

Predicción del promedio en modelos jerárquicos de dos niveles *

Fernando Velasco Luna **

Universidad Veracruzana

En este trabajo el objetivo es la predicción del promedio de la variable respuesta en datos los cuales presentan estructura jerárquica. Se trabaja con el modelo de regresión lineal de dos niveles con una variable explicatoria en cada nivel. Se propone un procedimiento heurístico y computacional para obtener la predicción del residual del nivel 2, el cual es necesario para obtener la predicción del promedio.

In this paper the goal is the prediction of the mean value of the response variable in a nonsampled unit for datum which present a hierarchical structure. A heuristic and computational procedure for predicting residuals in the second level is proposed, which is needed to obtain the predicted values for the mean.

Palabras clave: Áreas Pequeñas, Datos con Estructura Jerárquica, Modelos Lineales Jerárquicos, Teoría de la Predicción.

Keywords: Small Area, Hierarchical Structure, Hierarchical Linear Models, Prediction Theory.

1. Introducción

Los modelos lineales jerárquicos forman una clase general de modelos que permiten la modelación en una gran variedad de situaciones en las cuales se tienen datos que presentan una estructura jerárquica. Estos modelos tienen una gran variedad de aplicaciones en diversas áreas, tales como: investigación educativa, biología, investigación social, psicología, medicina, entre otras. Los modelos lineales jerárquicos tienen una gran historia, pero han recibido especial atención en los últimos años y sus áreas de aplicación se han multiplicado considerablemente [8, 9, 14, 15, 18]. Los modelos lineales jerárquicos son también conocidos en la literatura bajo una gran variedad de nombres, tales como modelos multinivel [7, 9], modelos de coeficientes aleatorios [15], modelos de componentes de la varianza y covarianza [19], o como modelos de efectos mixtos [13].

Las investigaciones en estimación en áreas pequeñas han recibido una considerable importancia en años recientes, debido a la creciente demanda de estadísticas confiables en tales áreas, que esencialmente son un subgrupo de una población de la cual la muestra ha sido obtenida. En estimación en áreas pequeñas el interés es por lo general la estimación de algún atributo del área tal como la media, un total o una proporción. La estimación en áreas pequeñas está relacionada con hacer uso de datos de una muestra usando datos de una gran población para hacer inferencia en subgrupos de la población. El objetivo es dar estimaciones confiables para áreas pequeñas, las

*Recibido el 12 de septiembre de 2007 y aceptado el 12 de septiembre de 2007

****Dirección postal:** Laboratorio de Investigación y Asesoría Estadística, LINAE Facultad de Estadística e Informática. Av. Xalapa Esq. Manuel Avila Camacho s/n. C.P. 91020. Xalapa Veracruz, México. **Correo electrónico:** fvelasco@uv.mx

cuales cuentan con pocas unidades muestreadas. Las encuestas dan poca información para un área pequeña en particular, la cual puede ser de sumo interés; ya que las encuestas por lo general son diseñadas para proporcionar estimaciones para grandes poblaciones. La estadística basada sólo en una muestra de un área pequeña es poco confiable debido al tamaño de la muestra para esta área. Con el fin de producir estimaciones en áreas pequeñas confiables, es necesario basarse en información de áreas pequeñas cercanas. (Ver [10]). Recientemente algunos investigadores [1, 2, 5, 6, 12] , han considerado el uso de modelos lineales jerárquicos para problemas de predicción en áreas pequeñas. En el área de la predicción del promedio de la variable respuesta en unidades de nivel 2, Battesse y Fuller [1, 2] proponen un modelo de regresión jerárquico, el modelo intercepto aleatorio para la estimación (o predicción) del número de hectáreas plantadas con maíz y soya. Dempster et al. [5] consideran un modelo lineal jerárquico con pendientes aleatorias. Fay y Herriot [6] hacen uso de variables explicatorias a nivel 2 con el fin de modelar el promedio de la variable respuesta. Holt y Moura [12] hicieron uso del modelo de regresión lineal jerárquico de dos niveles con una variable explicatoria a nivel 1 y de un modelo lineal jerárquico de dos niveles con variables explicatorias en ambos niveles.

En este trabajo la meta es la predicción del promedio \bar{Y} de la variable respuesta en una unidad de nivel 2 no muestreada en su variable respuesta. Se trabaja con el modelo de regresión lineal de dos niveles con una variable explicatoria en cada nivel. Se propone un procedimiento heurístico y computacional para obtener la predicción del promedio \bar{Y}_0 haciendo uso del valor de la variable explicatoria a nivel 2, w_0 . La obtención de \bar{Y}_0 se lleva a cabo tomando las k (5, 10, 15) unidades de nivel 2 más cercanas a la unidad de nivel 2 para la cual se desea obtener la predicción y tomando todas las unidades de nivel 2. La elección de estas k unidades más cercanas se hace utilizando el valor de la variable explicatoria a nivel 2, w_0 . Se proponen 5 criterios para obtener el residual u_0 , el cual es necesario para obtener la predicción del promedio \bar{Y}_0 , los cuales son: aleatorio, del mínimo, del máximo, de la mediana, y de la media aritmética.

El primer objetivo de este trabajo es saber si los tamaños de muestra de nivel 2 y de nivel 1, los valores de la varianza y de la covarianza de los errores de nivel 2, así como el criterio utilizado para obtener el residual u_0 , influyen en la elección de la cantidad k de unidades de nivel 2 que intervienen en la propuesta para llevar a cabo la predicción. En base a comparar una Estimación del Promedio del Error Cuadrático Medio de Predicción se tiene que ni los tamaños de muestra, ni los valores de la varianza y covarianza influyen en la elección de la cantidad k de unidades más cercanas, pero que el criterio sí influye en la elección de esta cantidad k . Un segundo objetivo es conocer si los tamaños de muestra de nivel 1 y de nivel 2, y los valores de la varianza y covarianza de los errores del nivel 2, influyen sobre la predicción; los resultados indican que los tamaños de muestra de nivel 1 y de nivel 2, así como el valor de la varianza sí influyen sobre la predicción, mientras que el valor de la covarianza no influye. Un tercer objetivo es conocer si el criterio el cual se utiliza para obtener el residual u_0 , y la cantidad k de unidades más cercanas influyen sobre la predicción; los resultados indican que tanto el criterio como la cantidad k de unidades más cercanas, sí influyen sobre la predicción. La conclusión principal es que al implementar la propuesta para obtener la predicción del promedio \bar{Y}_0 se deben utilizar los criterios de la mediana o de la media aritmética para obtener el residual u_0 , y que la cantidad k de unidades de nivel 2 más cercanas a utilizarse para obtener la predicción del promedio \bar{Y}_0 , cuando se hace uso de estos criterios, debe ser el de todas las unidades de nivel 2.

2. Modelo lineal general jerárquico

En esta sección se trata el modelo lineal general jerárquico. Para un tratamiento del modelo lineal general jerárquico ver [20]. Se tienen dos niveles en la estructura jerárquica, cada nivel tiene su propio submodelo.

Modelo en el nivel 1: El modelo para el nivel 1, con m variables explicatorias a nivel 1 $x_{1j}, x_{2j}, \dots, x_{mj}$ tiene la forma

$$\begin{aligned} y_{ij} &= \beta_{0j} + \beta_{1j}x_{1ij} + \beta_{2j}x_{2ij} + \dots + \beta_{mj}x_{mij} + e_{ij}, \\ i &= 1, 2, \dots, n_j; \quad j = 1, 2, \dots, J, \\ E(e_{ij}) &= 0, \quad \text{Var}(e_{ij}) = \sigma_e^2. \end{aligned} \tag{1}$$

Modelo en el nivel 2: El modelo para el nivel 2, con q variables explicatorias a nivel 2 $w_{1j}, w_{2j}, \dots, w_{qj}$, tiene la forma:

$$\begin{aligned} \beta_{0j} &= \beta_{00} + \beta_{01}w_{1j} + \beta_{02}w_{2j} + \dots + \beta_{0q}w_{qj} + u_{0j} \\ \beta_{1j} &= \beta_{10} + \beta_{11}w_{1j} + \beta_{12}w_{2j} + \dots + \beta_{1q}w_{qj} + u_{1j} \\ &\vdots \\ \beta_{mj} &= \beta_{m0} + \beta_{m1}w_{1j} + \beta_{m2}w_{2j} + \dots + \beta_{mq}w_{qj} + u_{mj}. \end{aligned} \tag{2}$$

El modelo combinado para la j -ésima unidad de nivel 2, del modelo nivel 1 con m variables explicatorias a nivel 1; y del modelo nivel 2 con q variables explicatorias a nivel 2, tiene la forma:

$$\begin{aligned} y_{ij} &= \beta_{00} + \beta_{01}w_{1j} + \beta_{02}w_{2j} + \dots + \beta_{0q}w_{qj} + u_{0j} + \\ &\quad + \beta_{10}x_{1ij} + \beta_{11}w_{1j}x_{1ij} + \dots + \beta_{1q}w_{qj}x_{1ij} + u_{1j}x_{1ij} \\ &\quad + \dots + \beta_{m0}x_{mij} + \dots + \beta_{mq}w_{qj}x_{mij} + u_{mj}x_{mij} + e_{ij}, \end{aligned} \tag{3}$$

reordenando términos se obtiene el modelo

$$\begin{aligned} y_{ij} &= \beta_{00} + \beta_{10}x_{1ij} + \dots + \beta_{m0}x_{mij} + \\ &\quad + \beta_{01}w_{1j} + \beta_{11}w_{1j}x_{1ij} + \dots + \beta_{m1}w_{1j}x_{mij} + \\ &\quad + \dots + \beta_{0q}w_{qj} + \beta_{1q}w_{qj}x_{1ij} \dots + \beta_{mq}w_{qj}x_{mij} + \\ &\quad + u_{0j} + u_{1j}x_{1ij} + \dots + u_{mj}x_{mij} + e_{ij}. \end{aligned} \tag{4}$$

Definiendo

$$\mathbf{Y}_j = \begin{bmatrix} y_{1j} \\ y_{2j} \\ \vdots \\ y_{n_jj} \end{bmatrix}; \quad \mathbf{X}_j = \begin{bmatrix} 1 & x_{11j} & x_{21j} & \dots & x_{m1j} \\ 1 & x_{12j} & x_{22j} & \dots & x_{m2j} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n_jj} & x_{2n_jj} & \dots & x_{mn_jj} \end{bmatrix};$$

$$\mathbf{e}_j = \begin{bmatrix} e_{1j} \\ e_{2j} \\ \vdots \\ e_{n_jj} \end{bmatrix}; \quad \boldsymbol{\beta}_j = \begin{bmatrix} \beta_{0j} \\ \beta_{1j} \\ \vdots \\ \beta_{mj} \end{bmatrix}$$

En forma matricial el modelo nivel 1 (1) toma la forma

$$\mathbf{Y}_j = \mathbf{X}_j \beta_j + \mathbf{e}_j; \quad j = 1, \dots, J, \quad (5)$$

donde \mathbf{Y}_j es el vector respuesta $n_j \times 1$, \mathbf{X}_j es la matriz de variables explicatorias a nivel 1 de orden $n_j \times (m+1)$, β_j es el vector de parámetros de orden $(m+1) \times 1$ y \mathbf{e}_j es un vector de errores aleatorios $n_j \times 1$. Se supone $E(\mathbf{e}_j) = \mathbf{0}$, $\text{Var}(\mathbf{e}_j) = \sigma_e^2 \mathbf{I}_{n_j}$. Definiendo

$$\mathbf{W}_j = \begin{bmatrix} 1 & w_{1j} & w_{2j} & \cdots & w_{qj} & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & \cdots & 1 & w_{1j} & w_{2j} & \cdots & w_{qj} \end{bmatrix}$$

y

$$\begin{aligned} \beta &= [\beta_{00}, \beta_{01}, \dots, \beta_{0q}, \beta_{10}, \beta_{11}, \dots, \beta_{1q}, \dots, \beta_{m0}, \beta_{m1}, \dots, \beta_{mq}]^T; \\ \mathbf{u}_j &= [u_{0j}, u_{1j}, \dots, u_{mj}]^T \end{aligned}$$

En forma matricial el modelo nivel 2 (2) tiene la forma

$$\beta_j = \mathbf{W}_j \beta + \mathbf{u}_j; \quad j = 1, \dots, J, \quad (6)$$

donde \mathbf{W}_j es la matriz de variables explicatorias a nivel 2, de orden $(m+1) \times (q+1)(m+1)$, β es el vector $(q+1)(m+1) \times 1$ de coeficientes fijos, y \mathbf{u}_j es el vector de errores aleatorios nivel 2 de orden $(m+1) \times 1$. Supongase $E(\mathbf{u}_j) = \mathbf{0}$ y

$$\text{Var}(\mathbf{u}_j) = \mathbf{\Omega} = \begin{bmatrix} \sigma_{u_0}^2 & \sigma_{u_01} & \cdots & \sigma_{u_0m} \\ \sigma_{u_01} & \sigma_{u_1}^2 & \cdots & \sigma_{u_1m} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{u_0m} & \sigma_{u_1m} & \cdots & \sigma_{um}^2 \end{bmatrix},$$

En forma matricial el modelo combinado para la j -ésima unidad de nivel 2 (4) tiene la forma

$$\begin{aligned} \mathbf{Y}_j &= \mathbf{X}_j \mathbf{W}_j \beta + \mathbf{X}_j \mathbf{u}_j + \mathbf{e}_j; \quad j = 1, \dots, J, \\ E(\mathbf{Y}_j) &= \mathbf{X}_j \mathbf{W}_j \beta, \quad \mathbf{V}_j = \text{Var}(\mathbf{Y}_j) = \mathbf{X}_j \mathbf{\Omega} \mathbf{X}_j^T + \sigma_e^2 \mathbf{I}_{n_j}. \end{aligned} \quad (7)$$

Definiendo

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_J \end{bmatrix}; \quad \mathbf{u} = \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \\ \mathbf{u}_J \end{bmatrix}; \quad \mathbf{e} = \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \vdots \\ \mathbf{e}_J \end{bmatrix}; \quad \mathbf{X} = \text{diag}(\mathbf{X}_j); \quad \mathbf{W} = \begin{bmatrix} \mathbf{W}_1 \\ \mathbf{W}_2 \\ \vdots \\ \mathbf{W}_J \end{bmatrix}$$

donde $\text{diag}(\mathbf{A}_j)$ representa los términos diagonales por matriz bloque, con \mathbf{A}_j en el j -ésimo bloque de la diagonal. Del modelo (7) se tiene

$$\mathbf{Y} = \mathbf{XW}\beta + \mathbf{Xu} + \mathbf{e}, \quad (8)$$

el cual se denomina modelo lineal general jerárquico. La matriz de varianzas y covarianzas para tal modelo tiene la forma

$$\mathbf{V} = \text{Var}(\mathbf{Y}) = \mathbf{X}\text{diag}(\boldsymbol{\Omega})\mathbf{X}^T + \text{diag}(\sigma_e^2\mathbf{I}_{n_j}) \quad (9)$$

Definiendo

$$\text{Var}(\mathbf{e}) = \mathbf{R} = \text{diag}(\sigma_e^2\mathbf{I}_{n_j}) \quad \text{y} \quad \text{Var}(\mathbf{u}) = \mathbf{G} = \text{diag}(\boldsymbol{\Omega}) \quad (10)$$

la matriz de varianzas y covarianzas para el modelo lineal general jerárquico (8) tiene la forma

$$\mathbf{V} = \text{Var}(\mathbf{Y}) = \mathbf{X}\mathbf{G}\mathbf{X}^T + \mathbf{R}. \quad (11)$$

3. Predicción en el modelo lineal general jerárquico

Un objetivo del análisis de datos con estructura jerárquica es la predicción de los efectos aleatorios. El método más conocido para la predicción es el Mejor Predictor Lineal Insesgado. Teoría de predicción en modelos jerárquicos es estudiada en [3], [4].

De Henderson (1975), (Ver [11]), el mejor predictor lineal insesgado para $\bar{\mathbf{Y}}_j = \bar{\mathbf{X}}_j\beta_j$ cuando σ_e^2 y $\boldsymbol{\Omega}$ son conocidas está dado por

$$\widehat{\bar{\mathbf{Y}}}_j = \bar{\mathbf{X}}_j\mathbf{W}_j\widehat{\beta}^* + \bar{\mathbf{X}}_j\widehat{\mathbf{u}}_j^* \quad (12)$$

donde $\widehat{\beta}^*$ es el estimador de mínimos cuadrados generalizados de β ; esto es:

$$\widehat{\beta}^* = \left(\sum_{j=1}^J \mathbf{W}_j^T \mathbf{X}_j^T \mathbf{V}_j^{-1} \mathbf{X}_j \mathbf{W}_j \right)^{-1} \left(\sum_{j=1}^J \mathbf{W}_j^T \mathbf{X}_j^T \mathbf{V}_j^{-1} \mathbf{Y}_j \right) \quad (13)$$

y

$$\widehat{\mathbf{u}}_j^* = \boldsymbol{\Omega} \mathbf{X}_j^T \mathbf{V}_j^{-1} \left(\mathbf{Y}_j - \mathbf{X}_j \mathbf{W}_j \widehat{\beta}^* \right) \quad (14)$$

Cuando σ_e^2 y $\boldsymbol{\Omega}$ son desconocidas, sean $\widehat{\sigma}_e^2$, $\widehat{\boldsymbol{\Omega}}$ y $\widehat{\beta}$ las estimaciones de Máxima Verosimilitud Restringida, así se obtiene que el promedio está dado por:

$$\widehat{\bar{\mathbf{Y}}}_j = \bar{\mathbf{X}}_j \mathbf{W}_j \widehat{\beta} + \bar{\mathbf{X}}_j \widehat{\mathbf{u}}_j, \quad (15)$$

donde

$$\begin{aligned} \widehat{\mathbf{u}}_j &= \widehat{\boldsymbol{\Omega}} \mathbf{X}_j^T \widehat{\mathbf{V}}_j^{-1} \left(\mathbf{Y}_j - \mathbf{X}_j \mathbf{W}_j \widehat{\beta} \right), \\ \widehat{\mathbf{V}}_j &= \mathbf{X}_j \widehat{\boldsymbol{\Omega}} \mathbf{X}_j^T + \widehat{\sigma}_e^2 \mathbf{I}_{n_j}. \end{aligned} \quad (16)$$

4. Procedimiento heurístico propuesto

En esta sección se presenta el procedimiento heurístico y computacional propuesto para obtener la predicción del promedio de la variable respuesta. Además se realiza un estudio por simulación para validar la propuesta.

4.1 El Algoritmo

En ocasiones no se cuenta con información de la variable de interés en la unidad de nivel 2 en la cual se desea obtener la predicción y una posible opción es la de basarse en información de unidades de nivel 2 cercanas a la unidad de interés. De la sección anterior se tiene que para obtener la predicción es necesario tener el residual del nivel 2. De (15) esta predicción del promedio está dada por

$$\widehat{\mathbf{Y}}_0 = \overline{\mathbf{X}}_0 \mathbf{W}_0 \widehat{\boldsymbol{\beta}} + \overline{\mathbf{X}}_0 \widehat{\mathbf{u}}_0 \quad (17)$$

donde \mathbf{W}_0 es la matriz de variables explicatorias a nivel 2 y $\overline{\mathbf{X}}_0$ es la matriz de promedios de las variables explicatorias a nivel 1. Para obtener la predicción $\widehat{\mathbf{Y}}_0$ se propone el siguiente algoritmo:

1. Realizar el análisis a través de la regresión jerárquica con la información disponible, con el fin de obtener los residuales del nivel 2, $\widehat{\mathbf{u}}_j = (\widehat{u}_{0j}, \widehat{u}_{1j})$
2. En base a los valores de la variable explicatoria a nivel 2, es decir, los valores w_j , los cuales son conocidos, obtener las k unidades de nivel 2 más cercanas a la unidad de nivel 2 de interés. Para este proposito se utiliza la distancia euclidiana.
3. A partir de estas k unidades de nivel 2 más cercanas a la unidad de nivel 2 de interés y de los residuales del nivel 2 $\widehat{\mathbf{u}}_j$, obtenidos en el paso 1, elegir los k residuales correspondientes a las k unidades de nivel 2 más cercanas.
4. Usando los k residuales de nivel 2 de cada una de estas k unidades seleccionadas, obtener el residual de nivel 2 de interés \mathbf{u}_0 bajo cada uno de los siguientes criterios
 - a) Una vez que se tienen los k residuales del nivel 2 $\widehat{\mathbf{u}}_1, \widehat{\mathbf{u}}_2, \dots, \widehat{\mathbf{u}}_k$, a partir de estos elegir en forma aleatoria uno el cual será el residual $\widehat{\mathbf{u}}_0$.
 - b) Elegir el mínimo de los k residuales, \widehat{u}_{0j} , así como de \widehat{u}_{1j} éstos formaran el residual $\widehat{\mathbf{u}}_0$.
 - c) Elegir el máximo de los k residuales, \widehat{u}_{0j} , así como de \widehat{u}_{1j} éstos formaran el residual $\widehat{\mathbf{u}}_0$.
 - d) Elegir la mediana de los k residuales, \widehat{u}_{0j} , así como de \widehat{u}_{1j} éstos formaran el residual $\widehat{\mathbf{u}}_0$.
 - e) Elegir la media aritmética de los k residuales \widehat{u}_{0j} , media aritmética así como de \widehat{u}_{1j} éstos formaran el residual $\widehat{\mathbf{u}}_0$.
5. A partir del residual, $\widehat{\mathbf{u}}_0$, obtenido bajo cada criterio, y haciendo uso de $\widehat{\mathbf{Y}}_0 = \overline{\mathbf{X}}_0 \mathbf{W}_0 \widehat{\boldsymbol{\beta}} + \overline{\mathbf{X}}_0 \widehat{\mathbf{u}}_0$ obtener la predicción $\widehat{\mathbf{Y}}_0$.

4.2 Estudio por simulación

Se investigó el funcionamiento del procedimiento de predicción propuesto por medio de un estudio por simulación, el cual incluyó diferentes valores para el número de unidades de nivel 2, para el número de unidades de nivel 1, para las varianzas a nivel 2, así como para la covarianza.

En la primera etapa se generaron los errores normales de nivel 2, tanto u_{0j} como u_{1j} , ambos tipos de errores se generaron con media cero y diferentes valores para sus varianzas $\sigma_{u_0}^2$, y $\sigma_{u_1}^2$, (1, 9, 25) así también se tomaron diferentes valores para la covarianza entre ellos (0.25, 0.5, 0.75). A partir de lo anterior se calculó el valor de la variable respuesta bajo el siguiente modelo de regresión de dos niveles con una variable explicatoria en cada nivel y con los coeficientes de regresión tomando los valores $\beta_{00} = 1, \beta_{10} = 1, \beta_{01} = 1$, y $\beta_{11} = 1$,

$$y_{ij} = \beta_{00} + \beta_{10}x_{1ij} + \beta_{01}w_{1ij} + \beta_{11}w_{1j}x_{1ij} + u_{0j} + u_{1j}x_{1ij} + e_{ij} \quad (18)$$

Se trabajo con 81 puntos de diseño: Unidades de nivel 2: 300, 100 y 50; Unidades de nivel 1: 10, 50 y 25; Varianzas 1, 9, 25; Covarianza 0.25, 0.5 y 0.75. En cada uno de los cuales se llevo a cabo la predicción a partir de los 5 criterios y para los 4 valores de k . Para cada uno de los puntos de diseño se realizaron 1000 replicaciones.

Con el fin de obtener la predicción se simuló una unidad de nivel 2 adicional, y las correspondientes unidades de nivel 1, se obtuvo el valor de la variable respuesta para cada una de las unidades correspondientes, y se obtuvo el promedio de la variable respuesta en esta unidad de nivel 2. Sea y_{i0} la variable respuesta de la i -ésima unidad de nivel 1, en la unidad de nivel 2 adicional, sea \bar{y}_0 el promedio de la variable respuesta en esta unidad de nivel 2. Aplicando la propuesta a las unidades de nivel 2, sin tomar en cuenta a la unidad de nivel 2 adicional ni a las unidades de nivel 1 correspondientes, se obtiene la predicción del promedio en la unidad de nivel 2 adicional, sea $\hat{\bar{y}}_0$ tal predicción. Se obtuvo el cuadrado de la diferencia entre la predicción del promedio y el promedio, es decir, se obtuvo la cantidad $(\bar{y}_0 - \hat{\bar{y}}_0)^2$. Lo anterior se realizó para cada una de las 1000 replicaciones de cada punto de diseño. Así se obtuvo $(\bar{y}_{0r} - \hat{\bar{y}}_{0r})^2$, $r = 1, 2, \dots, 1000$. De estas 1000 diferencias al cuadrado se obtuvo una Estimación del Error Cuadrático Medio de Predicción,

$$\widehat{ECMP} \left(l_{J_0, n_{j_0}, \sigma_{u_0}^2, \sigma_{u_1}^2, \sigma_{u_{01}}} , k_0, Cr_0 \right) = \frac{\sum_{r=1}^{1000} (\bar{y}_{0r} - \hat{\bar{y}}_{0r})^2}{1000}. \quad (19)$$

donde $\widehat{ECMP} \left(l_{J_0, n_{j_0}, \sigma_{u_0}^2, \sigma_{u_1}^2, \sigma_{u_{01}}} , k_0, Cr_0 \right)$ denota la estimación del error cuadrático medio de predicción correspondiente al punto de diseño $l_{J_0, n_{j_0}, \sigma_{u_0}^2, \sigma_{u_1}^2, \sigma_{u_{01}}}$, en el cual se tienen J_0 unidades de nivel 2, n_{j_0} unidades de nivel 1, un valor de la varianza $\sigma_{u_0}^2$, un valor de la covarianza $\sigma_{u_{01}}$, tomando las k_0 , unidades más cercanas y utilizando el criterio Cr_0 .

5. Resultados

En esta sección se presentan en primer lugar los resultados obtenidos del estudio por simulación respecto a la influencia de los tamaños de muestra, de los valores de las varianzas y covarianza y del criterio utilizado, en la elección de la cantidad k de unidades más cercanas. En segundo lugar se presentan los resultados de la influencia de los tamaños de muestra, de los valores de las varianzas y covarianza, de la cantidad k de unidades más cercanas y del criterio utilizado, sobre la predicción.

5.1 Cantidad k de unidades más cercanas

Se calculó el promedio de la estimación del error cuadrático medio de predicción para cada valor de J , para cada uno de los valores de k y cada uno de los criterios, el cual se basó en los 27 puntos de diseño del estudio por simulación para los cuales J tomó ese valor, así el promedio de la estimación del error cuadrático medio de predicción para cada valor de J está dado por

$$\widehat{PECMP}(J, k_0, Cr_0) = \frac{\sum_{l=1}^{27} \widehat{ECMP}(l_{J_0, n_{j_0}, \sigma_{u_0}^2, \sigma_{u_1}^2, \sigma_{u_{01}}}, k_0, Cr_0)}{27}. \quad (20)$$

donde $\widehat{PECMP}(J, k_0, Cr_0)$ denota el promedio de $\widehat{ECMP}(l_{J_0, n_{j_0}, \sigma_{u_0}^2, \sigma_{u_1}^2, \sigma_{u_{01}}}, k_0, Cr_0)$ correspondiente al punto de diseño l cuando el número de unidades de nivel 2 es J , la cantidad de unidades más cercanas es k_0 , y el criterio a utilizar es Cr_0 .

La tabla 5.1 presenta el $\widehat{PECMP}(J, k_0, Cr_0)$ para las cantidades J de unidades de nivel 2, cada uno de los valores de la cantidad k y para cada criterio.

Criterio	J	$k=5$	$k=10$	$k=15$	Todas
Aleatorio	300	16.28	16.37	16.18	16.28
	100	12.65	12.35	12.63	12.43
	50	11.68	11.67	11.99	11.97
Máximo	300	10.45	15.80	19.98	61.25
	100	14.52	21.89	27.03	55.54
	50	15.50	22.81	27.98	44.57
Mínimo	300	37.64	51.44	60.12	127.9
	100	26.03	37.07	44.49	81.52
	50	20.62	29.70	35.70	54.62
Mediana	300	17.60	5.74	5.34	4.23
	100	4.23	2.46	2.04	1.03
	50	3.51	1.78	1.37	0.55
Media Aritmética	300	6.48	5.29	4.89	4.21
	100	3.16	1.99	1.61	0.96
	50	2.52	1.36	0.97	0.42

Tabla 5.1. $\widehat{ECMP}(J, k_0, Cr_0)$ para el número de unidades de nivel 2.

De la tabla 5.1 se tiene que cuando se utiliza el criterio aleatorio para obtener el residual $\hat{\mathbf{u}}_0$ la elección de la cantidad k de unidades más cercanas depende del número J de unidades de nivel 2. Sin embargo la diferencia entre los valores del $\widehat{PECMP}(J, k_0, Cr_0)$ no es mucha. Mientras que en cada uno de los demás criterios el menor valor del $\widehat{PECMP}(J, k_0, Cr_0)$ se presenta para la misma cantidad k de unidades utilizadas. Aunque esta cantidad k no es la misma para todos los criterios. Por lo anterior se tiene que el número J de unidades de nivel 2 no influye en la elección de la cantidad k de unidades más cercanas al implementar la propuesta.

El cálculo del promedio de la estimación del error cuadrático medio de predicción para cada valor de n_j está dado por

$$\widehat{PECMP}(n_j, k_0, Cr_0) = \frac{\sum_{l=1}^{27} \widehat{ECMP}(l_{J_0, n_j, \sigma_{u_0}^2, \sigma_{u_1}^2, \sigma_{u_{01}}}, k_0, Cr_0)}{27}. \quad (21)$$

La tabla 5.2 presenta el $\widehat{PECMP}(n_j, k_0, Cr_0)$ para las diferentes cantidades n_j de unidades de nivel 1, cada uno de los valores de la cantidad k utilizadas al implementar la propuesta y para cada criterio.

Criterio	n_j	$k=5$	$k=10$	$k=15$	Todas
Aleatorio	100	13.46	13.07	13.49	13.36
	50	13.32	13.60	13.28	13.38
	25	13.83	13.72	14.02	13.94
Máximo	100	12.09	17.24	20.93	41.06
	50	13.84	20.90	25.83	54.81
	25	14.53	22.36	28.23	65.48
Mínimo	100	23.53	32.42	38.34	67.97
	50	27.97	39.64	47.13	88.29
	25	32.78	46.15	54.83	107.8
Mediana	100	4.88	3.08	2.68	1.67
	50	5.03	3.29	2.88	1.94
	25	5.43	3.61	3.19	2.20
Media Aritmética	100	3.80	2.61	2.24	1.60
	50	3.98	2.87	2.48	1.87
	25	4.36	3.17	2.76	2.12

Tabla 5.2. $\widehat{ECMP}(n_j, k_0, Cr_0)$ para el número de unidades de nivel 1.

De la tabla 5.2 se tiene que si se utiliza el criterio aleatorio para obtener el residual $\hat{\mathbf{u}}_0$ la elección de la cantidad k de unidades más cercanas depende del número n_j de unidades de nivel 1, pero los valores del $\widehat{PECMP}(n_j, k_0, Cr_0)$ varían poco. Mientras que en cada uno de los demás criterios el menor valor del $\widehat{PECMP}(n_j, k_0, Cr_0)$ se presenta para la misma cantidad k de unidades. Aunque esta cantidad k no es la misma para todos los criterios. Así se tiene que el número n_j de unidades de nivel 1 no influye en la elección de la cantidad k de unidades más cercanas.

El cálculo del promedio de la estimación del error cuadrático medio de predicción para cada valor de σ_u^2 de la varianza está dado por

$$\widehat{PECMP}(\sigma_u^2, k_0, Cr_0) = \frac{\sum_{l=1}^{27} \widehat{ECMP}(l_{J_0, n_{j_0}, \sigma_{u_0}^2, \sigma_{u_1}^2, \sigma_{u_{01}}, k_0, Cr_0)}{27}. \quad (22)$$

La tabla 5.3 presenta el $\widehat{PECMP}(\sigma_u^2, k_0, Cr_0)$ para cada uno de los valores σ_u^2 de la varianza, cada uno de los valores de la cantidad k y para cada criterio.

Criterio	σ_u^2	$k=5$	$k=10$	$k=15$	Todas
Aleatorio	1	2.54	2.55	2.54	2.55
	9	10.94	10.09	10.82	11.04
	25	27.13	26.94	27.44	27.09
Máximo	1	0.86	0.88	0.99	2.02
	9	9.14	13.74	17.12	37.80
	25	30.46	45.88	56.88	121.5
Mínimo	1	5.53	7.05	7.95	12.79
	9	23.77	33.12	39.05	72.42
	25	55.00	78.05	93.31	178.9
Mediana	1	1.89	1.75	1.71	1.64
	9	4.35	2.99	2.67	1.92
	25	9.10	5.24	4.37	2.26
Media Aritmética	1	1.81	1.72	1.68	1.63
	9	3.55	2.65	2.35	1.87
	25	6.80	4.29	3.45	2.09

Tabla 5.3. $\widehat{ECMP}(\sigma_u^2, k_0, Cr_0)$ para los valores de la varianza.

De la tabla 5.3 se tiene que no importando el valor de la varianza la elección de la cantidad k de unidades más cercanas al implementar la propuesta es la misma en cada uno de los criterios. Aunque esta cantidad k no es la misma para todos los criterios. Así se tiene que el valor de la varianza no influye en la elección de la cantidad k de unidades más cercanas.

El cálculo del promedio de la estimación del error cuadrático medio de predicción para cada valor de $\sigma_{u_{01}}$ de la covarianza está dado por

$$\widehat{PECMP}(\sigma_{u_{01}}, k_0, Cr_0) = \frac{\sum_{l=1}^{27} \widehat{ECMP}(l_{J_0, n_{j_0}, \sigma_{u_0}^2, \sigma_{u_1}^2, \sigma_{u_{01}}, k_0, Cr_0)}{27}. \quad (23)$$

La tabla 5.4 presenta el $\widehat{PECMP}(\sigma_{u_{01}}, k_0, Cr_0)$ para cada uno de los valores $\sigma_{u_{01}}$ de la covarianza, cada uno de los valores de la cantidad k y para cada criterio.

Criterio	σ_{u01}	$k=5$	$k=10$	$k=15$	Todas
Aleatorio	0.25	13.73	13.10	13.43	13.66
	0.5	13.48	13.60	13.72	13.53
	0.75	13.52	13.61	13.66	13.64
Máximo	0.25	13.27	19.92	24.73	53.76
	0.5	13.61	20.36	25.23	54.11
	0.75	13.56	20.18	25.08	53.49
Mínimo	0.25	28.00	39.41	46.64	87.33
	0.5	28.06	39.35	46.75	88.08
	0.75	28.12	39.46	46.86	88.63
Mediana	0.25	5.08	3.27	2.85	1.84
	0.5	5.17	3.40	2.98	1.98
	0.75	5.11	3.31	2.91	1.99
Media Aritmética	0.25	4.01	2.82	2.42	1.77
	0.5	4.09	2.93	2.54	1.91
	0.75	4.06	2.90	2.51	1.92

Tabla 5.4. $\widehat{ECMP}(\sigma_{u01}, k_0, Cr_0)$ para los valores de la covarianza.

De la tabla 5.4 se tiene que en cada uno de los criterios el menor valor del $\widehat{ECMP}(\sigma_{u01}, k_0, Cr_0)$ se presenta para la cantidad k de unidades más cercanas. Aunque esta cantidad k no es la misma para todos los criterios. Por lo anterior se tiene que el valor de la covarianza no influye en la elección de la cantidad k de unidades más cercanas al implementar la propuesta.

La tabla 5.5 presenta el $\widehat{PECMP}(k_0, Cr_0)$ para cada uno de los valores de la cantidad k de unidades más cercanas y para cada criterio.

Criterio	$k=5$	$k=10$	$k=15$	Todas
Aleatorio	13.54	13.46	13.60	13.56
Máximo	13.49	20.17	25.00	53.78
Mínimo	28.09	39.40	46.77	88.03
Mediana	5.11	3.33	2.92	1.94
Media Aritmética	4.05	2.88	2.49	1.86

Tabla 5.5. $\widehat{ECMP}(k_0, Cr_0)$ para cada criterio y cada valor de la cantidad k .

De la tabla 5.5 se tiene que cuando se utiliza el criterio aleatorio para obtener el residual \hat{u}_0 la diferencia entre los valores del $\widehat{PECMP}(k_0, Cr_0)$ para los valores de la cantidad k es poca, por lo que se elige cualquier cantidad k para llevar a cabo la predicción. Cuando se utiliza el criterio del máximo o del mínimo se eligen las 5 unidades más cercanas para llevar a cabo la predicción, mientras que cuando se utiliza el criterio de la mediana o de la media aritmética se eligen todas las unidades de nivel 2 para llevar a cabo la predicción. Por lo anterior se tiene que el criterio que se utiliza para obtener el residual \hat{u}_0 si influye en la elección de la cantidad k de unidades más cercanas utilizadas al implementar la propuesta.

5.2 Predicción

De la tabla 5.1 se tiene que cuando se utilizan los criterios aleatorio, del mínimo, del de la mediana o de la media aritmética, no importando la cantidad k de unidades más cercanas, el menor valor del $\widehat{PECMP}(J, k_0, Cr_0)$ se presenta cuando se tienen 50 unidades de nivel 2 y que al incrementarse el tamaño de muestra de nivel 2 se incrementa el valor del $\widehat{PECMP}(J, k_0, Cr_0)$. Mientras que cuando se hace uso del criterio del máximo, no importando la cantidad k de unidades más cercanas, el menor valor del $\widehat{PECMP}(J, k_0, Cr_0)$ se presenta cuando se tienen 300 unidades de nivel 2 y que al incrementarse el tamaño de muestra de nivel 2 disminuye el valor del $\widehat{PECMP}(J, k_0, Cr_0)$. Así se tiene que el número de unidades de nivel 2 si influye sobre la predicción.

De la tabla 5.2 se tiene que cuando se hace uso del criterio aleatorio no existe diferencia entre los valores del $\widehat{PECMP}(n_j, k_0, Cr_0)$. Para todos los demás criterios, no importando la cantidad k de unidades más cercanas, el menor valor del $\widehat{PECMP}(n_j, k_0, Cr_0)$ se presenta cuando se tienen 100 unidades de nivel 1 y que al incrementarse el tamaño de muestra de nivel 1 disminuye el valor del $\widehat{PECMP}(n_j, k_0, Cr_0)$. Así se tiene que el número de unidades de nivel 1 si influye sobre la predicción.

De la tabla 5.3 se tiene que cuando el valor de la varianza es igual a 1, el valor del $\widehat{PECMP}(\sigma_u^2, k_0, Cr_0)$ es menor y que este valor se incrementa a medida que se incrementa el valor de la varianza. Así se tiene que el valor de la varianza si influye sobre la predicción.

De la tabla 5.4 se tiene que el valor del $\widehat{ECMP}(\sigma_{u01}, k_0, Cr_0)$ es semejante para cualquier valor de la covarianza en cada uno de los criterios. Aunque este valor del $\widehat{ECMP}(\sigma_{u01}, k_0, Cr_0)$ no es el mismo para todos los criterios. Así se tiene que el valor de la covarianza no influye sobre la predicción.

La tabla 5.6 presenta el $\widehat{PECMP}(k_0, Cr_0)$ para cada uno de los valores de la cantidad k de unidades más cercanas y para cada criterio.

k	Aleatorio	Máximo	Mínimo	Mediana	Media Aritmética
5	13.54	13.49	28.10	5.11	4.05
10	13.46	20.17	39.40	3.33	2.88
15	13.60	24.99	46.77	2.92	2.49
Todas	13.56	53.78	88.03	1.94	1.86

Tabla 5.6. $\widehat{ECMP}(k_0, Cr_0)$ para valores de la cantidad k por criterio.

Respecto a la influencia de la cantidad k de unidades más cercanas se tiene de la tabla 5.6 que cuando se eligen todas las unidades de nivel 2 se presentan los menores valores del $\widehat{PECMP}(k_0, Cr_0)$ cuando se utilizan los criterios de la mediana o de la media aritmética, mientras que cuando se eligen las 5 unidades más cercanas se presentan los menores valores del $\widehat{PECMP}(k_0, Cr_0)$ cuando se utilizan los criterios

del mínimo o del máximo. Así se tiene que la cantidad k de unidades más cercanas si influye sobre la predicción.

Al considerar la información de la tabla 5.5 y 5.6 se tiene que cuando se utilizan los criterios del máximo o del mínimo al incrementarse la cantidad k de unidades más cercanas se incrementa el valor del $\widehat{PECMP}(k_0, Cr_0)$, mientras que cuando se utilizan los criterios de la mediana o de la media aritmética al incrementarse la cantidad k de unidades más cercanas el valor del $\widehat{PECMP}(k_0, Cr_0)$ disminuye.

6. Conclusiones

En primer lugar se tiene que ni los tamaños de muestra, ni los valores de la varianza y covarianza influyen en la elección de la cantidad k de unidades más cercanas, pero que el criterio utilizado para obtener la predicción de \mathbf{u}_0 si influye.

Los resultados indican que los tamaños de muestra de nivel 1 y de nivel 2, así como el valor de la varianza si influyen sobre la predicción, mientras que el valor de la covarianza no influye.

Respecto a la influencia del criterio el cual se utiliza para obtener la predicción de \mathbf{u}_0 , y de la cantidad k de unidades de nivel más cercanas, sobre la predicción; los resultados indican que tanto el criterio como la cantidad k de unidades más cercanas, si influyen sobre la predicción.

La conclusión principal es que al implementar la propuesta para obtener la predicción del promedio $\bar{\mathbf{Y}}_0$ se deben utilizar los criterios de la mediana o de la media aritmética para obtener el residual \mathbf{u}_0 , y que la cantidad k de unidades de nivel 2 a utilizarse para obtener la predicción del promedio $\bar{\mathbf{Y}}_0$, cuando se hace uso de estos criterios, debe ser el de todas las unidades de nivel 2.

Referencias

- [1] Battese, G. E. and Fuller, W. A. "Prediction of County Group Area using Survey and Satellite Data". *Survey section proceedings, American Statistical* , 1981, 500-505.
- [2] Battese, G. E. Harter, R. M. and Fuller, W. A. "An Error-Component Model for Prediction of County Crop Areas using Survey and Satellite Data". *Journal of the American Statistical Association*, **83**, 1988, 28-36.
- [3] Bolfarine, H. and Zacks, S. *Prediction Theory for Finite Populations*. Springer Verlag, New York, 1992.
- [4] Christensen, R. *Plane Answers to Complex Questions: The Theory of Linear Models*. Springer Verlag, New York, 1996.
- [5] Dempster, A.P., Rubin, D.B. and Tsutakawa, R.K. "Estimation in Covariance Components Models". *Journal of the American Statistical Association* , **76** 1981, 341-353.
- [6] Fay, R.E., and Herriot, R.A. "Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data". *Journal of the American Statistical Association* , **74** 1979, 269-277.
- [7] Goldstein, H. "Efficient Statistical Modeling of Longitudinal Data". *Annals of Human Biology*, **13**(2), 1986, 129-141.

- [8] Goldstein, H. *Multilevel Models in Educational and Social Research*. Griffin, London, 1987.
- [9] Goldstein, H. *Multilevel Statistical Models*. Second Edition. Halsted Press, New York, 1995.
- [10] Ghosh, M., and Rao, J. N. K. "Small Area Estimation: An Appraisal". *Statistical Science*, **9**, 1994, 55-93.
- [11] Henderson, C.R., "Best Linear Unbiased Estimation and Prediction under a Selection Model". *Biometrics*, **31**, 1975, 423-447.
- [12] Holt, D. and Moura F. A. S., "Small Area Estimation Using Multi-level Models". *Survey Methodology*, **25**, 1999, 73-80.
- [13] Laird, N. M. and Ware, J. H. "Random Effects Models for Longitudinal Data". *Biometrics*, **38**, 1982, 963-974.
- [14] Longford, N. T. *Random Coefficient Models*. Oxford: University Press, New York. 1993.
- [15] Longford, N. T. *Random Coefficient Models*. In *Handbook of Statistical Models for the Social and Behavioral Sciences*, Plenum Press, New York, 1995.
- [16] Prasad, N. G. N. and Rao, J. N. K. "On the Estimation of Mean Squared Error of Small Area Predictors". *Technical Report 97*. Carleton University, Laboratory for Research in Statistics and Probability: Ottawa. 1986.
- [17] Prasad, N. G. N. and Rao, J. N. K. "The Estimation of the Mean Squared Error of Small-Area Estimators". *Journal of the American Statistical Association*, **85**, 1990, 163-171.
- [18] Raudenbush, S. W. and Bryk, A. S. *Hierarchical Linear Models: Applications and Data Analysis Methods*. Sage Publications, NewburyPark, 2002.
- [19] Searle, S. R., Casella, G., and McCulloch, C. E. *Variance Components*. John Wiley, New York, 1992.
- [20] Velasco-Luna, F. "Modelo Lineal General Jerárquico". *Revista de Ciencias Básicas, UJAT*, **5(2)**:20-28, Diciembre 2006.