^* exists and it is given by

$$
m_2^* = G_2^{-1}(g-G_1m_1^*) \tag{18}
$$

Keeping in mind (16), (17) and (18), and reordering the columns of the incidence matrices appropriately, then (15) can be rewritten as

$$
\begin{pmatrix} Y_1\\ Y_2\\ Y_3\end{pmatrix} = \begin{pmatrix} \omega_1 & \vdots & \omega_2\\ U_1 & \vdots & U_2\\ S_1 & \vdots & S_2\end{pmatrix}\begin{pmatrix} m_1^*\\ m_2^*\end{pmatrix}+\begin{pmatrix} e_1\\ e_2\\ e_3\end{pmatrix} \tag{19}
$$

Substituting (18) for (19), we see that

$$
\begin{pmatrix} Y_1\\ Y_2\\ Y_3\end{pmatrix} = \begin{pmatrix} \omega_2\\ U_2\\ S_2\end{pmatrix}G_2^{-1}g+\begin{pmatrix} \omega_1-\omega_2G_2^{-1}G_1\\ U_1-U_2G_2^{-1}G_1\\ S_1-S_2G_2^{-1}G_1\end{pmatrix}m_1^*+\begin{pmatrix} e_1\\ e_2\\ e_3\end{pmatrix} \tag{20}
$$

Making $Y_1^* = Y_1 - \omega_2 G_2^{-1} g$, $Y_2^* = Y_2 - U_2 G_2^{-1} g$, $Y_3^* = Y_3 - S_2 G_2^{-1} g$, $V = \omega_1 - \omega_2 G_2^{-1} G_1$, $U = U_1 - U_2 G_2^{-1} G_1$, y $S = S_1 - S_2 G_2^{-1} G_1$. Then (20) is written as

$$\begin{pmatrix} Y_1^* \\ Y_2^* \\ Y_3^* \end{pmatrix} = \begin{pmatrix} V \\ U \\ S \end{pmatrix} m_1^* + \begin{pmatrix} e_1 \\ e_2 \\ e_3 \end{pmatrix} \tag{21}$$

In (21), if some empty cells are connected, then some of the cells where there is the additional information corresponding to the empty cells can be written as linear combination of the connected cells, which means that some rows from S are linear combination of the rows from V. In this way, if Y_3^* and S are partitioned in connected and unconnected information, then from the additional information it is obtained that

$$\begin{pmatrix} Y_2^* \\ Y_3^* \end{pmatrix} = \begin{pmatrix} Y_2^* \\ Y_C \\ Y_F \end{pmatrix} = \begin{pmatrix} Y_M \\ Y_F \end{pmatrix} \text{ and } \begin{pmatrix} U \\ S \end{pmatrix} = \begin{pmatrix} U \\ C \\ F \end{pmatrix} = \begin{pmatrix} M \\ F \end{pmatrix} \tag{22}$$

where Y_C corresponds to the vector of observations of the connected cells, Y_F corresponds to the vector of observations of the non-connected cells, C is the matrix of indexes associated with the connected cells and F is the matrix of indexes associated to the non-connected cells.

Substituting (22) in (21) the model can be written as

$$\begin{pmatrix} Y_1^* \\ Y_M \\ Y_F \end{pmatrix} = \begin{pmatrix} V \\ M \\ F \end{pmatrix} m_1^* + \begin{pmatrix} e_1 \\ e_M \\ e_F \end{pmatrix} \tag{23}$$

In order to find the solution of the group of parameters m_1^*, it is necessary to consider lemma 1, which will be used in the demonstration of theorem 2.

Lemma 1. For $T = \begin{pmatrix} T_1 \\ T_2 \end{pmatrix}$ with $f(T_2) \subset f(T_1)$ (it means that the row space of T_2 is contained in the row space of T_1) and $rank(T) = rank(T_1)$, a generalized inverse of $T^t T = T_1^t T_1 + T_2^t T_2$ is given by

$$(T_1^t T_1 + T_2^t T_2)^- = (T_1^t T_1)^- - (T_1^t T_1)^- T_2^t [I + T_2 (T_1^t T_1)^- T_2^t]^{-1} T_2 (T_1^t T_1^-)^- \tag{24}$$

with $(T_1^t T_1)(T_1^t T_1)^- (T_1^t T_1) = T_1^t T_1$.

Proof. See López and Rincón (1999).

The following theorem gives us a solution for m_1^* in (23) and in this way, we find a solution for the estimation of parameters μ, when we have additional information in the cells. In the solution, we had to keep only the initial estimations for updating the parameters.

Theorem 2. When there is additional information in the cells of model (15) under restriction (2), the upgrade in the estimates of the parameters is given by

$$\widehat{\mu} = \begin{pmatrix} \widehat{m}_1^* & \vdots & \widehat{m}_2^* \end{pmatrix} \tag{25}$$

Where,

$$\begin{aligned} \hat{m}_1^* &= \{I - (V^tV + F^tF)^- M^t R^{-1} M\}(V^tV + F^tF)^-[(V^tV)\hat{m}_1 \\ &\quad + M^tY_M + F^tY_F] \end{aligned} \tag{26}$$

$$\hat{m}_2^* = G_2^{-1}(g - G_1\hat{m}_1^*) \tag{27}$$

with,

$$R = I_M + M(V^tV + F^tF)^- M^t \tag{28}$$

Proof. Working with model (15) under restriction $G\mu = g$, we find equation (23) following a similar procedure to that in section 2.2. In this way, the maximum likelihood solution for m_1^* in equation (23) is given by

$$\widehat{m}_1^* = \left(V^tV + M^tM + F^tF\right)^- \left(V^tY_1^* + M^tY_M + F^tY_F\right) \tag{29}$$

According to the lemma 1, as $W = \begin{pmatrix} V \\ M \\ F \end{pmatrix}$ with $f(M) \subset f\begin{pmatrix} V \\ F \end{pmatrix}$ and $rank(W) = rank(V) + rank(F)$, a generalized inverse of $V^tV + M^tM + F^tF$ is given by

$$\begin{aligned} (V^tV + M^tM + F^tF)^- = \; & (V^tV + F^tF)^- - (V^tV + F^tF)^- M^t[I_M \\ & + M(V^tV + F^tF)^- M^t]^{-1} M(V^tV + F^tF)^- \end{aligned} \tag{30}$$

or equivalently

$$(V^tV + M^tM + F^tF)^- = I - (V^tV + F^tF)^- M^t R^{-1} M(V^tV + F^tF)^- = K \tag{31}$$

With R as in (28). Substituting (31) in (29), it is obtained that

$$\hat{m}_1^* = K(V^tY_1^* + M^tY_M + F^tY_F) \tag{32}$$

and like

$$V^tY_1^* = (V^tV)(V^tV)^- V^tY_1^* = (V^tV)\hat{m}_1 \tag{33}$$

With $\hat{m}_1$ the solution obtained in (10) should simply be considered an inverse generalized matrix instead of an inverse of a full rank matrix. In this way, replacing

(33) in (32) we get (26) and substituting this result in (18), the estimator for m_2 is obtained as it was given in (27).

Finally, from (26) and (27) we have (25), completing the proof.

Explicit forms to upgrade the estimate of the parameters when additional information arrives to the experiment are presented in the following corollaries.

Corollary 1. If in model (15), the additional information only comes from the cells that can be connected, when imposing restriction (2), then the upgrade in the estimate of parameters μ, in terms of initial parameter estimates is given as in (25), where for this case

$$\hat{m}_1^* = \left\{ I_1 - (V^tV)^- M^t \left[I_M + M(V^tV)^{-1} M^t \right]^{-1} M \right\} \left[\hat{m}_1 + (V^tV)^- M^t Y_M \right] \quad (34)$$

and $\hat{m}_2^*$ is like in (27).

Proof. If the additional information only comes from the cells that can be connected when imposing the restriction (2), then in expression (23) $F = 0$, and therefore (26) in theorem 2, can be expressed as

$$\hat{m}_1^* = \{ I - (V^tV)^- M^t (R^*)^{-1} M \} (V^tV)^- \left[(V^tV)\hat{m}_1 + M^t Y_M \right] \quad (35)$$

where

$$R^* = I_M + M(V^tV)^- M^t \quad (36)$$

Corollary 2. If model (15) is not subject to restriction (2), then the upgrade in the parameter estimated μ in terms of the initial parameter estimates is given by

$$\begin{aligned} \widehat{m}_1^* = \widehat{\mu} = \quad & \left\{ I_1 - (W_1^tW_1)^- W_2^t \Pi^{-1} W_2 \right\} \left[\widehat{m}_1 + (W_1^tW_1)^- W_2^t Y_2 \right] \\ & + (W_3^tW_3)^- W_3^t Y_3 \end{aligned} \quad (37)$$

where

$$\Pi = I_2 + W_2 \left(W_1^tW_1 \right)^- W_2^t \quad (38)$$

Proof. The solution by the method of maximum likelihood for model (23) is given by (26). When the model is unconstrained, then $G = 0$ and $g = 0$, therefore, $V = W_1$, $M = W_2$ and $F = W_3$. Thus (26) can be written as

$$\begin{aligned} \widehat{m}_1^* = \quad & \left\{ I_1 + (W_1^tW_1 + W_3^tW_3)^- W_2^t \Pi^{-1} W_2 \right\} (W_1^tW_1 + W_3^tW_3)^- \\ & \left[(W_1^tW_1)\, \widehat{m}_1 + W_2^tY_2 + W_3^tY_3 \right] \end{aligned} \quad (39)$$

with

$$\Pi = I_2 + W_2 \left(W_1^tW_1 + W_3^tW_3 \right)^- W_3^t \quad (40)$$

This corresponds to the estimate of μ, since non restriction has been imposed. Also, $\widehat{m}_1 = (W_1^t W_1)^- W_1^t Y_1$.

In this case, W_1 and W_3 are orthogonal (that is, $W_1^t W_3 = 0$). Also as W_1 comes from the matrix of initial information and W_3 from the matrix of additional information that arrives to the cells where initially there was no information, then

$$\left(W_1^t W_1 + W_3^t W_3\right)^- = \left(W_1^t W_1\right)^- + \left(W_3^t W_3\right)^- \tag{41}$$

In addition, W_2 and W_3 are orthogonal too (that is, $W_2^t W_3 = 0$) because W_2 comes from the matrix of additional information that arrives to the cells where initially there was information. With these results and after replacing (41) and carrying out the different products in the expression (39) we obtain

$$\begin{array}{rl} \widehat{m}_1^* = \widehat{\mu} = & \left[I_1 - (W_1^t W_1)^- W_2^t \Pi^{-1} W_2\right] (\widehat{m}_1 + (W_1^t W_1)^- W_2^t Y_2) \\ & + (W_3^t W_3)^- W_3^t Y_3 \end{array} \tag{42}$$

and after factoring the matrix in (42), we obtain (37), completing the proof.

Also, it can be proven that $\pi^t \widehat{\mu}$, with $\widehat{\mu}$ obtained as in (25), is a BLUE, that is, $BLUE(\pi^t \mu) = \pi^t \widehat{\mu}$.

The expected value of the previous linear combination is

$$E(\pi^t \widehat{\mu}) = \pi^t \mu \tag{43}$$

and its variance and covariance matrix is given by

$$V\left(\pi^t \widehat{\mu}\right) = \pi^t \begin{pmatrix} I_1 \\ -G_2^{-1} G_1 \end{pmatrix} V\left(\widehat{m}_1^*\right) \begin{pmatrix} I_1 & \vdots & -G_1^t (G_2^{-1})^t \end{pmatrix} \pi \tag{44}$$

where

$$V\left(\widehat{m}_1^*\right) = K\sigma^2 \tag{45}$$

In the following section, the estimate of the residual variance is upgraded when there is additional information in the experiment.

3.2 Upgrade the estimation of the residual variance with additional information

Similarly to the parameter estimates, in this section, an attempt is made to upgrade the estimate of the residual variance when there is additional information, without necessity of having knowledge about the initial information. The results of this estimate are summarized in theorem 3.

Theorem 3. When there is additional information in the cell means model (15) under restriction (2), the upgrade in the estimate of the residual variance in terms of initial estimate of the residual variance is given by

$$\begin{array}{rl}\hat{\sigma}^2 = & \frac{1}{(n-r)}(k-r_v)\hat{\sigma}_1^2 + SSM_1 - [(V^tV)\hat{m}_1 + F^tY_F + M^tY_M]^tK \\ & [(V^tV)\hat{m}_1 + F^tY_F + M^tY_M] + Y_M^tY_M + Y_F^tY_F\end{array} \tag{46}$$

Where K is obtained as in (31), $r = r_V + r_F$, r_F and r_V are, respectively, the rank of the matrices V and F,

$$\hat{\sigma}_1^2 = \frac{1}{k-r_V}Y_1^{*t}[I_1 - V(V^tV)^-V^t]Y_1^* \tag{47}$$

and

$$SSM_1 = Y_1^{*t}V(V^tV)^-V^tY_1^* \tag{48}$$

Proof. When we work with model (15) and imposing restriction $G\mu = g$, we obtain the equation (23) following a procedure similar to the one of section 2.2. In this way, the estimate of the residual variance is upgraded. In this model, the maximum likelihood solution for σ^2 is

$$\begin{array}{l}(n-r)\hat{\sigma}^2 = (Y_1^{*t}Y_M^tY_F^t) \\ \qquad\qquad \left[I - \begin{pmatrix}V\\M\\F\end{pmatrix}(V^tV + M^tM + F^tF)^- \begin{pmatrix}V^t & M^t & F^t\end{pmatrix}\right]\begin{pmatrix}Y_1^*\\Y_M\\Y_F\end{pmatrix}\end{array} \tag{49}$$

Using lemma 1 and carrying out the respective products in (49), we obtain

$$\begin{array}{rl}(n-r)\hat{\sigma}^2 = & (Y_1^{*t}[I_1 - VKV^t]Y_1^* + Y_M^t(I_M - MKM^t)Y_M \\ & +Y_F^t(I_F - FKF^t)Y_F - 2Y_1^{*t}VKM^tY_M \\ & -2Y_1^{*t}VKF^tY_F - Y_M^tMKF^tY_F\end{array} \tag{50}$$

Adding and subtracting $Y_1^{*t}V(V^t)^-V^tY_1^*$ in (50), the variance can be rewritten as

$$\begin{array}{rl}(n-r)\hat{\sigma}^2 = & (Y_1^{*t}[I_1 - V(V^tV)^-V^t]Y_1^{*t}[V(V^tV)^-V^t + VKV^t]Y_1^* \\ & -2Y_1^{*t}VKF^tY_FY_M^t(I_M - MKM^t)Y_M - Y_M^tMKF^tY_F \\ & +Y_F^t(I_F - FKF^t)Y_F - 2Y_1^{*t}VKM^tY_M\end{array} \tag{51}$$

Substituting (10), (33), (47) and (48) into (51), we see that

$$\begin{array}{rl}(n-r)\hat{\sigma}^2 = & (k-r_V)\hat{\sigma}_1^2 + SSM_1 - \hat{m}_1^t(V^tV)K(V^tV)\hat{m}_1 + Y_M^t(I_M \\ & -MKM^t)Y_M + Y_F^t(I_F + FKF^t)Y_F - 2\hat{m}_1^t(V^tV)KF^tY_F \\ & -2\hat{m}_1^t(V^tV)KM^tY_M - Y_M^tMKF^tY_F\end{array} \tag{52}$$

When the appropriate factoring is carried out and dividing both sides of the equation (52) by $(n-r)$ we have (46), which completes the proof.

In corollary 3 and 4 two expressions for variances are presented, when additional information comes to cells that can be connected and when no restriction is imposed in the experiment.

Corollary 3. If in model (15), the additional information reaches to the connected cells under restriction (2), then the upgrade in the estimate of residual variance σ^2 in terms of the initial estimate of residual variance σ_1^2 is given by

$$\hat{\sigma}^2 = \frac{1}{n-r}\Big\{(k-r_V)\hat{\sigma}_1^2 + (Y_M - M\hat{m}_1)^t (R^*)^{-1}(Y_M - M\hat{m}_1)\Big\} \tag{53}$$

where; $\hat{\sigma}_1^2$ and R^* are such as in (49) and (36), respectively, and $r = r_V$.

Proof. If the additional information only reaches to the cells that can be connected under restriction (2), then in expression (23) $F = 0$ and $Y_F = 0$; therefore, (46) in theorem 3 can be expressed as

$$\begin{aligned} \widehat{\sigma}^2 = \quad & \frac{1}{n-r}\Big\{(k-r_V)\widehat{\sigma}_1^2 + SSM_1 - \left[(V^tV)\widehat{m}_1 + M^tY_M\right]^t K_1 \\ & \left[(V^tV)\widehat{m}_1 + M^tY_M\right] + Y_M^tY_M\Big\} \end{aligned} \tag{54}$$

where $K_1 = \left[I - (V^tV)^- M^t (R^*)^{-1} M\right](V^tV)^-$ and R^* is the same as in (36). Using (54) and carrying out the respective products, we come to

$$\begin{aligned} (n-r)\widehat{\sigma}^2 \quad = \quad & (k-r_V)\widehat{\sigma}_1^2 + SSM_1 - \widehat{m}_1^t(V^tV)K_1(V^tV)\widehat{m}_1 + Y_M^t \\ & \left[I_M - MK_1M^t\right]Y_M - 2\widehat{m}_1^t(V^tV)K_1M^tY_M \end{aligned} \tag{55}$$

Substituting the value of K_1, applying properties of generalized inverse and carrying out the respective products into (55), we obtain that

$$\begin{aligned} (n-r)\widehat{\sigma}^2 \quad = \quad & (k-r_V)\widehat{\sigma}_1^2 + SSM_1 - \widehat{m}_1^t(V^tV)\widehat{m}_1 + \widehat{m}_1^tM^t(R^*)^{-1}M\widehat{m}_1 \\ & + Y_M^tY_M - \left[Y_M^tM(V^tV)^-M^t(R^*)^{-1} + 2\widehat{m}_1^tM^t(R^*)^{-1}\right] \\ & \left[R^* - M(V^tV)^-M^t\right]Y_M \end{aligned} \tag{56}$$

As $SSM_1 = \widehat{m}_1^t(V^tV)\widehat{m}_1$ and substituting (36) into (56), we see that

$$\begin{aligned} (n-r)\widehat{\sigma}^2 \quad = \quad & (k-r_V)\widehat{\sigma}_1^2 + (M\widehat{m}_1)^t(R^*)^{-1}(M\widehat{m}_1) - 2(M\widehat{m}_1)^t(R^*)^{-1}Y_M \\ & + Y_M^t(R^*)^{-1}Y_M \end{aligned} \tag{57}$$

The proof is completed carrying out the factoring correspondents and dividing both sides of the equation (57) by $(n-r)$. In this way, we come to (53).

Corollary 4. If model (15) is unconstrained, then the upgrade in the estimate of residual variance σ^2 in terms of the initial residual variance estimate, σ_1^2, is given by

$$\widehat{\sigma}^2 = \frac{1}{n-r}\Big\{(k-r_V)\widehat{\sigma}_1^2 + (Y_2 - W_2\widehat{m}_1)^t\Pi^{-1}(Y_2 - W_2\widehat{m}_1) + Y_3^t \\ [I_3 - W_3(W_3^tW_3)^-W_3^t]\,Y_3\Big\} \tag{58}$$

where $\widehat{\sigma}_1^2$ is such as (47), but replacing V by W_1, Π is given by (40) and $r = r_{W_1} + r_{W_3}$.

Proof. If model (15) is unconstrained, then $G = 0$ and $g = 0$, with this be satisfied that $V = W_1$, $M = W_2$ and $F = W_3$. In this way, the solution of maximum likelihood for σ^2 presented in equation (46) can be written as

$$\widehat{\sigma}^2 = \frac{1}{n-r}\Big\{(k-r_{W_1})\widehat{\sigma}_1^2 + SSM_1 - \left[\left(W_1^tW_1\right)\widehat{m}_1 + W_3^tY_3 + W_2^tY_2\right]^t K_2 \\ \left[\left(W_1^tW_1\right)\widehat{m}_1 + W_3^tY_3 + W_2^tY_2\right] + Y_2^tY_2 + Y_3^tY_3 \tag{59}$$

where K_2 is obtained by replacing (38) into (31), this is

$$K_2 = \left(W_1^tW_1\right)^- + \left(W_3^tW_3\right)^- - \left(W_1^tW_1\right)^- W_2^t\Pi^{-1}W_2\left(W_1^tW_1\right)^-$$

and Π such as (38). Solving the respective products in (59)), we obtain

$$(n-r)\widehat{\sigma}^2 = (k-r_{W_1})\,\widehat{\sigma}_1^2 + SSM_1 - \widehat{m}_1^t\left(W_1^tW_1\right)K_2\left(W_1^tW_1\right)\widehat{m}_1 + Y_2^t \\ \left[I_2 - W_2K_2W_2^t\right]Y_2 + Y_3^t\left[I_3 - W_3K_2W_3^t\right]Y_3 - 2\widehat{m}_1^t\left(W_1^tW_1\right) \\ K_2\left(W_2^tY_2 + W_3^tY_3\right) - 2Y_2^tW_2K_2W_3^tY_3 \tag{60}$$

As W_3 is orthogonal to W_1 and W_2 given the justification of corollary 2, applying properties of generalized inverse and carrying out the respective products in (60), we come to

$$(n-r)\widehat{\sigma}^2 = (k-r_{W_1})\,\widehat{\sigma}_1^2 + SSM_1 - \widehat{m}_1^t\left(W_1^tW_1\right)\widehat{m}_1 + \widehat{m}_1^tW_2^t\Pi^{-1}W_2\widehat{m}_1 \\ +Y_2^tW_2\left(W_1^tW_1\right)^-W_2^t\Pi^{-1}W_2\left(W_1^tW_1\right)^-W_2^tY_2 + Y_3^t[I_3 - W_3 \\ \left(W_3^tW_3\right)^-W_3^t]Y_3 - 2\widehat{m}_1^tW_2^tY_2 + 2\widehat{m}_1^tW_2^t\Pi^{-1}W_2\left(W_1^tW_1\right)^- \\ W_2^tY_2 + Y_2^t[I_2 - W_2\left(W_1^tW_1\right)^-W_2^t]Y_2 \tag{61}$$

Replacing $SSM_1 = \widehat{m}_1^t\left(W_1^tW_1\right)\widehat{m}_1$ and factoring (61), it is found that

$$(n-r)\widehat{\sigma}^2 = (k-r_{W_1})\,\widehat{\sigma}_1^2 + \widehat{m}_1^tW_2^t\Pi^{-1}W_2\widehat{m}_1 + Y_2^tY_2 + Y_3^t[I_3 - W_3 \\ \left(W_3^tW_3\right)^-W_3^t]Y_3 - [Y_2^tW_2\left(W_1^tW_1\right)^- + 2\widehat{m}_1^t]W_2^t\Pi^{-1} \\ [\Pi - W_2\left(W_1^tW_1\right)^-W_2^t]Y_2 \tag{62}$$

Substituting (38) into (62) and factoring, we obtain

$$\begin{aligned}(n-r)\widehat{\sigma}^2 &= (k-r_{W_1})\,\widehat{\sigma}_1^2 + (W_2\widehat{m}_1)^t\,\Pi^{-1}\,(W_2\widehat{m}_1) - 2\,(W_2\widehat{m}_1)^t\,\Pi^{-1}Y_2 \\ &\quad + Y_2^t\Pi^{-1}Y_2 + Y_3^t\left[I_3 - W_3\left(W_3^tW_3\right)^-W_3^t\right]Y_3 \end{aligned} \tag{63}$$

Dividing both sides by $(n-r)$ into (63) and carrying out the appropriate factors, we finally come to (58), which completes the proof.

Finally, it can be proven that estimator $\widehat{\sigma}^2$ (obtained as in theorem 3) is a BLUE.

In this case, we find that

$$E(\widehat{\sigma}^2) = \sigma^2 \tag{64}$$

and the variance of the variance estimator is

$$V\left(\widehat{\sigma}^2\right) = \frac{2\sigma^2}{n-r} \quad ; \quad n > r \tag{65}$$

4 Application

The theoretical results presented in the previous sections are illustrated in the following asymmetric factorial arrangement 2x3x4. The data groups were taken from Myers and Montgomery (1995). In the experiment, concerning the factors that influence the terminal surface of a metallic particle, these factors were: corrosion rate (inch/min) (A), court depth (inches) (B) and material type (C).

We assume that the structure of the data with initial information in the experiment is presented in Table 1. Studies in the area had outlined, prior to the development our study, that the interaction between the three involved factors is not possible.

Table 1 - Initial arrangement of information in the factorial design $2x3x4$

Corrosion	Material I				Material II			
	Cut depth				Cut depth			
rate	0.15	0.20	0.30	0.40	0.15	0.20	0.30	0.40
0.20	74	79			63		77	101
	78				68		79	
0.25	98		98	105		85	83	
	91			102		81	87	
0.30		115		133	100			118
				138				122

The model associated with the factorial design without interaction of third order is given by

$$y_{ijkl} = \mu_{ijk} + e_{ijkl} \tag{66}$$

with $i = 1, 2$, $j = 1, 2, 3$, $k = 1, 2, 3, 4$ y $l = 1, 2, \ldots, n_{ijk}$ and where y_{ijkl} is the l-th observation related to the i-th material, j-th corrosion rate and k-th cut depth; μ_{ijk} is the ijk-th cell mean and e_{ijkl} is the random error component such that $e_{ijkl} \sim N(0, \sigma^2)$.

For model (66), the restrictions of non interaction between factors ABC, are given by

$$\{\mu_{ijk} - \mu_{ij'k} - \mu_{ijk'} + \mu_{ij'k'}\} - \{\mu_{i'jk} - \mu_{i'j'k} - \mu_{i'jk'} + \mu_{i'j'k'}\} = 0; \tag{67}$$

for all i, j and k, with $i \neq i', j \neq j', k \neq k'$.

Of the abc possible contrasts in (67), there are only $(a-1)(b-1)(c-1)$ contrasts linearly independent for interaction ABC. Such a group of linearly independents restrictions is obtained by fixing $i' = a = 2, j' = b = 3$ y $k' = c = 4$, in this way,

$$\mu_{1jk} - \mu_{13k} - \mu_{1j4} + \mu_{134} - \mu_{2jk} - \mu_{23k} - \mu_{2j4} + \mu_{234} = 0 \tag{68}$$

with $j = 1, 2$ and $k = 1, 2, 3$. Matrix G for contrasts (68) is then

$$G\mu_1 = (D_2 \otimes D_3 \otimes D_4)\mu_1 = 0 \tag{69}$$

where $D_2 = \begin{pmatrix} 1 & -1 \end{pmatrix}$, $D_3 = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix}$ and $D_4 = \begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{pmatrix}$.

Therefore, matrix G is 6×24 of rank 6. Carrying out a partition of matrix G into two submatrices G_2 of 6×6 and G_1 of 6×18, where G_2 is non singular and its columns correspond to cell means [111, 112, 113, 121, 122, 123]. The columns of G_1 correspond to the remain cell means, which are [114, 124, 131, 132, 133, 134, 211, 212, 213, 214,221, 222, 223, 224, 231, 232, 233, 234].

Similarly, the rows of μ_1 and the columns of W_1 are reordered according to the partition of G.

In this case, matrix V does not have full rank ($r_V = 14$). Then considering the restriction of non interaction of second order among three factors (ABC), the previous factorial arrangement is not connected. To connect the arrangement, it becomes necessary to impose another restriction to the design, for example some restriction of non double interaction. However, for effects of the study any additional restriction was not imposed. In consequence, we developed an analysis with this information and the restrictions given by (68).

The vectors of parameters m_1 and m_2 are estimated from expressions (10) and (11) respectively, using a generalized inverse due to the form of matrix V; the results for these estimates are

$$\hat{m}_1 = (120.5\ 103.5\ 106.5\ 115\ -75\ 135.5\ 65.5\ -40\ 78\ 101\ 0\ 83\ 85\ 0\ 100\ 0\ 0\ 120)^t$$

and

$$\hat{m}_2 = (76\ \ 79\ \ 7\ \ 94.5\ \ 286\ \ 98)^t$$

and the initial variance residual is estimated from the expression (47) as follows

$$\hat{\sigma}_1^2 = \frac{1}{23-14} Y_1^t [I_{23} - V(V^t V)^- V^t] Y_1 = 9.8$$

In Table 2, the analysis of variance for the test of hypothesis $H_0 : Vm_1 = 0$ with the initial information is presented. This hypotheses is emphatically rejected with any reasonable probability of Type-I error based on the small $p - value$.

Table 2 - Analysis of variance for hypothesis $H_0 : Vm_1 = 0$ with the initial information

Source	DF	SS	MS	F	$p - value$
Treatments	13	8925.65	686.59	70.20	<0.0001
Material	1	872.98	872.98	89.28	<0.0001
Corrosion	2	6269.05	3134.52	320.58	<0.0001
Material*Corrosion	2	304.28	152.14	15.56	0.0012
Cut	3	1182.85	394.28	40.32	<0.0001
Material*Cut	3	163.84	54.61	5.59	0.0193
Cut*Corrosion	2	132.65	66.32	6.78	0.0160
Error	9	88.00	9.78		
C-Total	22	9013.65			

In order to know what estimable function are testable in the analysis of variance of Table 2, we will employ the following expression

$$(V_2^t E_1 V_2)(V_2^t E_1 W_1)\mu_1 \tag{70}$$

with $E_1 = I_1 - V_1(V_1^t V_1)^- V_1^t$ and where matrices V_1 and V_2 are given in the modified cell means model (9) with the appropriate restriction in (3). Specifically, these matrices are obtained through the sequential fit of models. In this way, matrix V_2 contains the effect of interest and effects that preceded it, which are in V_1.

For example, the estimable functions for the Cut Depth are given in Table 3, these hypotheses are not easy to interpret. By repeating the same process over the other effects involved in the experiment, the rest of estimable functions can be built (see Melo, 2001).

Table 3 - Type-I estimable functions for the Cut Depth with initial information

Means	Row 1	Row 2	Row 3
μ_{111}	0.12	-0.27	-0.04
μ_{112}	-0.12	0.27	0.04
μ_{121}	0.36	0.13	0.05
μ_{123}	0.02	0.15	0.32
μ_{124}	-0.38	-0.27	-0.38
μ_{132}	0.13	0.40	0.19
μ_{134}	-0.13	-0.40	-0.19
μ_{211}	0.27	0.01	-0.15
μ_{213}	-0.04	0.18	0.44
μ_{214}	-0.24	-0.19	-0.29
μ_{222}	-0.02	0.32	-0.23
μ_{223}	0.02	-0.32	0.23
μ_{231}	0.25	0.13	0.14
μ_{234}	-0.25	-0.13	-0.14

In general, Type-I estimable functions and their sums of squares depend on the order in which the effects are added in the model. According to Table 2 all the effects are significant, but researchers should be careful with hypotheses testing through the different sums of square because not all hypotheses are testable.

4.1 Illustration of the method for the upgrade in the estimates when information is added to the experiment.

We suppose now that additional information arrived to the group of data with initial information presented in Table 1. It is presented in Table 4.

Table 4 - Additional information in the factorial design $2x3x4$

Corrosion	Material I				Material II			
	Cut depth				Cut depth			
rate	0.15	0.20	0.30	0.40	0.15	0.20	0.30	0.40
0.20		82		102				103
0.25			105		74			
0.30	114						111	
	108						107	

In this case, as in model (21) some empty cells are connected, then some cells,

which have additional information corresponding to the empty cells, can be written as linear combination of the connected cells with the initial information. For that, keeping in mind theorem 2 and equations (26) and (27), we have the following estimates for m_1^* and m_2^*,

$$\hat{m}_1^* = (105\ 103.5\ 112.5\ 115\ 120.5\ 134\ 67\ -26.5\ 78\ 100.5\ 74\ 83\ 85\ 86\ 97\ 0\ 109\ 121.5)^t$$

and

$$\hat{m}_2^* = (74.5\ \ 80.5\ \ 81.5\ \ 94.5\ \ 203\ \ 101.5)^t$$

As we observe above, the non-connected cell means produced negative parameter estimations.

The upgraded residual variance with the additional information is obtained from theorem 3 by means of expression (46) as follows

$$\hat{\sigma}^2 = \frac{1}{32-17}(88 + 214605 - 307171 + 53018 + 39650) = 12.67$$

In Table 5, the analysis of variance for the test of the hypothesis $H_0 : Vm_1^* = 0, Mm_1^* = 0, Fm_1^* = 0$ with all information (initial and additional information) is presented. This hypothesis is emphatically rejected with any reasonable probability of Type-I error based on the small $p-value$. This result is similar to the one obtained above.

Table 5 - Upgrade of analysis of variance for hypothesis $H_0 : Vm_1^* = 0, Mm_1^* = 0, Fm_1^* = 0$ with the additional information

Source	DF	SS	MS	F	$p-value$
Treatments	16	10528.46	658.03	51.95	<0.0001
Material	1	830.28	830.28	65.55	<0.0001
Corrosion	2	6398.90	3199.28	252.59	<0.0001
Material*Corrosion	2	379.22	189.61	14.97	0.0003
Cut	3	2507.62	835.87	65.99	<0.0001
Material*Cut	3	159.17	53.06	4.19	0.0243
Cut*Corrosion	5	253.28	50.66	4.00	0.0166
Error	15	190.00	12.67		
C-Total	31	10718.46			

To know what estimable functions are testing in the analysis of variance in Table 5, we will employ the following expression

$$\left[\begin{pmatrix} V_2^t & M_2^t & F_2^t \end{pmatrix} E_2 \begin{pmatrix} V_2 \\ M_2 \\ F_2 \end{pmatrix} \right]^{-} \begin{pmatrix} V_2^t & M_2^t & F_2^t \end{pmatrix} E_2 \begin{pmatrix} W_1 \\ W_2 \\ W_3 \end{pmatrix} \mu$$

with

$$E_2 = \begin{pmatrix} I_1 & 0 & 0 \\ 0 & I_M & 0 \\ 0 & 0 & I_F \end{pmatrix} - \begin{pmatrix} V_1 \\ M_1 \\ F_1 \end{pmatrix} (V_1^t V_1 + M_1^t M_1 + F_1^t F_1)^{-} \begin{pmatrix} V_1^t & M_1^t & F_1^t \end{pmatrix}$$

and where matrices V_1, V_2, M_1, M_2, F_1 and F_2 were given in the modified cell means model (21) with the appropriate restriction in (17). These matrices are obtained as in (70). As above, by repeating the same process over the other effects involved in the experiment, the rest of estimable functions can be built.

For example, the estimable functions with additional information for the Cut Depth are given in the Table 6.

Table 6 - Type-I estimable functions for the Cut Depth with additional information

Means	Row 1	Row 2	Row 3
μ_{111}	0.15	-0.16	0.04
μ_{112}	-0.01	0.38	0.06
μ_{114}	-0.14	-0.22	-0.10
μ_{121}	0.22	0.10	-0.02
μ_{123}	0.01	0.07	0.28
μ_{124}	-0.22	-0.17	-0.25
μ_{131}	0.21	0.00	0.09
μ_{132}	0.02	0.27	0.05
μ_{134}	-0.23	-0.28	-0.14
μ_{211}	0.22	0.10	-0.02
μ_{213}	0.01	0.07	0.28
μ_{214}	-0.22	-0.17	-0.25
μ_{221}	0.07	-0.10	-0.06
μ_{222}	-0.01	0.34	-0.11
μ_{223}	-0.06	-0.24	0.17
μ_{231}	0.13	0.06	-0.01
μ_{232}	0.05	0.09	-0.27
μ_{234}	-0.18	-0.15	-0.26

We can see that the estimable functions for this effect are not the same as those found with initial information (see Table 3). Therefore, the hypotheses testable and their associated sums of square are testing combinations of different effects. In this

way, although according to Table 5 all the effects are significant, this analysis is not similar to the analysis with initial information because another set of hypotheses is testing here.

Additionally, in the presence of some significant interactions, as in this case, these hypotheses may not be of general interest and more specialized hypotheses might be considered. However, hypotheses which depend on particular cell frequencies seem to be very difficult to justify.

Conclusions

This article carried out theoretical developments that lead to the estimation of parameters and the computation of the residual variance as a function of the initial estimations for cell means models for N-ways of classification with fixed effects, when additional information is available for the cells. Some properties of those new estimators, such as unbiasedness and minimum variance were proven.

Acknowledgement by researchers about the prior restrictions in an experimental design where there are unbalanced and empty cells can lead to estimability through the connectedness of estimable functions studied. Otherwise, if we do not know those restrictions, the analysis of the information can be made with the methodology presented for experiments with or without additional information.

Acknowledgments

The authors gratefully acknowledge the comments and suggestions of two anonymous referees that helped improve the article immensely. Also they are indebted to the investigation project: Applied Statistics to the Experimental Investigation, Industry and Biotechnology. National University of Colombia.

MELO, O.; LÓPEZ, L.; MELO, S. Modelos de médias de caselas com informação adicional: Estimação e análise. *Rev. Mat. Estat.,* São Paulo, v.24, n.3, p.115-137, 2006.

- RESUMO: Neste trabalho, novos desenvolvimentos teóricos foram feitos para a estimação de parâmetros e ajuste da variância residual, em função de estimações iniciais nos modelos de médias de caselas de efeitos fixos, a N-vias de classificação quando se tem informação adicional nas caselas. Também foram provadas as propriedades dos estimadores, como dos estimadores serem não-viciados e com variância minima. No ajuste dos novos parâmetros não foi necessário o conhecimento da informação inicial para obter as novas estimações. Este problema pode-se apressentar em pesquisas da industria, da biologia, da agricultura entre outras. Finalmente, apresenta-se uma aplicação com delineamentos factoriais.

- PALAVRAS-CHAVE: Delineamentos desbalanceados; modelos de médias de caselas; informação adicional; conectores.

References

HOCKING, R. R. *The analysis of linear models*. Monterrey: Brooks/Cole, 1985. 385p.

HOCKING, R. R. *Methods and applications of linear models*. New York: John Wiley & Sons, 1996. 731p.

HOCKING, R. R. *Methods and applications of linear models*. 2nd.ed. New York: John Wiley & Sons, 2002. 731p.

IEMMA, A. F. *Análisis de varianza de experimentos con celdas vacías*. Piracicaba: Escuela Superior de Agricultura "Luiz De Queiroz" da Universidade De São Paulo-USP, 1993. 90p.

JOHN, P. *Incomplete block design*. New York: Marcel Dekker, 1974. 101p.

LÓPEZ, L. A.; RINCÓN, L. F. *Modelos lineales - Notas de clase*. Universidad Nacional de Colombia. Bogotá: Departamento de Matemáticas y Estadística, 1999. 129p.

LÓPEZ, L. A. Los modelos de medias de celdas, una herramienta fundamental en la estadística industrial. In: SIMPOSIO DE ESTADÍSTICA, 1999. Rionegro. *Control Estadístico de la Calidad*, Bogotá: Universidad Nacional de Colombia, 1999. 65p.

MELO, C.; MELO, O.; LÓPEZ, A. Hipóteses efetivas em modelos de médias de caselas, construção através do método de Murray-Smith. *Multiciência*, São Carlos, v.4, n.2, p.104-123, 2001.

MELO, O.; LOZANO, A.; LÓPEZ, L. Funciones estimables en modelos de clasificación con datos desbalanceados a través del algoritmo de Cholesky. *Multiciência*, São Carlos, v.3, n.2, p.131-147, 1999.

MELO, S.; LÓPEZ, L.; MELO, O. Comparação de métodos para detectar desconexões em modelos de classificação múltipla sem interação. *Multiciência*, São Carlos, v.4, n.1, p.174-194, 2000.

MURRAY, L. W.; SMITH, D. W. Estimability, testability and connectedness in the cell means model. *Commun. Stat.*, New York, v.14, n.8, p.1889-1917, 1985.

MURRAY, L. W. Estimation of missing cells in randomized block and latin square designs. *Am. Stat.*, Washington, v.40, n.4, p.289-293, 1986.

RAO, C. R. On the linear combination of observations and the general theory of least squares. *Sankhyâ*. p.237-256, 1945.

SPEED, F. M.; HOCKING, R. R.; HACKNEY, O. P. Methods of analysis of linear models with unbalanced data. *J. Am. Stat. Assoc.*. New York, v.73, p.105-112, 1978.

SEARLE, S. R. *Linear models*. New York: John Wiley & Sons, 1971. 532p.

SEARLE, S. R. *Linear models for unbalanced data*. New York: John Wiley & Sons, 1987. 533p.

WEEKS, D. L.; WILLIAMS, D. R. A Note on the determination of connectedness in an N-Way cross classification. *Technometrics*. Washington, v.6, n.3, p.319-324, 1964.

ZYSKIND, G. *Teoría de las hipótesis lineales*. Colegio de Postgraduados. Chapingo: Escuela Nacional de Agricultura-Centro de Estadística y Calculo, 1980. 210p.

Recebido em 20.04.2006.

Aprovado após revisão em 28.09.2006.