Revista Colombiana de Estadística
Volumen 23 (2000) N^o 2, páginas 15 *a* 27

ESTIMACIÓN DE PARÁMETROS Y CALCULO DE LA VARIANZA RESIDUAL EN MODELOS DE MEDIAS DE CELDAS CON INFORMACIÓN ADICIONAL

OSCAR O. MELO M. *
LUIS A. LÓPEZ P.**

Resumen

Este artículo proporciona algunos desarrollos teóricos que permiten la estimación de parámetros y el cálculo de la varianza residual para modelos de medias de celdas a N-vías de clasificación con efectos fijos, cuando existe información adicional en las celdas. El conocimiento de los valores iniciales de las celdas, no se requiere para la estimación de los parámetros si existe nueva información en el experimento.La actualización de las estimaciones puede obtenerse teniendo en cuenta únicamente las estimaciones iniciales de los parámetros.La situación antes descrita suele presentarse frecuentemente en experimentación industrial, agrícola y biológica, entre otras áreas del conocimiento científico.

*Palabras Claves:*Diseños desbalanceados, modelos de medias de celda, información adicional , modelos conectados.

*Profesor Asistente, Departamento de Estadística, Universidad Nacional de Colombia; Sede Medellín
**Profesor Asociado, Departamento de Matemáticas y Estadística, Universidad Nacional de Colombia

Abstract

This paper provides some theoretical developments which allow the estimation of parameters and residual variance for cell means models with N-classification ways with fixed effects when there is additional information in the cells. For the estimation of the parameters, the knowledge of the initial values for the cells is not required if there is new information in the experiment. To update them, only the estimation obtained with the initial information of parameters are needed. This estimation problem is common in the industrial, agricultural and biological experimentation, among other areas of the scientific knowledge.

1. Introducción

Generalmente, un investigador planea un experimento de forma balanceada, con el objeto de facilitar el análisis e interpretación de la información, pero también puede planear de antemano un diseño desbalanceado o puede suceder que por accidente o por algunas otras razones ajenas al experimento mismo haya perdida de información, provocando esto último un desbalanceamiento en el diseño. El desbalanceamiento se traduce en la pérdida de simetría del diseño, dificultad en la construcción de la base de funciones estimables y la identificación de las hipótesis efectivas de interés a la investigación.

Para dar solución a la presencia de celdas vacías o datos faltantes, se han propuesto en la literatura estadística varias salidas; se pueden citar principalmente las siguientes: i) repetir el experimento bajo las mismas condiciones iniciales, ii) estimar los datos faltantes minimizando la suma de cuadrados del error introduciendo variables auxiliares o mediante cualquier otro método de inputación de información faltante o iii) usar métodos iterativos de estimación y analizar los datos tal como están, es decir, realizar un análisis de la información en presencia de celdas vacías.

También, se puede partir de un diseño balanceado, pero debido a condiciones propias del experimento se incluye información adicional, en algunas o todas las combinaciones de celdas, conllevando esto último a un desbalanceamiento del diseño; o puede ocurrir que al inicio del experimento no se haya obtenido información en algunos puntos del diseño y cuando se reali zan dichas repeticiones adicionales en condiciones similares pero en diferentes tiempos, manteniendo las mismas condiciones iniciales del experimento, pueda completarse la información en las celdas; o que por el contrario se tenga nueva información en celdas en las que inicialmente no se tenía. La inclusión de esta información adicional al diseño conlleva a la estimación de un nuevo conjunto de parámetros y en consecuencia un análisis apropiado para dicho problema debe realizarse.

El tener información adicional en las celdas, puede ocasionar desbalancemiento en el experimento, en esta dirección es de gran utilidad los recientes estudios propuestos por López (1992), Iemma (1993), Mondardo (1994), Iemma, et. al. (1999), López (1999) y Melo, et al (2000), en donde se llama la atención sobre los problemas a que debe enfrentarse el investigador cuando se abordan modelos de clasificación con estructura desbalanceada y aún en presencia de celdas vacías. En estas referencias, se resalta el cuidado que debe tenerse en la identificación de las verdaderas hipótesis probables.

2. Metodología

El marco metodológico para el desarrollo teórico del problema son los modelos lineales de medias de celda y modelos de medias de celda modificado, los cuales se resumen en las secciones 2.1 y 2.2.

2.1. Modelos de medias de celda - Modelo M

Siguiendo a Speed, et. al. (1978), se define el modelo de medias de celda o modelo M, como el modelo lineal:

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

donde, Y_1 es un vector de variables aleatorias de tamaño $kx1$, W_1 es una matriz de bloques de tamaño kxq con el $ij\cdots s$-ésimo bloque diagonal correspondiente a un vector columna de unos de tamaño $n_{ij\cdots s}$ (denotado por $Jn_{ij\cdots s}$), en donde $n_{ij\cdots s}$ es el número de observaciones de la $ij\cdots s$-ésima celda, μ_1 es un vector de medias poblacionales de tamaño $qx1$ y e_1 es un vector de variables aleatorias no observables de tamaño $kx1$ tal que $e_1\tilde{\ }N(\widetilde{0},\sigma_1^2 I_1)$.

Si no hay celdas pérdidas, W_1 tiene rango columna completo; si por el contrario hay celdas con información faltante, entonces W_1 tiene una columna cero por cada celda perdida por tanto, esta matriz no va a tener rango columna completo .Por su parte μ_1 se sigue tomando como si todas las celdas hubieran sido observadas.

En este contexto el sistema tiene como solución a $\widehat{\mu_1} = (W_1^t W_1)^- W_1^t Y_1$, la cual coincide con la estimación mínimos cuadrados para μ_1. Así, el Mejor Estimador Lineal Insesgado (MELI) de $\mu_{ij\cdots s}$ es la media de la celda $ij\cdots s$, es decir MELI$(\mu_{ij\cdots s}) = \overline{Y}_{ij\cdots s}$.

Si al modelo (1) se le impone una restricción del tipo:

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

se le conoce como modelo de medias de celda con restricción, siendo G es una matriz de contrastes desconocidos de orden sxq de rango s.

2.2. Modelo de medias de celda modificado.

Si en (2) $g = 0$, entonces el modelo de medias de celda es caracterizado en Hocking (1985, 1996) y Speed, et. al. (1978) como el modelo (1), sujeto a la restricción $G\mu_1 = 0$ donde G es especificada como en la sección 2.1 y va a representar las relaciones lineales conocidas sobre las medias de celda. Generalmente se van a especificar contrastes de no-interacción en el modelo de medias de celda, como puede verse en Hocking (1985, 1996), Searle (1987) y, Murray y Smith (1985), entre otros.

Las columnas de G pueden ser reordenadas y G particionada en dos submatrices, es decir $G = \left[\begin{array}{ccc} G_1 & \vdots & G_2 \end{array}\right]$ con G_2 de orden sxs de rango s y G_1 de orden $sx(q-s)$. La partición de μ_1 es independiente de los datos obtenidos; en particular, es independiente del número de celdas perdidas (digamos f) y su localización. La partición solamente depende de G y las relaciones lineales entre las medias involucradas en esta matriz; esta partición depende de como el experimento se concibió y no como se realizo. Teniendo en cuenta la partición hecha en G., las columnas de μ_1 se pueden reordenar de la siguiente manera:

$$\mu_1^t = \begin{pmatrix} m_1^t & m_2^t \end{pmatrix} \tag{3}$$

y así(2) se puede escribir como

$$G\mu_1 = G_1 m_1 + G_2 m_2 = g \tag{4}$$

La escogencia de la submatriz G_2 es arbitraria y cuando es de rango completo, existe una solución única para m_2 en términos de m_1 y esta dada por:

$$m_2 = G_2^{-1}\left(g - G_1 m_1\right) \tag{5}$$

Teniendo en cuenta las particiones (4) y (5), la matriz de incidencia W_1 debe reordenarse de acuerdo con la partición de G y μ_1. De esta manera, de (1) se obtiene el modelo de medias de celda modificado de Murray y Smith (1985).

$$Y_1 = \begin{pmatrix} \omega_1 & | & \omega_2 \end{pmatrix} \begin{pmatrix} m_1 \\ m_2 \end{pmatrix} + e_1 \tag{6}$$

Al sustituir (6) en (7), el modelo (1) se puede reescribir como:

$$Y_1 - \omega_2 G_2^{-1} g = \left(\omega_1 - \omega_2 G_2^{-1} G_1\right) m_1 + e_1 \tag{7}$$

al hacer $Y_1^* = Y_1 - \omega_2 G_2^{-1} g$ y $V = \omega_1 - \omega_2 G_2^{-1} G_1$ una matriz transformada de tamaño $kx(q-s)$ de rango $(q-s)$, entonces (8) lo escribimos como:

$$Y_1^* = V m_1 + e_1 \tag{8}$$

Por ser V una matriz no singular , se puede aplicar el método usual de mínimos cuadrados para el modelo de rango completo sin restricción, obteniéndose el mejor estimador lineal insesgado para m_1, como:

$$Meli(m_1) = \widehat{m}_1 = \left(V^t V\right)^{-1} V^t Y_1^* \tag{9}$$

y cuando sustituímos (10) en (6) se sigue que el mejor estimador lineal insesgado para para m_2 es:

$$Meli(m_2) = \widehat{m}_2 = G_2^{-1}\left[g - G_1 \left(V^t V\right)^{-1} V^t Y_1^*\right] \tag{10}$$

Si no hay celdas con información faltante, entonces el rango de V es con seguridad $(q-s)$. Si hay celdas vacías, V puede aún tener rango $(q-s)$, en este caso (10) y (11) son los únicos MELIS de m_1 y m_2. Si el rango de V es menor de $(q-s)$, entonces esencialmente se tiene el mismo problema del modelo superparametrizado (ver Searle 1971) cuando la matriz X no es de rango completo .

En el modelo (9), al imponer restricciones de no interacción, se puede establecer de manera sencilla si es posible conectar celdas vacías con información de celdas donde hay información y poder asíestablecer finalmente la base de funciones estimables. Por la importancia de estos dos conceptos se presenta en la sección 2.2.1 una síntesis de estimabilidad y conectés.

2.2.1. Estimabilidad y conectés

En Searle (1987) se plantea que un diseño es conectado, si es posible unir todas las parejas de tratamientos por un camino involucrando celdas llenas en las cuales alternativamente se puede pasar en la misma fila (tratamiento) o en la misma columna (bloque). Asílos tratamientos i y r en las celdas (i,j) y (r,t) pueden ser conectados por el camino

$$(i,j) \to (i,v) \to (u,v) \to (u,t) \to (r,t) \tag{11}$$

En Hocking (1985, 1996), Searle (1971, 1987) se da una definición de conectes para modelos de dos vías sin interacción, la cual había sido generalizada por Weeks y Williams (1964), quienes presentan una propuesta para el estudio de conectes en modelos lineales a N-vías de clasificación cruzada sin interacción. En la definición 1 se caracteriza en forma sencilla si un modelo de medias de celdas es conectado.

Definición. Un experimento conformado por un conjunto de datos asocia do a un modelo de medias de celda, se dice conectado si μ_1 es linealmente estimable de manera única.

Un conjunto particular de datos puede ser conectado por un modelo y no por otro. Por ejemplo, en el modelo de clasificación con tres factores se tienen varios niveles de conectes dependiendo de la restricción necesaria para alcanzar estimabilidad de μ_1. El modelo sin restricción requiere $n_{ijk} > 0$. La condición de no-interacción entre los tres factores puede ser suficiente para permitir la estimación de todos los μ_{ijk} en el caso de celdas perdidas. Si esta no es suficiente, entonces la imposición de restricción de no-interacción de menor orden puede conducir a la estimabilidad.

Se debe enfatizar que la búsqueda de estimabilidad de los parámetros no justifica el supuesto en las restricciones. Este supuesto se debe hacer antes de recoger la información y debe basarse en información a-priori respecto a las relaciones entre las medias de celdas. Es decir, se esta interesado con el análisis como inicialmente se concibió y no en análisis basados en algunas conveniencias dictadas por los datos.

Murray y Smith (1985) presentan un nuevo criterio para conectes basándose en el modelo de medias de celda modificado (9) y en la definición 1, el cual es resumido en el siguiente teorema:

Teorema 1. Para el modelo (1) y (2) el experimento es conectado si y solo si V tiene rango columna completo, es decir el rango de V es $q - s < k$. (ver prueba en Murray y Smith 1985)

Si V tiene rango columna completo, entonces el experimento es conectado, μ_1 es linealmente estimable en su totalidad, y el análisis original puede ser llevado a cabo como se planeo. Si V no tiene rango columna completo, entonces el experimento no es conectado, μ_1 no es linealmente estimable y el análisis original no puede realizarse. En este caso el modelo de medias de celda (1) y (2) no es de rango completo.

En muchas situaciones dependiendo de la naturaleza de la información, algunas de las medias no-observadas pueden ser estimadas haciendo uso de la restricción (3) y las medias observadas; es decir pueden conectarse. En otras, solamente algunas combinaciones lineales de medias no observadas pueden ser estimadas.

3. Actualización de los parámetros y de la varianza del residual.

En esta sección, se supone que llega información adicional a cualquiera de las celdas, es decir puede suceder que llegue información adicional en celdas donde ya había información o en otras donde inicialmente las celdas no reportaban información. Para la actualización tanto de los parámetros como de la varianza residual cuando hay información adicional en cualquiera de las celdas, reescribimos el modelo (1) como:

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

donde,

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

y, con: Y_1 de tamaño $kx1$ vector de variables aleatorias que va a contener la información inicial, Y_2 de tamaño $tx1$ vector de variables aleatorias de la información adicional que llega a las mismas celdas donde ya se tenía información, Y_3 de tamaño $(n-k-t)x1$ vector de variables aleatorias de la información adicional que llega a las celdas donde inicialmente no se tenía información, $\sigma_1^2 I_1$ de tamaño kxk matriz de varianzas de las variables aleatorias de la información inicial, $\sigma_2^2 I_2$ de tamaño txt matriz de varianzas de las variables aleatorias de la información adicional que llega a las celdas en donde ya se tenía información, $\sigma_2^2 I_3$ de tamaño $(n-k-t)x(n-k-t)$ matriz de varianzas de las variables aleatorias de la información adicional en las celdas en donde no se tenía información, W_1de tamaño kxk, W_2 de tamaño txt y W_3 de tamaño $(n-k-t)x(n-k-t)$, son matrices de bloques con el $ij\cdots s$-ésimo bloque diagonal correspondiente a un vector columna de unos de tamaño $n_{ij\cdots s}$ (denotado por $Jn_{ij\cdots s}$), en donde $n_{ij\cdots s}$ es el número de observaciones de la $ij\cdots s$-ésima población tanto para la información inicial como para información adicional y μ es un vector de medias poblacionales.

4. Estimación de los parámetros.

Para llevar a cabo la estimación de parámetros y el cálculo de la varianza residual en el modelo (12), suponemos distribución normal de los errores de tal forma que al considerar el logaritmo de la función de verosimilitud:

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

Para dar solución a (13) y para efectos de este trabajo, se supone que la varianza de la información inicial es la misma varianza de la información adicional, con este supuesto se obtiene las ecuaciones normales:

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

o equivalentemente, el modelo se puede escribir como:

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

Si en (15) se le impone la restricción del tipo (2), donde según Murray y Smith (1985) la matriz de restricción de interacción para una interacción dada es:

$$G=M_1\otimes M_2\otimes\cdots\otimes M_s \tag{16}$$

donde $M_i,\ i=1,2,\ldots,s$ hace referencia a la presencia de un factor de interés y $\otimes$ denota el producto Kronecker.

Además las matrices M_i son obtenidas por la expresión:

$$M_i=\begin{cases} D_i & \textit{si la interacción involucra el factor } i\\ J_i^t & \textit{si la interacción no involucra el factor } i\end{cases} \tag{17}$$

donde, $D_i=\begin{pmatrix} I_{(i-1)} & \vdots & -J_{(i-1)}\end{pmatrix}$ con $I_{(i-1)}$ una matriz identidad de dimensión $(i-1)$ y $J_{(i-1)}$ un vector columna de unos de longitud $(i-1)$. El uso de estas matrices D es muy conveniente para describir la restricción de interacción de altos ordenes. Esta regla puede también ser usada para generar matrices con el fin de hacer pruebas de hipótesis sobr efectos principales o interacciones.

En forma semejantea los desarrollos hechos en la sección 2.2, se tiene que el equivalente a (4) en el modelo (12) es :

$$\mu^t = \begin{pmatrix} m_1^{*t} & m_2^{*t} \end{pmatrix} \tag{18}$$

y así(18) se puede reescribir como

$$G\mu = G_1 m_1^* + G_2 m_2^* = g \tag{19}$$

donde la escogencia de la submatriz G_2 es arbitraria Una solución única para m_2^* en términos de m_1^* existe y esta dada por:

$$m_2^* = G_2^{-1}\left(g - G_1 m_1^*\right) \tag{20}$$

Teniendo en cuenta (18), (19) y (20), y reordenando apropiadamente las columnas de las matrices de incidencia W_1, W_2 y W_3 entonces (17) se puede reescribir como:

$$\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} = \begin{pmatrix} \omega_1 & \vdots & \omega_2 \\ U_1 & \vdots & U_2 \\ S_1 & \vdots & S_2 \end{pmatrix} \begin{pmatrix} m_1^* \\ m_2^* \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ e_3 \end{pmatrix} \tag{21}$$

Al reemplazar (20) en (21), se obtiene:

$$\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} = \begin{pmatrix} \omega_2 \\ U_2 \\ S_2 \end{pmatrix} G_2^{-1} g + \begin{pmatrix} \omega_1 - \omega_2 G_2^{-1} G_1 \\ U_1 - U_2 G_2^{-1} G_1 \\ S_1 - S_2 G_2^{-1} G_1 \end{pmatrix} m_1^* + \begin{pmatrix} e_1 \\ e_2 \\ e_3 \end{pmatrix} \tag{22}$$

Haciendo: $Y_1^* = Y_1 - \omega_2 G_2^{-1} g$, $Y_2^* = Y_2 - U_2 G_2^{-1} g$, $Y_3^* = Y_3 - S_2 G_2^{-1} g$, $V = \omega_1 - \omega_2 G_2^{-1} G_1$, $U = U_1 - U_2 G_2^{-1} G_1$, y $S = S_1 - S_2 G_2^{-1} G_1$.,entonces (22) es escrito como:

$$\begin{pmatrix} Y_1^* \\ Y_2^* \\ Y_3^* \end{pmatrix} = \begin{pmatrix} V \\ U \\ S \end{pmatrix} m_1^* + \begin{pmatrix} e_1 \\ e_2 \\ e_3 \end{pmatrix} \tag{23}$$

En (23), si algunas celdas vacías son conectadas según lo descrito en la sección 2.2.1, entonces algunas de las celdas donde se encuentra la información adicional correspondientes a las celdas vacías se puede escribir como combinación lineal de las celdas conectadas, es decir, algunos de las filas de S pueden ser escritas como combinación lineal de las filas de V. De esta manera, si se particiona Y_3^* y S en información conectada y no conectada, se obtiene para la información adicional que:

$$\begin{pmatrix} Y_2^* \\ Y_3^* \end{pmatrix} = \begin{pmatrix} Y_2^* \\ Y_C \\ Y_F \end{pmatrix} = \begin{pmatrix} Y_M \\ Y_F \end{pmatrix} y \begin{pmatrix} U \\ S \end{pmatrix} = \begin{pmatrix} U \\ C \\ F \end{pmatrix} = \begin{pmatrix} M \\ F \end{pmatrix} \tag{24}$$

En (24) Y_C corresponde al vector de observaciones de las celdas conectadas y Y_F corresponde al vector de observaciones de las celdas no conectadas.

Sustituyendo (24) en (23) se sigue:

$$\begin{pmatrix} Y_1^* \\ Y_M \\ Y_F \end{pmatrix} = \begin{pmatrix} V \\ M \\ F \end{pmatrix} m_1^* + \begin{pmatrix} e_1 \\ e_M \\ e_F \end{pmatrix} \tag{25}$$

Bajo el supuesto de normalidad, las soluciónes máximo verosimil para el modelo (25) son:

$$\widehat{m}_1^* = \left(V^tV + M^tM + F^tF\right)^- \left(V^tY_1^* + M^tY_M + F^tY_F\right) \tag{26}$$

La inversa de la suma de matrices en la ecuación anterior, se soluciono haciendo uso del lema 10.4, de López, A. y Rincón, F. (1999) ,teniendo entonces que :

$$\begin{aligned}\left(V^tV + M^tM + F^tF\right)^- &= \left(V^tV + F^tF\right)^- - \left(V^tV + F^tF\right)^- M^t \\ &\quad [I_2 + M\left(V^tV + F^tF\right)^- M^t]^{-1} M \left(V^tV + F^tF\right)^-\end{aligned} \tag{27}$$

Al sustituír (27) en (26), se obtiene la actualización de parámetros para m_1^* en términos de las estimaciones iniciales como:

$$\begin{aligned}\widehat{m}_1^* &= \left\{I_1 - \left(V^tV + F^tF\right)^- M^t \left[I_2 + M\left(V^tV + F^tF\right)^- M^t\right]^{-1} M\right\} \\ &\quad \left(V^tV + F^tF\right)^- \left[(V^tV)\widehat{m}_1 + M^tY_M + F^tY_F\right]\end{aligned} \tag{28}$$

con $\widehat{m}_1$ obtenido como en (10). Al sustituír (28) en (20), se obtiene:

$$\widehat{m}_2^* = G_2^{-1}\left(g - G_1\widehat{m}_1^*\right) \tag{29}$$

Obteniendose finalmente de (28) y (29) que:

$$\widehat{\mu}^t = \begin{pmatrix} \widehat{m}_1^{*t} & \widehat{m}_2^{*t} \end{pmatrix} \tag{30}$$

Al estimador dado en (30) se le comprueban las propiedades de insesgamiento y de mínima varianza, es decir:

$$E(\pi^t \widehat{\mu}) = \pi^t \mu \tag{31}$$

y

$$V\left(\pi^t \widehat{\mu}\right) = \pi^t \begin{pmatrix} V\left(\widehat{m}_1^*\right) & Cov\left(\widehat{m}_1^*, \widehat{m}_2^*\right) \\ Cov\left(\widehat{m}_1^*, \widehat{m}_2^*\right) & V\left(\widehat{m}_2^*\right) \end{pmatrix} \pi \tag{32}$$

donde,

$$\begin{aligned} V\left(\widehat{m}_1^*\right) &= \left\{ I_1 - \left(V^t V + F^t F\right)^- M^t \left[I_2 + M\left(V^t V + F^t F\right)^- M^t \right]^{-1} M \right\} \\ &\quad (V^t V + F^t F)^- \sigma^2 \end{aligned} \tag{33}$$

$$V\left(\widehat{m}_2^*\right) = G_2^{-1} G_1 V\left(\widehat{m}_1^*\right) G_1^t \left(G_2^{-1}\right)^t \tag{34}$$

$$Cov\left(\widehat{m}_1^*, \widehat{m}_2^*\right) = -G_2^{-1} G_1 V\left(\widehat{m}_1^*\right) \tag{35}$$

5. Estimación de la varianza residual.

Utilizando (25) para actualizar la estimación de la varianza residual, se obtiene:

$$\begin{aligned} (n-r)\widehat{\sigma}^2 = (k - r_V)\widehat{\sigma}_1^2 + SCM_I - \left((V^t V)\widehat{m}_1 + F^t Y_F + M^t Y_M\right)^t K \\ \left((V^t V)\widehat{m}_1 + F^t Y_F + M^t Y_M\right) + Y_M^t Y_M^t + Y_F^t Y_F \end{aligned} \tag{36}$$

Donde, $\widehat{\sigma}_1^2 = \frac{1}{k - r_V} Y_1^{*t} \left[I_1 - V\left(V^t V\right)^- V^t \right] Y_1^*$, $SCM_I = Y_1^{*t} V (V^t V)^- V^t Y_1^*$ y

$$K = \left\{ I_1 - (VV + FF)^- M^t \left[I_2 + M\left(V^t V + F^t F\right)^- M^t \right]^{-1} M \right\} (VV + FF)^- \tag{37}$$

Al estimador dado en (36) se le comprueba las propiedades de insesgamiento y de mínima varianza, encontrándose que:

$$E\left(\widehat{\sigma}^2\right) = \sigma^2 y V\left(\widehat{\sigma}^2\right) = \frac{2\sigma^2}{n-r} \; con \; n > r \tag{38}$$

6. Conclusiones

En este artículo se presentaron algunos desarrollos teóricos que permiten la estimación de parámetros y el cálculo de la varianza residual para modelos de medias de celdas a N-vías de clasificación con efectos fijos, cuando existe información adicional en las celdas. El conocimiento de los valores iniciales de las celdas, no se requiere para la estimación de los parámetros si existe nueva información en el experimento.La actualización de las estimaciones puede obtenerse teniendo en cuenta únicamente las estimaciones iniciales de los parámetros.

Referencias

[1] HOCKING, R. R.(1985). *The Analysis of Linear Models.* Brooks/Cole . Monterrey

[2] HOCKING, R. R. (1996) *Methods and Applications of Linear Models.* John Wiley & Sons, Inc. New York.

[3] EMMA, A.(1993)*F. Análisis de Varianza de Experimentos con Celdas Vacías.* Escuela Superior De Agricultura “Luiz De Queiroz”. Universidade De Sao Paulo. S. P-Brasil . Cap: II, III, V.

[4] IEMMA, A. F., LÓPEZ, L. A. y RINCÓN, L. F.(1999) *Proyectores Ortogonales Especiales.* Revista de Investigación Operacional. Universidad de la Habana - Cuba .Vol **20**. 107-114.

[5] LÓPEZ, L. A. (1992) *Sumário; Uso da Notação R( ) em dados Desbalanceados.* Escuela Superior de Agricultura ”Luiz De Queiroz”. Universidad de Sao Paulo. S. P.-Brasil Pág. 1-18.

[6] LÓPEZ, L. A y RINCON L. F.(1999) *Modelos Lineales - Notas de Clase.*Universidad Nacional de Colombia. Departamento de Matemáticas y Estadística. Bogotá D.C.

[7] LÓPEZ, L.A. (1999) *Los modelos de medias de celdas, una herramienta fundamental en la estadística industrial.* Simposio de Estadística. Rionegro, Antioquía .

[8] MELO, O. O, LOZANO, A. R. y LOPEZ, L. A.(2000)*Funciones Estimables en Modelos de Clasificación con Datos Desbalanceados a través del Algoritmo de Cholesky.* Revista Multiciencia. São Carlos, Brasil Vol.3, No.2, Pág 131-147.

[9] MONDARDO, M.(1994)*Estimabilidade de Funções Paramétricas com dados Desbalanceados através do PROC GLM do SAS: Aplicações a Pesquisa Agropecuaria.* Escuela Superior De Agricultura " Luiz De Queiroz" Universidad de Sao Paulo. S. P. - Brasil

[10] MURRAY, L. W. and SMITH, D. W.(1985)*Estimability, Testability and Connectedness in the Cell Means Model.* Communications in Statistics. New York Pág. 1889-1917.

[11] MURRAY, L. W. (1986) *Estimation of Missing Cells in Randomized Block and Latin Square Designs.* The American Statistician , Vol **40** (4). Pág. 289 - 293.

[12] RAO, C. R.(1945)*On the Linear Combination of Observations an the General Theory of Least Squares.* Sankhyâ Pág. 237-256.

[13] SEARLE, S. R.(1971) *Linear Models.* John Wiley & Sons. New York

[14] SEARLE, S. R.(1987) *Linear Models for Unbalanced Data.* John Wiley & Sons. New York

[15] SPEED, F. M., HOCKING, R. R. and HACKNEY, O. P. (1978)*Methods of Analysis of Linear Models With Unbalanced Data.* Journal of the American Statistical Association. Vol. **73**. Pág. 105-112.

[16] WEEKS, D. L, and WILLIAMS, D. R. (1964) *A Note on the Determination of Connectedness in an N-Way Cross Classification.* Technometrics. Vol. **6** No.3, Pág 319-324.