Representación del BLUP en el Modelo Lineal General Mixto, Caso Balanceado

Fernando Velasco Luna[1], Mario Miguel Ojeda Ramirez[2]
[1] Facultad de Estadística e Informática. UV. fvelasco@uv.mx
[2] Facultad de Estadística e Informática. UV. mojeda@uv.mx

Resumen
El modelo lineal general mixto engloba una clase amplia de modelos que permiten tratar situaciones con datos de estructura jerárquica. Es de vital importancia para realizar procesos de inferencia usando estos modelos contar con resultados de estimación y prueba de hipótesis para los efectos fijos y aleatorios. Aunque en la literatura se conocen suficientes resultados acerca de la teoría del álgebra lineal relacionada con la teoría de estimación y de prueba de hipótesis en el modelo lineal general (MLG), hace falta material que muestre la relación existente del álgebra lineal y la teoría de la predicción en el modelo lineal general mixto (MLGM) $Y = X\beta + Zu + e$. En este trabajo se considera la representación del Mejor Predictor Lineal Insesgado (*BLUP*) de un efecto mixto en términos del operador el proyector orthogonal $P_z = Z(Z^t Z)^{-1} Z^t$ sobre el espacio $S(Z)$ y del operador proyector oblicuo $P_{X;XV^{-1}} = X(X^t V^{-1} X)^{-} X^t V^{-1}$, sobre el espacio $S(X)$. Se presenta la condición que debe cumplir la matriz de diseño Z bajo la cual el $BLUP$ de un efecto mixto se puede expresar en términos de los operadores mencionados. Se consideran casos particulares del modelo lineal general mixto con dos componentes de la varianza, con lo cual estos resultados se pueden aplicar a modelos particulares.

Palabras Clave: Componentes de la varianza, efecto mixto, modelación multinivel, modelo lineal jerárquico, operador proyector.

1. Introducción

Los modelos lineales jerárquicos forman una clase general que permite abordar situaciones con datos que presentan una estructura jerárquica, en diversas áreas de aplicación como: investigación educativa (efectividad de escuela, logro escolar), biología (curvas de crecimiento, estudios genéticos), investigación social (análisis de encuestas, estudios de mercado), psicología (análisis de conducta), medicina (ajuste de datos de medidas repetidas y estudios de centros hospitalarios), entre otras. Recientes desarrollos en cómputo han permitido que se incremente el uso de estos modelos, que son también conocidos como modelos multinivel (De Leeuw, 2008), modelos de componentes de la varianza y covarianza (Searle *et al*., 2006), o como modelos de efectos mixtos (West *et al*, 2007). Un tratamiento y abundantes referencias acerca de estos modelos se pueden encontrar en (De Leeuw, 2008, Gelman and Hill, 2007; West *et al*, 2007).

La teoría de espacios vectoriales de dimensión finita proporciona un marco para trabajar conceptos de inferencia en el MLG $Y = X\beta + e$, $E(e) = 0$, $Cov(e) = \sigma^2 I$ donde $Y \epsilon R^n$, X es una matriz de constantes de orden $n \times p$, $\beta \epsilon R^p$ es un vector de parámetros desconocidos, y $e \epsilon R^n$ es un vector de errores aleatorios no observables. Conceptos como subespacio columna, operador proyector ortogonal u oblicuo, matriz inversa generalizada, entre otros, juegan un papel de suma importancia en el estudio del MLG. Los proyectores asociados con un estimador en particular en el MLG juegan un papel importante en la caracterización de las propiedades estadísticas del estimador. En el criterio de mínimos cuadrados para definir el mejor ajuste, el estimador de mínimos cuadrados ordinarios del vector de coeficientes en el MLG se puede expresar en términos del operador proyector orthogonal $P_x = X(X^tX)^{-1}X^t$ sobre el espacio $S(X)$; así mismo el estimador de mínimos cuadrados generalizados se puede expresar en términos del operador proyector oblicuo $P_{X:XV^{-1}} = X(X^tV^{-1}X)^{-}X^tV^{-1}$. Para una revisión de la aplicación del operador proyector en modelos lineales ver Christensen (2002). Kala y Pordzik, (2006) presentan propiedades de operadores lineales y su relación con el Mejor Estimador Lineal Insesgado (*BLUE*). Peng *et al*. (2005) presentan algunas propiedades del operador oblicuo.

Considérese el modelo $Y = X\beta + e$, $E(e) = 0$, $Cov(e) = \sigma^2 I$ donde $Y \epsilon R^n$, X es una matriz de constantes de orden $n \times p$, $\beta \epsilon R^p$ es un vector de parámetros desconocidos, y $e \epsilon R^n$ es un vector de errores aleatorios no observables. Es conocido que $\hat{\beta}$ es un estimador mínimos cuadrados de β si, y sólo si, $X\hat{\beta} = P_X Y$. Si se considera el modelo $Y = X\beta + e$, $E(e) = 0$, $Cov(e) = \sigma^2 I$, donde V es una matriz definida positiva, entonces $\hat{\beta}$ es un estimador mínimos cuadrados de β si, y sólo si, $X\hat{\beta} = P_{X:XV^{-1}}Y$; en este caso $\hat{\beta}$ se denomina el estimador mínimos cuadrados generalizado de β (ver Christensen (2002) para demostraciones).

2. Mejor Predictor Lineal Insesgado y Estimaciónde Efectos Mixtos

Considérese el MLGM

$$Y = X\beta + Zu + e \tag{1}$$

donde $Y \epsilon R^n$, X y Z son matrices de constantes de orden $n \times p$ y $n \times q$ respectivamente, $\beta \epsilon R^p$ es un vector de parámetros fijos desconocidos, $u \epsilon R^q$ es un vector de efectos aleatorios, $u \sim N(0,\ \sigma_u^2 I_m)$, y $e \epsilon R^n$ es un vector de errores aleatorios no observables, $e \sim N(0,\ \sigma_e^2 I_n)$. La matriz de varianzas y covarianzas de Y está dada por:

$$V = Cov(Y) = \sigma_u^2 ZZ^t + \sigma_e^2 I\,. \tag{2}$$

En una serie de trabajos Henderson desarrolla el método del mejor predictor lineal insesgado para modelos mixtos. Concretamente Henderson *et al*. (1959) desarrollaron un conjunto de ecuaciones que simultáneamente proporcionan el *BLUE* de $X\beta$ y el *BLUP* de u . Estas ecuaciones son derivadas por la maximización de la densidad conjunta de Y y u, la cual está dada para, $Var(e) = R$ y $Var(u) = G$ por:

$$f(Y,u) = \frac{\exp\left\{-\left[(Y - X\beta - Zu)^t R^{-1}(Y - X\beta - Zu) + u^t G^{-1} u\right]/2\right\}}{(2\pi)^{\frac{(N+q.)}{2}}|R|^{1/2}|G|^{1/2}}. \tag{3}$$

Igualando a cero las derivadas parciales de (3) con respecto a los elementos de β y de u, se obtienen las ecuaciones del MLGM

$$\begin{bmatrix} X^t R^{-1} X & X^t R^{-1} Z \\ Z^t R^{-1} X & Z^t R^{-1} Z + G^{-1} \end{bmatrix} \begin{bmatrix} \beta \\ u \end{bmatrix} = \begin{bmatrix} X^t R^{-1} Y \\ Z^t R^{-1} Y \end{bmatrix}. \tag{4}$$

Las estimaciones pueden ser escritas como:

$$\hat{\beta} = \left(X^t V^{-1} X\right)^{-1} X^t V^{-1} Y; \qquad \hat{u} = GZ^t V^{-1}\left(Y - X\hat{\beta}\right). \tag{5}$$

En este contexto es necesaria la estimación de funciones de la forma $k^t\beta + m^t u$, para vectores específicos de constantes κ y m. Henderson (1975) obtiene el $BLUP$ de $k^t\beta + m^t u$ que está dado por:

$$BLUP\left(k^t\beta + m^t u\right) = k^t\hat{\beta} + m^t\hat{u}. \tag{6}$$

A continuación se presenta la condición que debe cumplir la matriz de diseño Z bajo la cual el $BLUP$ del efecto mixto $X\beta + Zu$ se puede expresar en términos de los operadores $P_{X:VX^{-1}}$, y P_Z, sobre los subespacios $S(X)$ y $S(Z)$ respectivamente.

Teorema 1. Bajo el modelo (1), si existe una constante d tal que $ZZ^t = dP_Z$, entonces la inversa V^{-1} de V se puede expresar en términos del operador proyector ortogonal P_z y de su complemento Q_z por medio de $V^{-1} = \frac{P_Z}{(d\sigma_u^2+\sigma_e^2)} + \frac{Q_Z}{\sigma_e^2}$.

Demostración: Se tiene que bajo el modelo (1), si existe una constante d tal que $ZZ^t = dP_z$, entonces

$$\begin{aligned} V &= \sigma_u^2 ZZ^t + \sigma_e^2 I = d\sigma_u^2 P_Z + \sigma_e^2\left(P_Z + Q_Z\right) \\ &= d\sigma_u^2 P_Z + \sigma_e^2 P_Z + \sigma_e^2 Q_Z = \left(d\sigma_u^2 + \sigma_e^2\right)P_Z + \sigma_e^2 Q_Z, \end{aligned}$$

de lo cual la inversa V^{-1} se puede expresar por medio de $V^{-1} = \frac{P_Z}{(d\sigma_u^2+\sigma_e^2)} + \frac{Q_Z}{\sigma_e^2}$.

Teorema 2. Bajo el modelo (1), si existe una constante d tal que $ZZ^t = dP_z$, entonces el $BLUP$ del efecto mixto $X\beta + Zu$ se puede expresar en términos de los operadores $P_{X:VX^{-1}}$ y P_z, por medio de $P_{X:VX^{-1}}Y + cP_zY - cP_ZP_{X:VX^{-1}}Y$, donde $c = \frac{d\sigma_u^2}{(d\sigma_u^2+\sigma_e^2)}$.

Demostración: Por la teoría de Henderson el $BLUP$ de $X\beta + Zu$ está dado por medio de $X\hat{\beta} + Z\hat{u}$ donde $\hat{\beta}$ es el estimador de mínimos cuadrados generalizados de β y $\hat{u}$ es el $BLUP$ de u. Se sabe que el estimador mínimos cuadrados generalizado se relaciona con el operador proyector $P_{X:VX^{-1}}$ por medio de $X\hat{\beta} = P_{X:VX^{-1}}Y$, y que $\hat{u} = GZ^tV^{-1}\left(Y - X\hat{\beta}\right)$. Así el $BLUP$ de $X\beta + Zu$ se puede expresar como:

$$
\begin{aligned}
X\hat{\beta}+Z\hat{u} &= P_{X:VX^{-1}}Y + ZGZ^{t}V^{-1}\left(Y - X\hat{\beta}\right) \\
&= P_{X:VX^{-1}}Y + \sigma_u^2 ZZ^t\left(\frac{P_Z}{\left(d\sigma_u^2+\sigma_e^2\right)}+\frac{Q_Z}{\sigma_e^2}\right)\left(Y - X\hat{\beta}\right) \\
&= P_{X:VX^{-1}}Y + d\sigma_u^2 P_Z\left(\frac{P_Z}{\left(d\sigma_u^2+\sigma_e^2\right)}+\frac{Q_Z}{\sigma_e^2}\right)\left(Y - X\hat{\beta}\right) \\
&= P_{X:VX^{-1}}Y + \frac{d\sigma_u^2 P_Z}{\left(d\sigma_u^2+\sigma_e^2\right)}\left(Y - P_{X:VX^{-1}}Y\right) \\
&= P_{X:VX^{-1}}Y + c\,P_Z Y - cP_Z P_{X:VX^{-1}}Y,
\end{aligned}
$$

donde $c = \frac{d\sigma_u^2}{(d\sigma_u^2+\sigma_e^2)}$.

3. Aplicaciones a Algunos Modelos Particulares

Casos particulares del MLGM que satisfacen la condición $ZZ^t = dP_Z$ son: (a) *Modelo solo intercepto sin variables explicatorias*. Este es el caso más simple de un modelo lineal jerárquico; el modelo para la i-ésima unidad de nivel 1 en la j-ésima unidad de nivel 2 está dado por:

$$y_{ij} = \mu + u_j + e_{ij}, \quad i = 1,\dots,d, \quad j = 1,\dots,k, \tag{7}$$

donde μ es un parámetro fijo u_j ; es el efecto aleatorio; u_j y e_{ij} son independientes, con $u \sim N(0,\ \sigma_u^2)$ y $e \sim N(0,\ \sigma_e^2)$. El modelo para la j-ésima unidad de nivel 2 es

$$y_j = 1_d\,\mu + 1_d u_j + e_j, \quad j = 1,\dots,k, \tag{8}$$

tomando $\beta = \mu$, $X = 1_k \otimes 1_d$, $Z = I_k \otimes 1_d$ y $u = (u_1,\dots,u_k)^t$ el modelo (8) es de la forma $y = X\beta + Zu + e$. En este caso se tiene que:

$$
\begin{aligned}
ZZ^t &= (I_k \otimes 1_d)(I_k \otimes 1_d)^t \\
&= (I_k \otimes 1_d)(I_k^t \otimes 1_d^t) \\
&= (I_k I_k^t \otimes 1_d 1_d^t) \\
&= (I_k \otimes 1_d 1_d^t);
\end{aligned}
$$

así:

$$\begin{aligned}
P_Z &= Z\left(Z^t Z\right)^{-} Z^t = \left(I_k \otimes 1_d\right)\left[\left(I_k \otimes 1_d\right)^t \left(I_k \otimes 1_d\right)\right]^{-} \left(I_k \otimes 1_d\right)^t \\
&= \left(I_k \otimes 1_d\right)\left[\left(I_k^t \otimes 1_d^t\right)\left(I_k \otimes 1_d\right)\right]^{-} \left(I_k \otimes 1_d\right)^t = \left(I_k \otimes 1_d\right)\left[\left(I_k^t I_k \otimes 1_d^t 1_d\right)\right]^{-} \left(I_k \otimes 1_d\right)^t \\
&= \left(I_k \otimes 1_d\right)\left[\left(I_k \otimes 1/d\right)\right]\left(I_k \otimes 1_d\right)^t = \left[\left(I_k \otimes 1_d/d\right)\right]\left(I_k \otimes 1_d\right)^t \\
&= I_k \otimes \frac{1_d 1_d^t}{d} = \frac{1}{d}\left(I_k \otimes 1_d 1_d^t\right) = \frac{1}{d} Z Z^t,
\end{aligned}$$

cumpliéndose la condición $dP_Z = ZZ^t$.

(b) *Modelo sólo intercepto con variables explicatorias*. Considérese el modelo:

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + u_j + e_{ij}, \quad i = 1,\ldots,d, \quad j = 1,\ldots,k, \tag{9}$$

donde β_0, β_1 y β_2 son parámetros fijos; u_j es el efecto aleatorio; u_j y e_{ij} son independientes, con $u_j \sim N\left(0, \sigma_u^2\right)$ y $e_{ij} \sim N\left(0, \sigma_e^2\right)$. El modelo para la j-ésima unidad de nivel 2 tiene la forma

$$y_j = 1_d \beta_0 + X_{1j}\beta_1 + X_{2j}\beta_2 + 1_d u_j + e_j, \quad j = 1,\ldots,k, \tag{10}$$

tomando $\beta = \left(\beta_0, \beta_1, \beta_2\right)^t$, $X_j = \left(1_d \vdots X_{1j} \vdots X_{2j}\right)$, $Z_j = 1_d$, el modelo (10) toma la forma $y_j = X_j\beta + Z_j u_j + e_j$, $j = 1,..,k$. Definiendo a X_1 y a X_2 como las matrices de las variables explicatorias, y $X = \left(1_k \otimes 1_d \vdots X_1 \vdots X_2\right)$, $Z = I_k \otimes 1_d$ y $u = \left(u_1,\ldots,u_k\right)^t$ el modelo (10) es de la forma $y = X\beta + Zu + e$, y se cumple la condición $ZZ^t = dP_Z$.

(c) *Modelo de dos vías*. Este modelo es de la forma

$$y_{ij} = \mu + \alpha_j + \beta_i + e_{ij}, \quad i = 1,\ldots,d, \quad j = 1,\ldots,k, \tag{11}$$

donde μ y β_i son parámetros fijos; α_j es el efecto aleatorio; α_j y e_{ij} son independientes, con $\alpha_j \sim N\left(0, \sigma_\alpha^2\right)$ y $e_{ij} \sim N\left(0, \sigma_e^2\right)$. Definiendo $\beta = \left(\mu, \beta_1,\ldots,\beta_d\right)^t$, $X_j = \left(1_d \vdots I_d\right)$, $Z_j = 1_d$ y $u_j = \alpha_j$ el modelo para la j-ésima unidad de nivel 2 está dado por medio de $y_j = X_j\beta + Z_j u_j + e_j$, $j = 1,\ldots,k$. Tomando $\beta = \left(\mu, \beta_1,\ldots,\beta_d\right)^t$, $u = \left(\alpha_1,\ldots,\alpha_k\right)$, $X = \left(1_k \otimes 1_d \vdots 1_k \otimes I_d\right)$ y $Z = I_k \otimes 1_d$ el modelo (11) está dado por medio de $y = X\beta + Zu + e$. En este caso también se cumple la condición $ZZ^t = dP_Z$.

4. Aplicación a datos reales

La representación del *BLUP* obtenida en términos de los operadores se aplico a datos relacionados con las calificaciones de estudiantes, se obtuvo el *BLUP* del efecto mixto en términos del operador proyector, y se comparo con el *BLUP* obtenido en base al resultado de Henderson. Los datos GPA son un conjunto de datos longitudinales, donde 100 estudiantes de una escuela han tenido seguimiento durante los 6 semestres de su escolaridad. El modelo usado fue el modelo solo intercepto sin variables explicatorias:

$$y_{ij} = \mu + u_j + e_{ij},\ \text{con}\ u_j \sim N\left(0, \sigma_u^2\right)\ \text{y}\ e_{ij} \sim N\left(0, \sigma_e^2\right).$$

Se realizó la predicción del efecto mixto $X\beta + Zu$ obtenida en base a la teoría de Henderson del BLUP $X\hat{\beta} + Z\hat{u}$ y a la basada en proyectores $P_{X;VX^{-1}}Y + cP_zY - cP_zP_{X;VX^{-1}}Y$, encontrándose que no existe diferencia entre las predicciones obtenidas.

5. Conclusiones

Se hace la representación del mejor predictor lineal insesgado del efecto mixto $X\beta + Zu$, basado en el MLGM, en términos de los operadores $P_{X;VX^{-1}}$ y Pz, sobre los subespacios $S(X)$ y $S(Z)$, respectivamente. La predicción obtenida se compara con la predicción obtenida basada en la teoría de Henderson, haciendo uso del modelo solo intercepto sin variables explicatorias, encontrándose que no existe diferencia entre las predicciones obtenidas. Está representación permitirá poder estudiar desde el punto de vista del álgebra lineal la teoría de predicción en el MLGM y las propiedades distribucionales de los predictores.

REFERENCIAS

Christensen, R. (2002). *Plane Answers to Complex Questions. The Theory of Linear Models*. 3rd Ed. New York: Springer.

Gelman, A. and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.

Henderson, C.R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics* **31**, 423-447.

Henderson, C.R., Kempthorne, O., Searle, S.R. and von Krosigk, C.N. (1959). Estimation of environmental and genetic trends from records subject to culling. *Biometrics*. **15**, 192-218.

Kala, R. and Pordzik. (2006). Two operators and the BLUE. *Linear Algebra and its Applications* **417**, 134-139.

De Leeuw, J. and Meijer, E. (2008). *Handbook of Multilevel Analysis*. New York: Springer.

Peng, C. and Zhang, X. (2005). On recursive oblique projectors. *IEEE Signal Processing Letters* **12**, 433-436.

Searle, S.R., Casella, G. and McCulloch, C.E. (2006). *Variance Components*. New York: John Wiley.

West B. T., Welch K. B., and Galecki, A. T. (2007). *Linear Mixed Models. A Practical Guide Using Statistical Software*. Boca Raton: Chapman & Hall/CRC.