Revista Colombiana de Estadística
Diciembre 2010, volumen 33, no. 2, pp. 233 a 250

Procedimiento y algoritmo de estimación en modelos multinivel para proporciones

Procedure and Estimation Algorithm in Multilevel Models for Proportions

Ernestina Castells[1,a], Mario M. Ojeda[2,b], Minerva Montero[3,c]

[1]Facultad de Matemática, Universidad Autónoma de Guerrero, Acapulco, México
[2]Facultad de Estadística e Informática, Universidad Veracruzana, Xalapa, México
[3]Departamento de Matemática, Instituto de Cibernética, Matemática y Física, La Habana, Cuba

Resumen

En este artículo se describe un procedimiento para la estimación de parámetros fijos y aleatorios en modelos multinivel para proporciones. El procedimiento de estimación se basa en el método de los mínimos cuadrados generalizados. Una vez que se formula el modelo, se demuestra que es posible aplicar la teoría asintótica de estimación en el marco del modelo lineal general. Se elabora un algoritmo que permite calcular los estimadores propuestos. La aplicación se ilustra con un ejemplo de meta-análisis. Se concluye que el procedimiento presentado puede ser una estrategia favorable en investigaciones aplicadas.

***Palabras clave*:** mínimos cuadrados generalizados iterativos, modelos multinivel, tablas de contingencia.

Abstract

This paper describes a procedure for the estimation of fixed and random parameters in multilevel model for proportions. The estimation procedure is developed using Iterative Generalized Least Squares. Once the model is formulated, we demonstrate that it is possible to apply the asymptotic estimation theory in the framework of the general lineal model. An algorithm to calculate the proposed estimators is elaborated. We illustrate the application using an example of meta-analysis. It is concluded that the proposed procedure can be favorable strategy to do applied research.

***Key words*:** Contingency tables, Iterative generalized least squares, Multilevel models.

[a]Profesora titular. E-mail: ernestinacg@yahoo.com
[b]Profesor titular. E-mail: mojeda@uv.mx
[c]Investigadora auxiliar. E-mail: minerva@icmf.inf.cu

1. Introducción

La flexibilidad de la modelación multinivel ha implicado un papel cada vez más importante de esta técnica dentro de la teoría estadística y las aplicaciones; sin embargo, la inferencia estadística acerca de los modelos multinivel para proporciones es una problemática que presenta dificultades teóricas y computacionales aún sin resolver. Para evitar la carga computacional y la inestabilidad asociada con la compleja integración numérica que implica la estimación de parámetros de modelos multinivel no lineales se han desarrollado varios métodos. Entre los más comunes se encuentran los que se basan en aproximaciones de la integral involucrada en la verosimilitud marginal, los métodos bayesianos basados en algoritmos de Monte Carlo mediante Cadena de Markov (MCMC), los métodos EM estocásticos y los métodos de verosimilitud simulada (Fotouhi 2003). Estos métodos, aunque en principio aplicables con cierto éxito, presentan diversos inconvenientes; una de las más notables desventajas es su alto costo computacional.

En la actualidad, continúa siendo de particular interés explorar nuevos métodos que le brinden al investigador herramientas adecuadas para la estimación de parámetros en modelos multinivel para datos categóricos. El objetivo de este artículo es fundamentar la racionalidad del método propuesto por Montero, Castell & Ojeda (2007). Este enfoque se basa fundamentalmente en la integración de tres estrategias estadísticas: 1) El uso, como base fundamental para elaborar métodos de inferencia para tablas de contingencia, de modelos asociados a la distribución asintótica gaussiana de un vector de transformaciones de las frecuencias observadas; 2) El uso de aproximaciones simplificadoras para la matriz de covarianza asintótica de los errores aleatorios de nivel-1 en el modelo multinivel propuesto y 3) La aplicación de mínimos cuadrados generalizados iterativos para la estimación de parámetros de los modelos gaussianos resultantes.

En la sección 2 se describe brevemente la filosofía del análisis multinivel para datos categóricos. En la sección 3 se formula un modelo multinivel para proporciones y en la sección 4 se demuestra que los estimadores propuestos tienen propiedades que permiten aplicar la teoría de estimación para muestras grandes. En la misma sección se demuestra la consistencia de los estimadores, tanto de los parámetros asociados a efectos fijos como a componentes de la varianza. En la sección 5 se presenta el algoritmo de estimación descrito en el anexo. Finalmente, en la sección 6, la aplicación del método se ilustra con un ejemplo.

2. Motivación para el análisis multinivel de proporciones

Muchos problemas prácticos tratan datos categóricos que conllevan al análisis de un conjunto de tablas de contingencia. Ejemplos importantes se pueden encontrar en problemas de meta-análisis (Glass 1976, Hedges & Olkin 1985), en el análisis de datos de panel (Hsiao 1995, Hamerle & Honning 1995), en los estudios de casos y controles multicentro (Lubin, Blot, Berrino, Flamant, Gillis, Kunze, Schmäwhl & Visco 1984, Fears & Brown 1986, Breslow & Zhao 1988) y en problemas de estimación en áreas pequeñas (Rudas 1986). En todos estos casos, los individuos se seleccionan de diferentes grupos y es posible reconocer que las observaciones pertenecientes a un mismo grupo son más parecidas entre sí que las que se encuentran en grupos diferentes. Ignorar la estructura de grupos puede provocar serios problemas inferenciales (Snijders & Bosker 1999).

Cuando se analizan datos en un conjunto de tablas de contingencia es posible establecer una estructura jerárquica de datos de dos niveles, en la que los grupos que conforman las tablas se reconocen como las unidades de nivel-2 y los individuos anidados dentro de los grupos se identifican como las unidades de nivel-1. Un enfoque que distingue los diferentes niveles de las variables explicativas y toma en cuenta explícitamente la varianza dentro y entre grupos es la modelación multinivel (Goldstein 1995). En los modelos multinivel, también conocidos como modelos jerárquicos (Brik & Raudenbush 1992) o modelos de coeficientes aleatorios (Longford 1995), algunos parámetros varían de grupo a grupo, lo que permite tratarlos como efectos aleatorios.

Frecuentemente, el interés del estudio en tablas de contingencia se centra en el análisis de un gran número de funciones de las probabilidades de las celdas. Uno de los modelos más utilizados es el de regresión logística multinivel (Efron 1996, Hartzel, Liu & Agresti 2001, Lee & Nelder 2002); sin embargo, la aparición de problemas cada vez más complejos exige que otras funciones (Forthofer & Lehnen 1981) diferentes de las conocidas funciones logit o probit, también sean tomadas en consideración.

La idea básica del enfoque descrito en este artículo es usar funciones de las probabilidades como las variables dependientes en un modelo lineal multinivel (Montero 2006). Este enfoque se basa en el uso de los mínimos cuadrados ponderados o mínimos cuadrados generalizados para datos categóricos. El empleo de esta metodología, presentada por primera vez por Grizzle, Starmer & Koch (1969) se había limitado sólo al caso donde los parámetros del modelo son fijos (Grizzle et al. 1969). En este artículo esta estrategía se extiende al caso multinivel.

La aplicación del método se ilustra a través de un ejemplo de meta-análisis, que puede considerarse como un problema estadístico multinivel, ya que la información dentro de los estudios se combina en presencia de una heterogeneidad potencial entre los estudios.

3. El modelo lineal multinivel para proporciones

En esta sección se describe la estructura jerárquica impuesta por un conjunto de tablas de contingencia y se formula un modelo multinivel para proporciones. Para simplicar la futura discusión de la teoría de estimación que se presenta en la sección 4 el modelo se expresa en términos matriciales.

Sea Y una variable respuesta con R categorías. Sean $X_1, X_2, \ldots, X_t$, un conjunto de variables explicativas. Sea G un factor de estratificación en J grupos (unidades de nivel-2). Sea $\mathcal{A}$ el conjunto formado por las diferentes combinaciones de los valores de $X_1, X_2, \dots X_t$. Sean I el cardinal del conjunto $\mathcal{A}$ y $C = \{1, 2, \cdots I\}$. A cada elemento de $\mathcal{A}$ se le hace corresponder biunívocamente un elemento c de $\mathcal{C}$. Todos los individuos (unidades de nivel-1) con la misma combinación de valores $(x_1, x_2, \dots x_t)$ se tratan como un subgrupo dentro de los grupos determinados por G. De cada subgrupo se seleccionan muestras aleatorias independientes de tamaño n_{ji}. Sea n_{jir} el número de individuos en la muestra del i-ésimo subgrupo del j-ésimo grupo, clasificados en la r-ésima categoría de respuesta ($i = 1, 2, \dots, I$; $j = 1, 2, \dots, J$; $r = 1, 2, \dots R$).

Sea $\pi_{r|ji} = P\left(Y = r \mid c = i, g = j\right)$. Se supone que $(n_{ji1}, n_{ji2}, \dots, n_{jir})$ para $i-j$ fijos sigue una distribución multinomial con probabilidades $\left(\pi_{1|ji}, \pi_{2|ji}, \dots, \pi_{R|ji}\right)$ y que las muestras correspondientes a diferentes subgrupos y/o diferentes grupos son independientes. La estructura de datos descrita anteriormente puede resumirse en J tablas de contingencia $I \times R$ como la que se muestra en la tabla 1.

TABLA 1: Tabla de contingencia para el j-ésimo grupo.

	Categorías de la respuesta				
Subgrupo	1	2	$\cdots$	R	Total
1	n_{j11}	n_{j12}	$\cdots$	n_{j1R}	n_{j1}
2	n_{j21}	n_{j22}	$\cdots$	n_{j2R}	n_{j2}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
I	n_{jI1}	n_{jI2}	$\cdots$	n_{jIR}	n_{jI}

El vector de probabilidades, π_j, asociado a la j-ésima tabla se puede escribir como $\boldsymbol{\pi}_j = \left(\boldsymbol{\pi}'_{j1}, \boldsymbol{\pi}'_{j2}, \dots, \boldsymbol{\pi}'_{jI}\right)'$, donde $\boldsymbol{\pi}_{ji} = \left(\pi_{1|ji}, \pi_{2|ji}, \dots, \pi_{R|ji}\right)'$ con $\sum\limits_{r=1}^{R} \pi_{r|ji} = 1$, para cada (i, j). Cada conjunto de probabilidades tiene $R - 1$ elementos funcionalmente independientes. Los J vectores de probabilidades en cada tabla de contingencia conforman un único vector $\boldsymbol{\pi} = (\boldsymbol{\pi}'_1, \boldsymbol{\pi}'_2, \dots, \boldsymbol{\pi}'_J)'$.

Sea $\boldsymbol{F}_m(\boldsymbol{\pi}_j)$, $m = 1, 2, \dots, a$ un conjunto de funciones de $\boldsymbol{\pi}_j$. Defínase $\boldsymbol{F}(\boldsymbol{\pi}_j) = [\boldsymbol{F}_1(\boldsymbol{\pi}_j), \boldsymbol{F}_2(\boldsymbol{\pi}_j), \dots, \boldsymbol{F}_a(\boldsymbol{\pi}_j)]'$ el vector de funciones de $\boldsymbol{\pi}_j$ para $j = 1, 2, \dots, J$ y $a \leq I(R-1)$. Estas funciones expresan la estructura relevante de los datos *dentro* de los grupos. La estructura *entre* grupos se toma en cuenta al definir un único vector $\boldsymbol{F}(\boldsymbol{\pi}) = [\boldsymbol{F}(\boldsymbol{\pi}_1)', \boldsymbol{F}(\boldsymbol{\pi}_2)', \dots, \boldsymbol{F}(\boldsymbol{\pi}_J)']'$ de funciones de probabilidades de orden $(aJ \times 1)$. Ya que se supone que la estructura *dentro* de los grupos es la misma para todos los grupos, se aplican las mismas funciones a cada uno de ellos.

Diferentes tipos de funciones pueden representarse en una manera relativamente simple usando notación matricial (Grizzle et al. 1969, Forthoper & Koch 1973). En la sección 6 se presenta un ejemplo con la función logit.

Para investigar acerca de la relación de la función de las probabilidades con las variables explicativas consideradas sobre la misma población objeto de estudio, Montero et al. (2007) proponen el siguiente modelo lineal multinivel.

Para cada uno de los J grupos se postula un modelo de nivel-1:

$$\boldsymbol{F}(\boldsymbol{\pi}_j) = \boldsymbol{X}_j \boldsymbol{\beta}_j, \quad j = 1, 2, \dots, J$$

donde $\boldsymbol{X}_j$ es una matriz de diseño de orden $(a \times t)$ y rango t; $\boldsymbol{\beta}_j = (\beta_{1j}, \beta_{2j}, \dots, \beta_{tj})'$ es un vector de parámetros aleatorios de orden $(t \times 1)$, donde β_{kj} es el coeficiente de la variable X_k en la la ecuación j.

La variabilidad de los J coeficientes $(\beta_{k1}, \beta_{k2}, \dots, \beta_{kJ})$ de la k-ésima variable $(k = 1, \dots, t)$ puede explicarse a través de un conjunto adicional de variables $Z_{1,} Z_2, \dots, Z_q$, medidas a nivel de grupo, así:

$$\beta_{kj} = \boldsymbol{Z}'_{kj} \boldsymbol{\Gamma}_k + u_{kj}, \qquad k = 1, \dots, t, \quad j = 1, 2, \dots, J \tag{1}$$

donde $\boldsymbol{\Gamma}_k$ es el vector de orden $(q_k \times 1)$ de coeficientes asociados a las variables medidas en el nivel-2, Z_{kj} representa las observaciones de las q_k variables en el nivel-2, en la j-ésima tabla y u_{kj} son los errores aleatorios no observables.

La ecuación en (1) puede describirse de una forma más compacta, de manera que el modelo en el nivel-2 se define como:

$$\boldsymbol{\beta}_j = \boldsymbol{Z}_j \boldsymbol{\Gamma} + \boldsymbol{u}_j, \quad j = 1, 2, \dots, J$$

donde $\boldsymbol{Z}_j = diag\left(\boldsymbol{Z}'_{1j}, \boldsymbol{Z}'_{2j}, \dots, \boldsymbol{Z}'_{tj}\right)$ es una matriz diagonal en bloques de orden $(t \times Q)$, con $Q = q_1 + q_2 + \dots + q_t$ y cuyos elementos son los valores de las variables explicativas en el nivel-2; $\boldsymbol{\Gamma}$ es el vector de efectos fijos de orden $(Q \times 1)$ y $\boldsymbol{u}_j$ es el vector de errores aleatorios de orden $(t \times 1)$. Se asume $E(\boldsymbol{u}_j) = 0$ y $Cov(\boldsymbol{u}_j)$ se denota por $\boldsymbol{\Omega}_{u_j}$. En la matriz de covarianza $\boldsymbol{\Omega}_{u_j}$ se utilizará $\sigma_{u_{kk^*}}$ $(k, k^* = 1, 2, \dots, t)$ para denotar las componentes de la varianza del nivel-2. Además, $Cov(\boldsymbol{u}_j, \boldsymbol{u}_{j^*}) = 0$, para $(j, j^* = 1, 2, \dots, J)$.

Una forma conveniente de expresar el modelo es mediante la formulación matricial dado en 2:

$$\boldsymbol{F}(\boldsymbol{\pi}) = \boldsymbol{A\Gamma} + \boldsymbol{Xu} \tag{2}$$

donde $\boldsymbol{A} = \left(\boldsymbol{Z}'_1 \boldsymbol{X}'_1, \boldsymbol{Z}'_2 \boldsymbol{X}'_2, \dots, \boldsymbol{Z}'_J \boldsymbol{X}'_J\right)'$, con $\boldsymbol{X}_j = \boldsymbol{X}_{j^*}$ para todo $j \neq j^*$, es una matriz de dimensión $(aJ \times Q)$ y $\boldsymbol{X} = diag(\boldsymbol{X}_1, \boldsymbol{X}_2, \dots, \boldsymbol{X}_J)$ es una matriz de dimensión $(aJ \times tJ)$. El vector $\boldsymbol{\Gamma}$ está formado por los Q efectos fijos y $\boldsymbol{u} = (u_1, u_2, \dots, u_J)'$.

Se supone que los efectos aleatorios de grupos diferentes son mutuamente independientes, $E(\boldsymbol{u}) = \boldsymbol{0}$ y $Cov(\boldsymbol{u}) = \boldsymbol{\Omega}_u$, con $\boldsymbol{\Omega}_u = \left[I_J \otimes \boldsymbol{\Omega}_{u_j}\right]$, donde $\otimes$ representa el producto de Kronecker. De aquí que $E\left[\boldsymbol{F}(\boldsymbol{\pi})\right] = \boldsymbol{A\Gamma}$ y $Var\left[\boldsymbol{F}(\boldsymbol{\pi})\right] = \boldsymbol{V}_{\boldsymbol{F}(\boldsymbol{\pi})}$, donde: $\boldsymbol{V}_{\boldsymbol{F}(\pi)} = \boldsymbol{X}\boldsymbol{\Omega}_u \boldsymbol{X}'$.

Es importante destacar que, a diferencia del caso en el que los parámetros del modelo son fijos y las inferencias conciernen sólo a los grupos especificados (clases, regiones, etc.), en el modelo multinivel para proporciones el interés de las inferencias se dirige hacia la población de grupos, no sólo a aquellos que pasan a representar la muestra. Los grupos de individuos que conforman las tablas pueden considerarse como unidades anónimas, de la misma forma que lo son las observaciones elementales.

Una vez formulado el modelo multinivel, es posible aplicar la teoría asintótica en el marco del modelo lineal general. En la próxima sección se muestran los resultados teóricos que permiten demostrar la validez de este enfoque.

4. Estimación

Para estimar los efectos fijos y los componentes de la varianza de modelos multinivel del tipo definido en la ecuación (2), se propone aplicar un enfoque basado en los mínimos cuadrados generalizados.

Supóngase que se han hecho observaciones en J grupos diferentes según la estructura de datos definida en la sección 3. Asociado a la muestra del i-ésimo subgrupo de la j-ésima tabla existe un vector de proporciones muestrales $\boldsymbol{p}_{ji} = (p_{ji1}, p_{ji2}, \ldots, p_{jiR})'$, donde $p_{jir} = n_{jir}/n_{ji}$. Sea $\boldsymbol{p}_j = [\boldsymbol{p}'_{j1}, \boldsymbol{p}'_{j2}, \ldots, \boldsymbol{p}'_{jI}]'$ la estimación muestral de $\boldsymbol{\pi}_j$. Cuando las muestras de los I subgrupos son independientes, la matriz de covarianza de $\boldsymbol{p}_j$ está dada por $\boldsymbol{V}_{\pi_j}$, la cual es una matriz diagonal en bloques de dimensión $(IR \times IR)$, con las matrices $\boldsymbol{V}_{\pi_{ji}} = 1/n_{ji}[\boldsymbol{D}_{\pi_{ji}} - \boldsymbol{\pi}_{ji}\boldsymbol{\pi}'_{ji}]$, para $i = 1, 2, \ldots, I$ y $j = 1, 2, \ldots, J$, sobre la diagonal principal, donde $\boldsymbol{D}_{\boldsymbol{\pi}_{ji}}$ es una matriz diagonal de dimensión $(I \times I)$ con elementos del vector $\boldsymbol{\pi}_{ji}$ sobre la diagonal principal.

Los J vectores de proporciones observadas en cada tabla de contingencia conforman un único vector $\boldsymbol{p} = [\boldsymbol{p}'_1, \boldsymbol{p}'_2, \ldots, \boldsymbol{p}'_J]'$ con media $\boldsymbol{\pi}$ y matriz de covarianza $V_{\boldsymbol{\pi}}$, la cual involucra a las matrices de covarianza de $\boldsymbol{p}_j$. Cuando las observaciones de diferentes grupos son independientes, la covarianza entre observaciones de diferentes grupos es cero, por tanto, la matriz de covarianza de $\boldsymbol{p}$ tiene forma de una matriz diagonal en bloques con las matrices $\boldsymbol{V}_{\boldsymbol{\pi}_j}$ sobre la diagonal principal.

Se puede demostrar por el teorema central del límite multivariado (Rao 1973, p. 128) que las propociones muestrales (y las proporciones muestrales condicionadas) tienen distribución asintóticamente normal. Por el método delta (Agresti 2002, p. 577), las funciones (y funciones condicionales) de tales proporciones, son también asintóticamente normales. El teorema 1, a continuación, especifica las distribuciones de tales funciones.

Teorema 1. *Sea $\boldsymbol{F}(\boldsymbol{p})$ la estimación muestral de $\boldsymbol{F}(\boldsymbol{\pi})$. Se supone que $\boldsymbol{F}$ tiene derivadas parciales continuas de segundo orden en una región abierta que contiene a $\boldsymbol{\pi}$. Sea $\boldsymbol{H}$ la matriz jacobiana de la función $\boldsymbol{F}(\boldsymbol{\pi})$, evaluada correspondientemente, la cual se supone no nula.*

Entonces:

$$i)\boldsymbol{F(p)} \mid \boldsymbol{u} \xrightarrow{d} N\left(\boldsymbol{A\Gamma} + \boldsymbol{Xu}, \boldsymbol{HV}_{\pi}\boldsymbol{H}'\right)$$

$$ii)\boldsymbol{F(p)} \xrightarrow{d} N\left(\boldsymbol{A\Gamma}, \boldsymbol{X\Omega}_u\boldsymbol{X}' + \boldsymbol{HV}_{\pi}\boldsymbol{H}'\right)$$

La demostración del teorema es consecuencia directa de aplicar el método delta.

En esta fase del análisis es posible postular el siguiente modelo en términos de las proporciones observadas:

$$\boldsymbol{F}(\boldsymbol{p}) = \boldsymbol{A\Gamma} + \boldsymbol{Xu} + \boldsymbol{e} \tag{3}$$

donde $\boldsymbol{e}$ se reconoce como el vector de errores aleatorios en el nivel-1. $\boldsymbol{A}$, $\boldsymbol{\Gamma}$, $\boldsymbol{X}$ y $\boldsymbol{u}$ se definen como en la ecuación en (2).

La parte aleatoria del modelo comprende dos errores: uno para cada nivel. Se supone que $\boldsymbol{u} \sim N(\mathbf{0}, \boldsymbol{\Omega}_u)$ y $\boldsymbol{e} \sim N(\mathbf{0}, \boldsymbol{H}\boldsymbol{V}_\pi\boldsymbol{H}')$, donde las matrices de covarianza para los errores de nivel-2 y los errores de nivel-1 son funciones de los parámetros $\boldsymbol{u}$ y $\boldsymbol{\pi}$, respectivamente. Sin embargo, como sugirió Goldstein (1987), se puede introducir una simplificación y requerir simplemente que las varianzas de los errores en el nivel-1 sean inversamente proporcional es a n_{ji}. Si además de la reparametrización se supone una variación simple aleatoria a través de las tablas, entonces se puede suponer que la varianza entre tablas es la misma para cada uno de los I subgrupos. En este caso $\boldsymbol{e} \sim N(\mathbf{0}, \boldsymbol{\Omega}_e)$, donde $\boldsymbol{\Omega}_e$ es una matriz diagonal con los elementos σ_e^2/n_{ji} en la diagonal principal.

El hecho de que hay más de un término de error añade una complicación a los procedimientos de estimación. Existen tres tipos de parámetros que pueden ser estimados: los efectos fijos, los coeficientes aleatorios del primer nivel (que pueden ser estimados o no, porque el modelo general que resulta de sustituir los modelos del nivel-2 en el modelo de nivel-1 no depende de ellos) y los componentes de la varianza y covarianza. A continuación se proponen los estimadores para los efectos fijos y los componentes de la varianza. El desarrollo para los efectos fijos y los componentes de la varianza hace uso de las ideas presentadas en Castells (1985).

4.1. Efectos fijos

Bajo las suposiciones de los errores del modelo en (3), éste se puede escribir como un modelo lineal con parámetros no aleatorios, tal que:

$$\boldsymbol{F}(\boldsymbol{p}) = \boldsymbol{A}\boldsymbol{\Gamma} + \boldsymbol{e}^* \quad \text{donde} \quad \boldsymbol{e}^* = \boldsymbol{X}\boldsymbol{u} + \boldsymbol{e} \tag{4}$$

Dado el modelo en (4) y de acuerdo con (Rao 1973, p. 306), se tiene que el mejor estimador lineal de $\boldsymbol{\Gamma}$ es el estimador mínimo cuadrado generalizado:

$$\widehat{\boldsymbol{\Gamma}} = \left(\boldsymbol{A}'\boldsymbol{V}_\lambda^{-1}\boldsymbol{A}\right)^{-1}\boldsymbol{A}'\boldsymbol{V}_\lambda^{-1}\boldsymbol{F}(\boldsymbol{p})$$

donde $\boldsymbol{V}_\lambda = \boldsymbol{X}\boldsymbol{\Omega}_u\boldsymbol{X}' + \boldsymbol{\Omega}_e$.

La existencia del estimador $\widehat{\boldsymbol{\Gamma}}$ está sujeta a la existencia de las inversas involucradas. Siendo el estimador $\widehat{\boldsymbol{\Gamma}}$ de la forma dada, está claro (Rao 1973, p. 307) que es insesgado para $\boldsymbol{\Gamma}$ y que

$$Var\left(\widehat{\boldsymbol{\Gamma}}\right) = \left(\boldsymbol{A}'\boldsymbol{V}_\lambda^{-1}\boldsymbol{A}\right)^{-1}$$

Sea $\overrightarrow{\mathbf{C}}$ el vector que resulta de escribir las columnas de una matriz cualquiera $\mathbf{C}$, una debajo de las otras. Entonces la matriz $\boldsymbol{V}_\lambda$ depende del parámetro $\boldsymbol{\lambda} = \begin{pmatrix} \overrightarrow{\boldsymbol{\Omega}_u} \\ \overrightarrow{\boldsymbol{\Omega}_e} \end{pmatrix}$ y si éste fuese conocido, $\widehat{\boldsymbol{\Gamma}}$ y su matriz de covarianza serían calculables. Suponer $\boldsymbol{\lambda}$ conocido significaría una restricción para la aplicación práctica de estos estimadores. Una situación mucho más realista sería suponer λ desconocido, estimar éste y sustituir las expresiones de sus estimadores en $\widehat{\boldsymbol{\Gamma}}$ logrando así un estimador en dos etapas para Γ de la forma:

$$\widehat{\widehat{\boldsymbol{\Gamma}}} = \left(\boldsymbol{A}'\widehat{\boldsymbol{V}}_{\lambda}^{-1}\boldsymbol{A}\right)^{-1}\boldsymbol{A}'\widehat{\boldsymbol{V}}_{\lambda}^{-1}\boldsymbol{F}(\boldsymbol{p})$$

donde $\widehat{\boldsymbol{V}}_{\lambda}$ es la estimación de $\boldsymbol{V}_{\lambda}$.

4.2. Componentes de la varianza

Un procedimiento para estimar las componentes de la varianza es el siguiente: el vector de errores $\boldsymbol{e}^*$ se puede escribir en la forma: $\boldsymbol{e}^* = \boldsymbol{\varpi}_1\,\boldsymbol{\xi}_1 + \boldsymbol{\varpi}_2\boldsymbol{\xi}_2$, donde: $\boldsymbol{\varpi}_1 = \boldsymbol{X}, \boldsymbol{\varpi}_2 = \boldsymbol{I}_{aJ}, \xi_1 = \boldsymbol{u}$ y $\xi_2 = \boldsymbol{e}$, con: $E(\boldsymbol{\xi}_i) = 0\ \forall\ i = 1, 2$; $Cov(\boldsymbol{\xi}_1) = \boldsymbol{\Omega}_u$; $Cov(\boldsymbol{\xi}_2) = \boldsymbol{\Omega}_e$; $Cov(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2) = Cov(\boldsymbol{\xi}_2, \boldsymbol{\xi}_1) = 0$; por tanto: $Var(\boldsymbol{e}^*) = \boldsymbol{\varpi}_1\boldsymbol{\Omega}_u\boldsymbol{\varpi}_1' + \boldsymbol{\Omega}_e$.

Queda claro que éste es un modelo de componentes de la varianza.

Sea $L(\boldsymbol{A})$ el espacio generado por las columnas de la matriz $\boldsymbol{A}$; $[L(\boldsymbol{A})]^{\perp}$ el espacio complemento ortogonal de $L(\boldsymbol{A})$; $\boldsymbol{R} = \left[\boldsymbol{I} - \boldsymbol{A}\left(\boldsymbol{A}'\boldsymbol{A}\right)^{-1}\boldsymbol{A}'\right]$ el proyector sobre $[L(\boldsymbol{A})]^{\perp}$, y $r(\mathbf{C})$ el rango de una matriz cualquiera $\mathbf{C}$. Entonces $r(\boldsymbol{R}) = aJ - r(\boldsymbol{A}) = n$. Sea $\boldsymbol{M}$ una matriz de orden $n \times aJ$ tal que $r(\boldsymbol{M}) = n$ y $\boldsymbol{MA} = \mathbf{0}$.

Defínase $\widetilde{\boldsymbol{F}} = \boldsymbol{MF}(\boldsymbol{p})$; se tiene que

$$E\left(\widetilde{\boldsymbol{F}}\right) = E(\boldsymbol{MF}(\boldsymbol{p})) = \boldsymbol{MA\Gamma} = \mathbf{0}$$

$$Cov\left(\widetilde{\boldsymbol{F}}\right) = Cov(\boldsymbol{MF}(\boldsymbol{p})) = \boldsymbol{MV}_{\lambda}\boldsymbol{M}'$$

Como $\boldsymbol{F}(\boldsymbol{p}) \xrightarrow{d} N(\boldsymbol{AF}, \boldsymbol{V}_{\lambda})$ entonces $\widetilde{\boldsymbol{F}}(\boldsymbol{p}) \xrightarrow{d} N\left(\mathbf{0}, \boldsymbol{MV}_{\lambda}\boldsymbol{M}'\right)$

Se tiene que:

$$E\left(\overrightarrow{\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}}'}\right) = \overrightarrow{\boldsymbol{MV}_{\lambda}\boldsymbol{M}'} = \left(\overrightarrow{\boldsymbol{M}\boldsymbol{\varpi}_1\boldsymbol{\varpi}_1'\boldsymbol{M}'} \vdots \overrightarrow{\boldsymbol{MM}'}\right)\begin{pmatrix}\overrightarrow{\boldsymbol{\Omega}_u}\\ \overrightarrow{\boldsymbol{\Omega}_e}\end{pmatrix} = \boldsymbol{Z}^*\boldsymbol{\lambda}$$

donde $\boldsymbol{z}^* = \left(\overrightarrow{\boldsymbol{M}\boldsymbol{\varpi}_1\boldsymbol{\varpi}_1'\boldsymbol{M}'} \vdots \overrightarrow{\boldsymbol{MM}'}\right)$ y $\boldsymbol{\lambda} = \left(\overrightarrow{\boldsymbol{\Omega}_u} \vdots \overrightarrow{\boldsymbol{\Omega}_e}\right)'$

Sea el modelo lineal dado por: $\overrightarrow{\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}'}} = \boldsymbol{Z}^*\boldsymbol{\lambda} + \boldsymbol{\mu}$, con $\boldsymbol{\mu} = \left(\overrightarrow{\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}'}} - \boldsymbol{Z}^*\boldsymbol{\lambda}\right)$, $E(\boldsymbol{\mu}) = \mathbf{0}$ y $Cov(\boldsymbol{\mu}) = \boldsymbol{V}_{\lambda}^*$, donde $\boldsymbol{V}_{\lambda}^*$ tiene la expresión:

$$\boldsymbol{V}_F^* = E\left[\left(\overrightarrow{\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}'}}\right)\left(\overrightarrow{\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}'}}\right)'\right] - \left[E\left(\overrightarrow{\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}'}}\right)E\left(\overrightarrow{\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}'}}\right)'\right]$$

pero

$$\left[\left(\overrightarrow{\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}'}}\right)\left(\overrightarrow{\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}'}}\right)'\right] = \begin{bmatrix} \widetilde{\boldsymbol{f}_1}\widetilde{\boldsymbol{f}_1}\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}'} & \widetilde{\boldsymbol{f}_1}\widetilde{\boldsymbol{f}_2}\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}'} & \dots & \widetilde{\boldsymbol{f}_1}\widetilde{\boldsymbol{f}_n}\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}'} \\ \widetilde{\boldsymbol{f}_2}\widetilde{\boldsymbol{f}_1}\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}'} & \widetilde{\boldsymbol{f}_2}\widetilde{\boldsymbol{f}_2}\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}'} & \dots & \widetilde{\boldsymbol{f}_2}\widetilde{\boldsymbol{f}_n}\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}'} \\ \vdots & \vdots & \ddots & \vdots \\ \widetilde{\boldsymbol{f}_n}\widetilde{\boldsymbol{f}_1}\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}'} & \widetilde{\boldsymbol{f}_n}\widetilde{\boldsymbol{f}_2}\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}'} & \dots & \widetilde{\boldsymbol{f}_n}\widetilde{\boldsymbol{f}_n}\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}'} \end{bmatrix}$$

donde $\widetilde{\boldsymbol{f}_i}$ es la i-ésima columna de la matriz $\widetilde{\boldsymbol{F}}$.

Calculando $E\left(\widetilde{\boldsymbol{f}_i}\widetilde{\boldsymbol{f}_j}\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}'}\right)$, empleando el momento de orden 4 de una distribución normal multivariada (ver Anderson 1958), se obtiene:

$$\begin{aligned} E\left(\widetilde{\boldsymbol{f}_i}\widetilde{\boldsymbol{f}_j}\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}'}\right) &= \tau_{ij}\boldsymbol{R} + \left[\tau_{i1}\boldsymbol{R}(j), \tau_{i2}\boldsymbol{R}(j), \dots, \tau_{in}\boldsymbol{R}(j)\right] \\ &+ \left[\tau_{j1}\boldsymbol{R}(i), \tau_{j2}\boldsymbol{R}(i), \dots, \tau_{jn}\boldsymbol{R}(i)\right] \\ &= \tau_{ij}\boldsymbol{R} + \boldsymbol{R}'(i) \otimes \boldsymbol{R}(j) + \boldsymbol{R}'(j) \otimes \boldsymbol{R}(i) \end{aligned}$$

En la expresión anterior, y con el objetivo de simplificar la formulación matemática, se ha denotado: $\boldsymbol{M}\boldsymbol{V}_\lambda\boldsymbol{M}' = \boldsymbol{R}$ y $\boldsymbol{R}(i)$ una columna cualquiera de $\boldsymbol{R}$.

Se ve fácilmente que:

$$\begin{aligned} E\left[\left(\overrightarrow{\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}'}}\right)\left(\overrightarrow{\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}'}}\right)'\right] &= \boldsymbol{R} \otimes \boldsymbol{R} + \left[\boldsymbol{R} \otimes \boldsymbol{R}(1), \boldsymbol{R} \otimes \boldsymbol{R}(2), \dots, \boldsymbol{R} \otimes \boldsymbol{R}(n)\right] \\ &+ \left[\boldsymbol{R}'(1) \otimes \overrightarrow{\boldsymbol{R}}, \boldsymbol{R}'(2) \otimes \overrightarrow{\boldsymbol{R}}, \dots, \boldsymbol{R}'(n) \otimes \overrightarrow{\boldsymbol{R}}\right] \\ &= \boldsymbol{R} \otimes \boldsymbol{R} + \left[\boldsymbol{R} \otimes \boldsymbol{R}(1), \boldsymbol{R} \otimes \boldsymbol{R}(2), \dots, \boldsymbol{R} \otimes \boldsymbol{R}(n)\right] + \overrightarrow{\boldsymbol{R}'} \otimes \overrightarrow{\boldsymbol{R}} \end{aligned}$$

Por otra parte:

$$\left[E\left(\overrightarrow{\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}'}}\right)E\left(\overrightarrow{\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}'}}\right)'\right] = \left(\overrightarrow{\boldsymbol{R}}\right)\left(\overrightarrow{\boldsymbol{R}}\right)' = \overrightarrow{\boldsymbol{R}'} \otimes \overrightarrow{\boldsymbol{R}}$$

Entonces:

$$\begin{aligned} Cov\left(\overrightarrow{\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}'}}\right) &= \left[\boldsymbol{R} \otimes \boldsymbol{R}\right] + \left[\boldsymbol{R} \otimes \boldsymbol{R}(1), \boldsymbol{R} \otimes \boldsymbol{R}(2), \dots, \boldsymbol{R} \otimes \boldsymbol{R}(n)\right] \\ &= \boldsymbol{V}^*_{\lambda_0} \end{aligned}$$

Minimizando, respecto a λ, la norma $\left\|\overrightarrow{\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}'}} - \boldsymbol{Z}^*\boldsymbol{\lambda}\right\|_{\mathbf{V}^{*-}_{\lambda_0}}$ con:

$$\begin{aligned} \boldsymbol{V}^*_{\lambda_0} &= \left[\left(\boldsymbol{M}\boldsymbol{V}_{\lambda_0}\boldsymbol{M}'\right) \otimes \left(\boldsymbol{M}\boldsymbol{V}_{\lambda_0}\boldsymbol{M}'\right)\right] \\ &+ \left[\left(\boldsymbol{M}\boldsymbol{V}_{\lambda_0}\boldsymbol{M}'\right) \otimes \left(\boldsymbol{M}\boldsymbol{V}_{\lambda_0}\boldsymbol{M}'\right)(1) \dots \left(\boldsymbol{M}\boldsymbol{V}_{\lambda_0}\boldsymbol{M}'\right) \otimes \left(\boldsymbol{M}\boldsymbol{V}_{\lambda_0}\boldsymbol{M}'\right)(n)\right] \end{aligned}$$

donde λ_0 es un valor inicial para $\boldsymbol{\lambda}$. Este valor podría obtenerse de la siguiente manera: se puede buscar el estimador mínimo cuadrado generalizado para $\boldsymbol{\Gamma}$, suponiendo $Cov\left(\boldsymbol{\xi}_1\right)=0$, y luego, examinando los residuos, obtener un valor inicial para λ.

El estimador para los componentes de la varianza tendrá la forma:

$$\widehat{\boldsymbol{\lambda}}=\left(\boldsymbol{Z}^{*'}\boldsymbol{V}_{\lambda_0}^{*-}\boldsymbol{Z}^{*}\right)\boldsymbol{Z}^{*'}\boldsymbol{V}_{\lambda_0}^{*-}\left(\overrightarrow{\widetilde{\boldsymbol{F}}\ \widetilde{\boldsymbol{F}'}}\right) \tag{5}$$

(se ha utilizado el símbolo $\boldsymbol{B}^{-}$ para indicar la inversa generalizada de una matriz $\mathbf{B}$ cualquiera). La expresión de $\widehat{\lambda}$ se da en función de la inversa generalizada porque las inversas involucradas nunca existen.

4.3. Propiedades de los estimadores

En esta sección se prueba que el estimador $\widehat{\boldsymbol{\lambda}}$ en 5 es un estimador del tipo $\boldsymbol{\lambda}_0$-MINQUE [1] propuesto por Kleffe (1976).

Teorema 2. *El estimador $\widehat{\boldsymbol{\lambda}}_{\lambda_0}$ es un estimador del tipo $\boldsymbol{\lambda}_0$-MINQUE.*

Demostración. Sólo hay que considerar el resultado del teorema 3 del anexo (dado por Castells 1985) y probar que $\widehat{\boldsymbol{\lambda}}_{\lambda_0}$ es una solución de la ecuación:

$$S_Z\left(\boldsymbol{\lambda}_0\right)\lambda=\mathbf{U}\left(\boldsymbol{\lambda}_0,\mathbf{F}\left(\mathbf{p}\right)\right)$$

donde

$$\boldsymbol{U}_i=\boldsymbol{F}\left(\boldsymbol{p}\right)'\boldsymbol{M}'\left(\boldsymbol{M}\boldsymbol{V}_{\lambda_0}\boldsymbol{M}'\right)^{-1}\boldsymbol{M}\widetilde{\mathbf{F}}_i\boldsymbol{M}'\left(\boldsymbol{M}\boldsymbol{V}_{\lambda_0}\boldsymbol{M}'\right)^{-1}\boldsymbol{M}\widetilde{\boldsymbol{F}}_jM'$$

con

$$S_M\left(\boldsymbol{\lambda}_0\right)_{ij}=tr\left(\boldsymbol{M}\boldsymbol{V}_{\lambda_0}\boldsymbol{M}'\right)^{-1}\boldsymbol{M}\widetilde{\mathbf{F}}_i\boldsymbol{M}'\left(\boldsymbol{M}\boldsymbol{V}_{\lambda_0}\boldsymbol{M}'\right)^{-1}\boldsymbol{M}\widetilde{\mathbf{F}}_j\boldsymbol{M}',\quad j=1,2$$

$\widetilde{\boldsymbol{F}}_1=\boldsymbol{I}_{aJ},\qquad \widetilde{\boldsymbol{F}}_2=\boldsymbol{\omega}_1\boldsymbol{\omega}_1'$

El elemento i del vector $\left(\boldsymbol{A}'\boldsymbol{V}_{\lambda_0}^{*-}\overrightarrow{\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}'}}\right)$ se puede escribir de la siguiente forma:

$$\begin{aligned}
\left(\boldsymbol{A}'\boldsymbol{V}_{\lambda_0}^{*-}\overrightarrow{\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}'}}\right)&=\left(\overrightarrow{\boldsymbol{M}_i\widetilde{\boldsymbol{F}}_i\boldsymbol{M}'}\right)'\boldsymbol{V}_{\lambda_0}^{*-}\overrightarrow{\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}'}}\\
&=\left(\overrightarrow{\boldsymbol{M}\widetilde{\boldsymbol{F}}_i\boldsymbol{M}'}\right)'\Theta^{-1}\overrightarrow{\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}'}}\\
&=\frac{1}{2}tr\left[\left(\boldsymbol{M}\boldsymbol{V}_{\boldsymbol{\lambda}_0}\boldsymbol{M}'\right)^{-1}\boldsymbol{M}\widetilde{\boldsymbol{F}}_i\boldsymbol{M}'\left(\boldsymbol{M}\boldsymbol{V}_{\lambda_0}\boldsymbol{M}'\right)^{-1}\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}'}\right]\\
&=\frac{1}{2}\widetilde{\mathbf{F}}'\left(\boldsymbol{M}\boldsymbol{V}_{\lambda_0}\boldsymbol{M}'\right)^{-1}\boldsymbol{M}\widetilde{\boldsymbol{F}}_i\boldsymbol{M}'\left(\boldsymbol{M}\boldsymbol{V}_{\lambda_0}\boldsymbol{M}'\right)^{-1}\widetilde{\boldsymbol{F}}
\end{aligned}$$

[1] En inglés: *minimum norm quadratic unbiased estimation.*

El elemento i-j de la matriz $\left(\boldsymbol{A}'\boldsymbol{V}_{\lambda_0}^{-}\boldsymbol{A}\right)$ viene expresado por:

$$\begin{aligned}\left(\boldsymbol{A}'\boldsymbol{V}_{\lambda_0}^{-}\boldsymbol{A}\right)_{ij} &= \left(\overrightarrow{\boldsymbol{M}\widetilde{\boldsymbol{F}}_i\boldsymbol{M}'}\right)'\boldsymbol{V}_{\boldsymbol{\lambda}_0}^{*-}\left(\overrightarrow{\boldsymbol{M}\widetilde{\boldsymbol{F}}_j\boldsymbol{M}'}\right)\\ &= \frac{1}{2}tr\left[\left(\boldsymbol{M}\widetilde{\boldsymbol{F}}_i\boldsymbol{M}'\right)\left(\boldsymbol{M}\boldsymbol{V}_{\boldsymbol{\lambda}_0}^{*}\boldsymbol{M}'\right)^{-1}\left(\boldsymbol{M}\widetilde{\boldsymbol{F}}_j\boldsymbol{M}'\right)\left(\boldsymbol{M}\boldsymbol{V}_{\boldsymbol{\lambda}_0}^{*}\boldsymbol{M}'\right)^{-1}\right]\\ &= \frac{1}{2}tr\left[\left(\boldsymbol{M}\boldsymbol{V}_{\lambda_0}^{*}\boldsymbol{M}'\right)\left(\boldsymbol{M}\widetilde{\boldsymbol{F}}_i\boldsymbol{M}'\right)^{-1}\left(\boldsymbol{M}\boldsymbol{V}_{\lambda_0}^{*}\boldsymbol{M}'\right)^{-1}\left(\boldsymbol{M}\widetilde{\mathbf{F}}_j\boldsymbol{M}'\right)\right]\end{aligned}$$

lo cual prueba que $\widehat{\lambda}_{\lambda_0}$ es un estimador del tipo definido $\boldsymbol{\lambda}_0$-MINQUE. □

5. Validación del algoritmo de estimación

Como parte de la estrategia de estimación, en este artículo se presentan (ver anexo) los pasos básicos de un algoritmo de mínimos cuadrados generalizados iterativos siguiendo el enfoque propuesto. La validez del procedimiento se ha venido explorando a través de varios estudios de simulación para conjuntos de datos balanceados (el tamaño de las muestras es el mismo para todos los subgrupos) y desbalanceados (el tamaño de las muestras de los subgrupo es diferente). En todos los casos se utilizó un modelo de regresión logística donde los logaritmos de la razón de riesgo (*log odds ratios*) se reconocen como efectos aleatorios de una población de grupos. El interés de la investigación se centró en el análisis de las estimaciones de los dos parámetros fijos y la varianza al nivel de grupo.

En el caso balanceado (Montero, Castell & Ojeda 2008) se examinó la influencia de diferentes tamaños de muestras y magnitudes de la varianza de los efectos aleatorios sobre la precisión de las estimaciones. Se consideraron setenta diseños diferentes, dados por las combinaciones de cinco números de tablas de contingencia (10, 25, 50, 75, 100), siete tamaños de muestra (10, 25, 50, 75, 100, 200, 300) de los subgrupos y dos magnitudes de varianza, una grande y otra pequeña. Los resultados indican que las estimaciones de los parámetros fijos son precisas para muestras de tamaño moderadamente pequeñas, pero se necesita un mayor número de observaciones para alcanzar un comportamiento razonable del estimador para la varianza de nivel-2. En general, para alcanzar mayor precisión en las estimaciones es más importante un número grande de individuos por subgrupos ($\geqslant 200$) que un número grande de grupos. La diferencia en precisión de las estimaciones para diseños con 50 o más tablas de contingencia no es sustancial.

En el caso desbalanceado se encontró que en ciertas situaciones, las estimaciones de la varianza producen sesgos grandes y estimaciones negativas. Para mejorar los resultados se propuso corregir el algoritmo aplicando una técnica basada en la descomposición de valores singulares truncados en la solución de los mínimos cuadrados generalizados para estimar los componentes de la varianza. Mediante simulación se mostró la efectividad de la técnica en cuanto a la reducción del sesgo de los estimadores (Montero & Guerra 2005). En el estudio se fijó el número de tablas. Los tamaños muestrales en los subgrupos se generaron a partir de tres distribuciones uniformes diferentes, dando lugar a tres tipos de diseños (ligeramente desbalanceado, moderadamente desbalanceado y muy desbalanceado). Se consideraron dos magnitudes de varianza de nivel-2, las mismas que en el caso balanceado. Para las especificaciones consideradas los resultados mostraron niveles aceptables de sesgo y precisión. Se observó además que la calidad de las estimaciones no se afectó por el grado de desbalance de los datos.

Para todos los casos se encontró que el método de estimación se comporta mejor cuando la varianza de los efectos aleatorios es pequeña.

6. Ejemplo

Para ilustrar el enfoque discutido en este artículo, se usaron los datos reportados en Turner, Omar, Yang, Goldstein & Thompson (2000), consistentes en 22 ensayos clínicos realizados para investigar el efecto de descontaminación selectiva del tracto digestivo, sobre el riesgo de infección del tracto respiratorio. Se seleccionaron aleatoriamente pacientes de unidades de cuidados intensivos, ya sea para recibir un tratamiento de combinación de antibióticos o para no recibir tratamiento. En la tabla 3 (ver anexo) se presenta la proporción de pacientes infectados en cada tratamiento para cada uno de los ensayos, así como los logaritmos de la razón de riesgo (*log odds ratios*) y sus varianzas.

Los datos pueden considerarse dentro de una estructura jerárquica, donde los subgrupos de pacientes (tratamiento y control) se anidan dentro de los ensayos. De esta manera, los subgrupos de pacientes se consideran como las unidades del nivel-1 y los ensayos, las unidades de nivel-2.

El meta-análisis de los ensayos con respuesta binaria se realizó ajustando un modelo de regresión logística en el que se fijan los efectos de los tratamientos y se permite que los logaritmos de la razón de riesgo (*log odds ratios*) varíen a través de los J ensayos (Turner et al. 2000):

$$logit\left(\pi_{ij}\right) = \beta_{0j} x_{ij} + \sum_{k=1}^{J} \beta_k D_{kij} \quad \text{en el nivel-1 (“dentro” de los ensayos)} \tag{6}$$

$$\begin{aligned} \beta_{0j} &= \theta + u_j \quad \text{en el nivel-2 (“entre” los ensayos)} \\ u_j &\sim N\left(0, \sigma_u^2\right) \end{aligned} \tag{7}$$

donde π_{ij} es la probabilidad de infectarse para los individuos del i-ésimo subgrupo en el j-ésimo ensayo, $x_{ij} = 0/1$ indica su subgrupo control/tratamiento, y el conjunto de $D_{1ij}, \ldots, D_{Jij} = 0/1$, su pertenencia al ensayo.

Sustituyendo 7 en 6 se obtiene el siguiente modelo:

$$logit\left(\pi_{ij}\right) = \left(\theta + u_j\right) x_{ij} + \sum_{k=1}^{J} \beta_k D_{kij} \tag{8}$$

donde θ (*log odds ratio*) representa el efecto promedio de interés.

Las estimaciones de los parámetros del modelo en (8), obtenidas mediante el enfoque de mínimos cuadrados generalizados (GLS[2]) propuesto en este artículo, se comparan con las mostradas por Turner et al. (2000), utilizando otros métodos disponibles en el sistema de programas MLwiN (Goldstein, Rasbash, Plewis, Draper, Browne, Yang, Woodhouse & MJR 1998): el método de cuasi-verosimilitud marginal (MQL), el método de cuasi-verosimilitud penalizada (PQL) y un método Bootstrap paramétrico (ver Goldstein & Rasbash 1996, Kuk 1995, para una discusión detallada de los métodos).

En la tabla 2 se muestran los valores estimados de θ y σ_u^2 y los correspondientes intervalos de confianza Wald. Nótese que la estimación GLS de θ es muy similar a la obtenida por el método Bootstrap de corrección de sesgos. En el caso de σ_u^2, la estimación GLS de σ_u^2 es la más extrema.

Tabla 2: Estimaciones del meta-análisis de los datos de infección del tracto respiratorio.

Método	Log OR θ	(IC 95 %)	Varianza entre ensayos σ_u^2	(IC 95 %)
MQL	$-1{,}43$	$(-1{,}80,\ -1{,}07)$	0,46	(0,08, 0,84)
PQL	$-1{,}49$	$(-1{,}90,\ -1{,}07)$	0,64	(0,14, 1,14)
PQL-Bootstrap	$-1{,}66$	$(-2{,}26,\ -1{,}05)$	0,71	(0,00, 1,13)
GLS	$-1{,}63$	$(-2{,}06,\ -1{,}04)$	1,06	(0,44, 1,69)

Para completar la comparación de enfoques, las estimaciones GLS se comparan con las obtenidas aplicando un método completamente bayesiano, que utiliza el muestreador Gibbs para Monte Carlo mediante Cadena de Markov (MCMC) (Schmidt, Spiegelhalter & Thomas 1995), y con las obtenidas mediante aproximaciones al modelo completamente bayesiano (Abrams & Sansó 1998), propuestas para reducir los cálculos requeridos. Estos métodos se pueden implementar en sistemas como el BUGS (Gilks, Thomas & Spiegelhalter 1994).

La estimación GLS de θ se presenta con el valor más extremo ($-1{,}63$ comparado con el $-1{,}49$ del completamente bayesiano y el $-1{,}50$ del bayesiano aproximado) y la estimación de σ_u^2 exhibe un valor muy cercano al obtenido mediante los otros métodos (1,06 comparado con el 1,09 del completamente bayesiano y 0,96 del bayesiano aproximado).

Como se ha podido observar, el procedimiento propuesto produjo estimaciones de los parámetros, razonablemente cercanas a las obtenidas por otros métodos ya conocidos. De esta forma, se pretende poner de manifiesto las importantes posibilidades que ofrece un método de estimación computacionalmente muy simple de aplicar en comparación con otros métodos basados en estrategias más complejas.

[2]En este artículo los métodos de estimación utilizados se identifican por sus siglas en inglés.

7. Conclusiones

El procedimiento de estimación presentado en este artículo ubica al análisis de un conjunto de tablas de contingencia en una clase de problemas que pueden tratarse utilizando mínimos cuadrados generalizados.

Una de las principales ventajas del enfoque propuesto es que puede usarse en situaciones donde otros métodos imponen la solución de complicadas expresiones matemáticas. Otra ventaja importante es la facilidad que le ofrece al investigador para construir una amplia familia de funciones de particular interés para el análisis multinivel de las proporciones esperadas en un conjunto de tablas de contingencia.

En general, el enfoque propuesto ofrece una herramienta muy útil para modelar una gran variedad de situaciones que ocurren frecuentemente en la práctica, constituyendo así una eficaz alternativa a la modelación multinivel para datos categóricos jerárquicos.

[Recibido: abril de 2009 — Aceptado: octubre de 2010]

Referencias

Abrams, K. & Sansó, B. (1998), 'Approximate Bayesian Inference for Random Effects meta-analysis', *Statistics in Medicine* **17**, 201–218.

Agresti, A. (2002), *Categorical Data Analysis*, Wiley, New York.

Anderson, T. W. (1958), *An Introduction to Multivariate Statistical Analysis*, Wiley, New York.

Breslow, N. E. & Zhao, L. P. (1988), 'Logistic Regression for Stratified Case-Control Studies', *Biometrics* **44**, 891–899.

Brik, A. S. & Raudenbush, S. W. (1992), *Hierarchical Linear Models: Applications and Data Analysis Methods*, Sage Publications, California.

Castells, E. (1985), Estimación en un modelo con parámetros aleatorios, Tesis de maestría, Facultad de Matemática, Universidad de La Habana, La Habana.

Efron, B. (1996), 'Empirical Bayes Methods for Combining Likelihoods', *Journal of the American Statistical Association* **96**(434), 538–565.

Fears, T. R. & Brown, C. C. (1986), 'Logistic Regression Methods for Retrospective Case-Control Dtudies using Complex Sampling Procedures', *Biometrics* **42**, 955–960.

Forthofer, R. N. & Lehnen, R. G. (1981), *Public Program Analysis: A New Categorical Data Approach*, Lifetime Learning Publications, Belmont, California.

Forthoper, R. N. & Koch, G. G. (1973), 'An Analysis for Compounded Functions of Categorical Data', *Biometrics* **29**, 143–157.

Fotouhi, A. R. (2003), 'Comparisons of Estimation Procedures for Nonlinear Multilevel Models', *Journal of Statistical Software* **8**(9), 1–39.

Gilks, W. R., Thomas, A. & Spiegelhalter, D. J. (1994), 'A language and Program for Complex Bayesian Modelling', *Statistician* **43**, 169–177.

Glass, G. V. (1976), 'Primary, Secondary and Meta-Analysis of Research', *Educational Researcher* **5**, 3–8.

Goldstein, H. (1987), *Multilevel Models in Educational and Social Research*, Charles Griffin, London.

Goldstein, H. (1995), *Multilevel Statistical Models*, 2 edn, Halsted Press, New York.

Goldstein, H. & Rasbash, J. (1996), 'Improved Approximations for Multilevel Models with Binary Responses', *Journal of the Royal Statistical Society. Series A* (57), 395–407.

Goldstein, H., Rasbash, J., Plewis, I., Draper, D., Browne, W., Yang, M., Woodhouse, G. & MJR, H. (1998), *A usert's guide to MLwiN*, Institute of Education.

Grizzle, J. E., Starmer, C. F. & Koch, G. G. (1969), 'Analysis of Categorical Data by Linear Models', *Biometrics* **25**, 489–504.

Hamerle, A. & Honning, G. (1995), Panel analysis for qualitative variables, *in* G. Arminger, C. C. Clogg & M. E. Sobel, eds, 'A Handbook for Statistical Modeling in the Social and Behavioral Sciences', Plenum, New York, pp. 401–451.

Hartzel, J., Liu, I.-M. & Agresti, A. (2001), 'Describing Heterogeneous Effects in Stratified Ordinal Contingency Tables, with Application to Multi-CCenter clinical trials', *Computational Statistics and Data Analysis* **35**, 429–499.

Hedges, L. V. & Olkin, I. (1985), *Statistical Methods for Meta-analysis*, Academic Press, New York.

Hsiao, C. (1995), *Analysis of Panel Data*, Cambridge University Press, New York.

Kleffe, J. (1976), 'A Note on MINQUE for Normal Models', *Mathematische Operationsforschung und Statistik* **7**, 107–114.

Kuk, A. Y. C. (1995), 'Asymtotically Unbiased Estimation in Generalized Linear Models with Random Effects', *Journal of the Royal Statistical Socity* **57**, 395–407.

Lee, Y. & Nelder, J. A. (2002), 'Analysis of Ulcer data Using Hierarchical Generalized Linear Models', *Statistics in Medicine* **21**, 191–202.

Longford, N. (1995), *Random coefficient models: Handbook of Statistiscal Modeling for the Social and Behavioral Sciences*, Plenum Press, New York.

Lubin, J. H., Blot, W. J., Berrino, F., Flamant, R., Gillis, C. R., Kunze, M., Schmäwhl, D. & Visco, G. (1984), 'Patterns of Lung Cancer Risg According to Type of Cigarrette Smoked', *International Journal of Cancer* **33**, 569–576.

Montero, M. (2006), Análisis de tablas de contingencia: un enfoque multinivel, Tesis de doctorado, Facultad de Matemática, Universidad de La Habana, La Habana.

Montero, M., Castell, E. & Ojeda, M. M. (2007), 'Fitting a Multilevel Model to a Sample of Contingency Tables using the GSK Approach', *Revista Investigación Operacional* **28**(3), 204–214.

Montero, M., Castell, E. & Ojeda, M. M. (2008), 'Analysis of a Contingency Tables sample: A Simulation Study', *Revista Ciencias Matemáticas* **24**, 83–92.

Montero, M. & Guerra, V. (2005), 'Estimating Multilevel Models for Categorical Data via Generalized Least Squares', *Revista Colombiana de Estadística.* **21**(8), 63–76.

Rao, C. R. (1973), *Linear Statistical Inference and Its Applications*, segunda edn, John Wiley & Sons, Inc., New York.

Rudas, T. (1986), 'A Monte Carlo Comparison of the Small Sample Behavior of the Pearson, the Likelihood Ratio and the Cressie-Read Statistics', *Journal of Statistical Computation and Simulation* (24), 107–120.

Schmidt, T. C., Spiegelhalter, D. J. & Thomas, A. (1995), 'Bayesian Approaches to Random'effects Meta-analysis: A Comparative Study', *Statistics in Medicine* **14**, 2685–2699.

Snijders, T. A. B. & Bosker, R. (1999), *Introduction to Basic and Advanced Multilevel Modelling*, Sage, London.

Turner, R. M., Omar, R. Z., Yang, M., Goldstein, H. & Thompson, S. G. (2000), 'Multilevel models for meta-analysis of clinical trials with binary outcomes', *Statistics in Medicine* **19**, 3417–3432.

Apéndice A.

Teorema 3. *Sean las matrices:*

$$\begin{aligned} &\boldsymbol{V}_\lambda^* = \boldsymbol{R} \otimes \boldsymbol{R} + [\boldsymbol{R} \otimes \boldsymbol{R}(1), \boldsymbol{R} \otimes \boldsymbol{R}(2), \ldots, \boldsymbol{R} \otimes \boldsymbol{R}(n)] \quad \boldsymbol{y} \\ &\Theta = 2(\boldsymbol{R} \otimes \boldsymbol{R}), \end{aligned}$$

y **A** *y* **B** *dos matrices simétricas cualesquiera de orden $n \times n$, entonces se cumple:*

$$\begin{aligned} i)\ & \boldsymbol{V}_\lambda^* \overrightarrow{\boldsymbol{B}} = \Theta \overrightarrow{\mathbf{B}} \\ ii)\ & \overrightarrow{\boldsymbol{A}}' \boldsymbol{V}_\lambda^{*-} \overrightarrow{\boldsymbol{B}} = \overrightarrow{\boldsymbol{A}}' \Theta^{-1} \overrightarrow{\boldsymbol{B}} \end{aligned} \tag{9}$$

Algoritmo de mínimos cuadrados generalizados iterativos

Tabla 3: Infecciones del tracto respiratorio en los grupos tratamiento y control de 22 ensayos para la descontaminación selectiva del tracto digestivo.

Ensayo	Infección/Total		*Odds ratio*	Log OR	Varianza (Log OR)
	Tratamiento	Control			
1	7/47	25/54	0,20	−1,59	0,24
2	4/38	24/41	0,08	−2,48	0,38
3	20/96	37/95	0,41	−0,89	0,11
4	1/14	11/17	0,04	−3,17	1,33
5	10/48	26/49	0,23	−1,46	0,21
6	2/101	13/84	0,11	−2,20	0,60
7	12/161	38/170	0,28	−1,27	0,12
8	1/28	29/60	0,04	−3,23	1,10
9	1/19	9/20	0,07	−2,69	1,26
10	22/49	44/47	0,06	−2,89	0,44
11	25/162	30/160	0,79	−0,23	0,09
12	31/200	40/185	0,66	−0,41	0,07
13	9/39	10/41	0,93	−0,07	0,28
14	22/193	40/185	0,47	−0,76	0,08
15	0/45	4/46	0,10	−2,27	2,27
16	31/131	60/140	0,41	−0,88	0,07
17	4/75	12/75	0,30	−1,22	0,36
18	31/220	42/225	0,71	−0,34	0,07
19	7/55	26/57	0,17	−1,75	0,23
20	3/91	17/92	0,15	−1,89	0,42
21	14/25	23/23	0,03	−3,62	2,20
22	3/65	6/68	0,50	−0,69	0,53

1. Obtener estimaciones iniciales de los parámetros fijos utilizando un ajuste de mínimos cuadrados generalizados para datos categóricos, asumiendo que los errores aleatorios en el nivel-2 tienen varianza 0.
2. A partir de estas estimaciones formar los residuos "crudos"
$$\widetilde{\boldsymbol{F}}(\boldsymbol{p}) = \boldsymbol{F}(\boldsymbol{p}) - \boldsymbol{A}\widehat{\boldsymbol{\Gamma}}$$
3. Calcular la matriz de productos cruzados: $\boldsymbol{F}(\boldsymbol{p})^{*} = \widetilde{\boldsymbol{F}}(\boldsymbol{p})\,\widetilde{\boldsymbol{F}}(\boldsymbol{p})'$.
4. Formar el vector: $\boldsymbol{F}(\boldsymbol{p})^{**} = \overrightarrow{\boldsymbol{F}(\boldsymbol{p})^{*}}$.
5. Similarmente, construir el vector $\overrightarrow{\boldsymbol{V}_{\lambda}}$.

6. La relación entre los vectores definidos en los pasos 4 y 5 se puede expresar mediante el modelo lineal: $E\left(\boldsymbol{F}\left(\boldsymbol{p}\right)^{**}\right) = \boldsymbol{Z}^{*}\boldsymbol{\lambda}$, donde $\boldsymbol{z}^{*}$ es la matriz de diseño para los parámetros aleatorios (o sea, los elementos de $\boldsymbol{\Omega}_u$ y $\boldsymbol{\Omega}_e$). Luego, en este paso es posible postular el modelo: $\boldsymbol{F}\left(\boldsymbol{p}\right)^{**} = \boldsymbol{Z}^{*}\boldsymbol{\lambda} + \boldsymbol{R}$.

7. Para estimar $\boldsymbol{\lambda}$ se aplica mínimos cuadrados generalizados, o sea:

$$\widehat{\lambda} = \left(\boldsymbol{Z}^{*'}\boldsymbol{V}_{\lambda}^{*-}\boldsymbol{Z}^{*}\right)^{-1}\boldsymbol{Z}^{*'}\boldsymbol{V}_{\lambda}^{*-}\boldsymbol{F}\left(\boldsymbol{p}\right)^{**}$$

donde $\boldsymbol{V}_{\lambda}^{*} = \widehat{\boldsymbol{V}}_{\lambda} \otimes \widehat{\boldsymbol{V}}_{\lambda}$.

8. Las estimaciones de $\boldsymbol{\Omega}_u$ y $\boldsymbol{\Omega}_e$ obtenidas en el paso anterior se sustituyen en:

$$\boldsymbol{V}_{\lambda} = \boldsymbol{X}\boldsymbol{\Omega}_u\boldsymbol{X}' + \boldsymbol{\Omega}_e$$

9. Se calculan nuevas estimaciones de los efectos fijos mediante el estimador mínimo cuadrado generalizado, utilizando la estimación en el paso 8 para la matriz de varianza y covarianza, obteniéndose así:

$$\widehat{\widehat{\boldsymbol{\Gamma}}} = \left(\boldsymbol{A}'\widehat{\boldsymbol{V}}_{\lambda}^{-1}\boldsymbol{A}\right)^{-1}\boldsymbol{A}'\widehat{\boldsymbol{V}}_{\lambda}^{-1}\boldsymbol{F}\left(\boldsymbol{p}\right)$$

10. Retornar al paso 2, pero utilizando $\widehat{\widehat{\boldsymbol{\Gamma}}}$ en lugar de $\widehat{\boldsymbol{\Gamma}}$.

El algoritmo alterna entre las estimaciones de los parámetros fijos y aleatorios hasta que el procedimiento converja.