

MAT-468: Modelos lineales con efectos mixtos

Felipe Osorio

fosorios.mat.utfsm.cl

Departamento de Matemática, UTFSM



- ▶ En la práctica, usualmente tenemos datos **desbalanceados**:
 - ▶ distinto número de mediciones por individuo.
 - ▶ mediciones no son tomadas en periodos de tiempo fijos.
- ▶ Técnicas de **regresión multivariadas** no son aplicables.
- ▶ Perfiles longitudinales pueden ser bien aproximados usando **regresión lineal**.



Esto lleva a una **formulación en 2-etapas**:

- Etapa 1:** Modelos de regresión lineal para cada individuo (**sujeto-específico**) separadamente.
- Etapa 2:** Explicar la variabilidad en los coeficientes de regresión sujeto-específico usando covariables.



Etapas 1:

- ▶ Respuestas Y_{ij} para el i -ésimo individuo medidas en el tiempo t_{ij} , $i = 1, \dots, n$; $j = 1, \dots, m_i$.

- ▶ Vector de respuestas \mathbf{Y}_i para el i -ésimo individuo

$$\mathbf{Y}_i = (Y_{i1}, \dots, Y_{im_i})^T.$$

- ▶ Modelo para la etapa 1:

$$\mathbf{Y}_i = \mathbf{Z}_i \boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i.$$

donde \mathbf{Z}_i es matriz de diseño $n_i \times q$ y $\boldsymbol{\beta}_i$ es un vector q -dimensional de coeficientes de regresión sujeto-específicos.

- ▶ Supuesto distribucional:

$$\boldsymbol{\epsilon}_i \sim N_{m_i}(\mathbf{0}, \mathbf{R}_i),$$

con $\mathbf{R}_i = \mathbf{R}_i(\boldsymbol{\tau})$. Frecuentemente $\mathbf{R}_i = \sigma^2 \mathbf{I}_{m_i}$.

- ▶ Note que el modelo anterior describe la variabilidad observada **dentro** de los individuos.



Etapla 2:

- ▶ Variabilidad **entre-individuos** puede ser estudiada relacionando β_i a covariables.
- ▶ Modelo para la etapa 2:

$$\beta_i = K_i \beta + b_i,$$

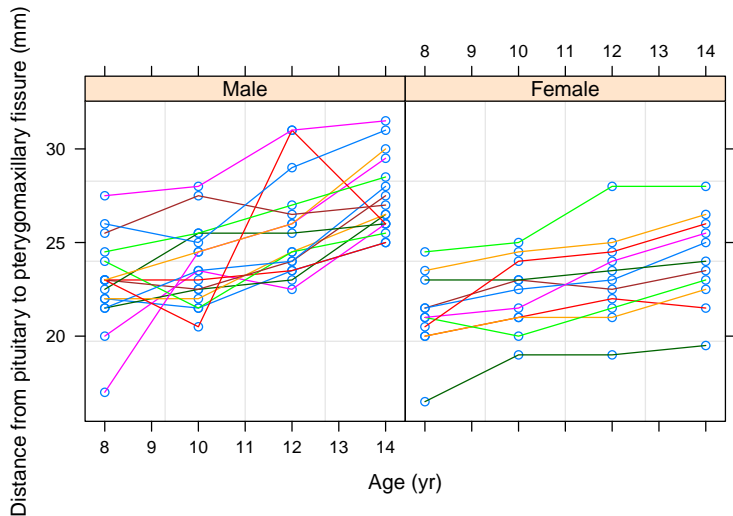
donde K_i es matriz de diseño $q \times p$.

- ▶ β es vector (común) de parámetros de regresión desconocidos.
- ▶ Supuesto distribucional:

$$b_i \sim N_q(0, \Psi).$$



Estudio de Ortodoncia (Potthoff y Roy, 1964)



- Observaciones tomadas en tiempos $t = 8, 10, 12, 14$.

- Modelo para la Etapa 1:

$$Y_{ij} = \beta_{0i} + \beta_{1i}\text{Sexo} + \beta_{2i}(\text{edad}_j - 11) + \beta_{3i}(\text{edad}_j - 11)\text{Sexo} + \epsilon_{ij},$$

donde $i = 1, \dots, 27$; $j = 1, \dots, 4$, con Sexo una variable indicadora,

$$\text{Sexo} = \begin{cases} 1, & \text{si femenino,} \\ 0, & \text{si masculino.} \end{cases}$$

- Notación matricial

$$Y_i = Z_i\beta_i + \epsilon_i.$$



- En la segunda etapa, los parámetros sujeto-específicos siguen el modelo:

$$\begin{aligned}\beta_{0i} &= \beta_0 + b_{1i}, & \beta_{1i} &= \beta_1 + b_{1i} \\ \beta_{2i} &= \beta_2 + b_{2i}, & \beta_{3i} &= \beta_3 + b_{2i},\end{aligned}$$

donde

$$\mathbf{b}_i = \begin{pmatrix} b_{1i} \\ b_{2i} \end{pmatrix} \sim N_2(\mathbf{0}, \Psi), \quad \Psi = \begin{pmatrix} \psi_{11} & \psi_{12} \\ \psi_{21} & \psi_{22} \end{pmatrix}$$

- Notación matricial:

$$\beta_i = \beta + \mathbf{b}_i.$$



- ▶ Parámetros de regresión β son de **interés primario**.
- ▶ Análisis en 2-etapas puede ser llevado a cabo de **forma explícita**.
- ▶ $\hat{\beta}_i$ son obtenidos del ajuste en la 1ra etapa, esto permite estimar β en la segunda etapa.
- ▶ Información contenida en Y_i es perdida al “resumirla” en $\hat{\beta}_i$.
- ▶ Se introduce variabilidad en el modelo para la etapa 2 al reemplazar β_i por $\hat{\beta}$.



El análisis en el modelo puede ser llevado a cabo de forma eficiente al **combinar ambas etapas**:

$$\begin{aligned}Y_i &= Z_i \beta_i + \epsilon_i, \\ \beta_i &= K_i \beta + b_i,\end{aligned}$$

esto es,

$$\begin{aligned}Y_i &= Z_i K_i \beta + Z_i b_i + \epsilon_i \\ &= X_i \beta + Z_i b_i + \epsilon_i,\end{aligned}$$

con $X_i = Z_i K_i$ matriz de diseño $m_i \times p$.



Definición 1: (Modelo lineal con efectos mixtos)

Considere

$$Y_i = X_i\beta + Z_i b_i + \epsilon_i, \quad i = 1, \dots, n,$$

donde

$$b_i \sim N_q(\mathbf{0}, \Psi), \quad \epsilon_i \sim N_{m_i}(\mathbf{0}, R_i), \quad b_i \perp \epsilon_i,$$

con X_i , Z_i matrices de diseño $m_i \times p$ y $m_i \times q$, respectivamente.

Terminología:

- ▶ El parámetro **común**, β es llamado vector de **efectos fijos**,
- ▶ “parámetros” **sujeto-específicos** b_i se denominan **efectos aleatorios**,
- ▶ elementos en Ψ y R_i son conocidos como **componentes de varianza**.

¹Biometrics 38, 963-974.



Modelos lineales con efectos mixtos

El modelo lineal con efectos mixtos (lme)

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \quad i = 1, \dots, n,$$

con

$$\mathbf{b}_i \sim N_q(\mathbf{0}, \boldsymbol{\Psi}), \quad \boldsymbol{\epsilon}_i \sim N_{m_i}(\mathbf{0}, \mathbf{R}_i), \quad \mathbf{b}_i \perp \boldsymbol{\epsilon}_i.$$

Puede ser re-escrito usando una **formulación jerárquica** como:

$$\mathbf{Y}_i | \mathbf{b}_i \stackrel{\text{ind}}{\sim} N_{m_i}(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i, \mathbf{R}_i), \quad \mathbf{b}_i \stackrel{\text{ind}}{\sim} N_q(\mathbf{0}, \boldsymbol{\Psi}),$$

para $i = 1, \dots, n$.

El modelo lme también es conocido como:

- ▶ jerárquico,
- ▶ multinivel,
- ▶ en dos etapas.



El modelo lme puede ser re-expresado como:

$$\begin{aligned} Y_i &= X_i\beta + Z_i b_i + \epsilon_i \\ &= X_i\beta + u_i, \quad u_i = Z_i b_i + \epsilon_i, \end{aligned}$$

como $b_i \perp \epsilon_i$, tenemos

$$u_i \sim N_{m_i}(0, V_i), \quad V_i = Z_i \Psi Z_i^T + R_i.$$

Finalmente, obtenemos el **modelo marginal**:

$$Y_i \sim N_{m_i}(X_i\beta, Z_i \Psi Z_i^T + R_i).$$



- ▶ De ahí que, supuestos muy específicos se realizan sobre la **media y covarianza marginal**.
 - ▶ Media implicada: $X_i\beta$,
 - ▶ Covarianza implicada: $V_i = Z_i\Psi Z_i^T + R_i$.
- ▶ Note que el modelo jerárquico implica el marginal y **NO** viceversa.



Modelos lineales con efectos mixtos

Considere el modelo jerárquico:

$$\mathbf{Y}_i | \mathbf{b}_i \stackrel{\text{ind}}{\sim} N_{m_i}(\mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i, \sigma^2 \mathbf{I}_{m_i}), \quad \mathbf{b}_i \stackrel{\text{ind}}{\sim} N_q(\mathbf{0}, \boldsymbol{\Psi}), \quad \boldsymbol{\Psi} = \boldsymbol{\Psi}(\boldsymbol{\theta}),$$

para $i = 1, \dots, n$.

La **función de verosimilitud** para el modelo lme asume la forma:

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2) = f(\mathbf{y}; \boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2),$$

donde $\mathbf{Y} = (\mathbf{Y}_1^T, \dots, \mathbf{Y}_n^T)^T$ es vector de respuestas N -dimensional, con $N = \sum_{i=1}^n m_i$.

Debido a que \mathbf{b}_i , $i = 1, \dots, n$, son **no observables**. Una alternativa es:

$$\begin{aligned} L(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2) &= \prod_{i=1}^n f(\mathbf{y}_i; \boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2) \\ &= \prod_{i=1}^n \int_{\mathbb{R}^q} f(\mathbf{y}_i | \mathbf{b}_i; \boldsymbol{\beta}, \sigma^2) f(\mathbf{b}_i; \boldsymbol{\theta}, \sigma^2) d\mathbf{b}_i \end{aligned}$$



Tenemos

$$f(\mathbf{y}_i | \mathbf{b}_i; \boldsymbol{\beta}, \sigma^2) = (2\pi\sigma^2)^{-n_i/2} \exp\{-\|\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i\|^2/2\sigma^2\},$$

y considere el **factor de precisión relativa**

$$\frac{\boldsymbol{\Psi}^{-1}}{1/\sigma^2} = \boldsymbol{\Delta}^T \boldsymbol{\Delta},$$

esto permite escribir:

$$\begin{aligned} f(\mathbf{b}_i; \boldsymbol{\theta}, \sigma^2) &= (2\pi)^{-q/2} |\boldsymbol{\Psi}|^{-1/2} \exp\{-\mathbf{b}_i^T \boldsymbol{\Psi}^{-1} \mathbf{b}_i/2\} \\ &= (2\pi\sigma)^{-q/2} |\boldsymbol{\Delta}^T \boldsymbol{\Delta}|^{1/2} \exp\{-\mathbf{b}_i^T \boldsymbol{\Delta}^T \boldsymbol{\Delta} \mathbf{b}_i/2\sigma^2\} \\ &= (2\pi\sigma)^{-q/2} |\boldsymbol{\Delta}| \exp\{-\|\boldsymbol{\Delta} \mathbf{b}_i\|^2/2\sigma^2\}. \end{aligned}$$



De este modo, obtenemos la verosimilitud

$$\begin{aligned} L(\beta, \theta, \sigma^2) &= \prod_{i=1}^n \frac{|\Delta|}{(2\pi\sigma^2)^{m_i/2}} \int_{\mathbb{R}^q} \frac{\exp\{-(\|Y_i - X_i\beta - Z_i b_i\|^2 + \|\Delta b_i\|^2)/2\sigma^2\}}{(2\pi\sigma^2)^{q/2}} db_i \\ &= \prod_{i=1}^n \frac{|\Delta|}{(2\pi\sigma^2)^{m_i/2}} \int_{\mathbb{R}^q} \frac{\exp\{-\|\tilde{Y}_i - \tilde{X}_i\beta - \tilde{Z}_i b_i\|^2/2\sigma^2\}}{(2\pi\sigma^2)^{q/2}} db_i \end{aligned}$$

donde

$$\tilde{Y}_i = \begin{pmatrix} Y_i \\ \mathbf{0} \end{pmatrix}, \quad \tilde{X}_i = \begin{pmatrix} X_i \\ \mathbf{0} \end{pmatrix}, \quad \tilde{Z}_i = \begin{pmatrix} Z_i \\ \Delta \end{pmatrix}.$$

son conocidos como **pseudo-datos**.



Note que

$$\begin{aligned}\|\tilde{\mathbf{Y}}_i - \tilde{\mathbf{X}}_i\boldsymbol{\beta} - \tilde{\mathbf{Z}}_i\mathbf{b}_i\|^2 &= \|\tilde{\mathbf{Y}}_i - \tilde{\mathbf{X}}_i\boldsymbol{\beta} - \tilde{\mathbf{Z}}_i\hat{\mathbf{b}}_i - \tilde{\mathbf{Z}}_i(\mathbf{b}_i - \hat{\mathbf{b}}_i)\|^2 \\ &= \|\tilde{\mathbf{Y}}_i - \tilde{\mathbf{X}}_i\boldsymbol{\beta} - \tilde{\mathbf{Z}}_i\hat{\mathbf{b}}_i\|^2 + (\mathbf{b}_i - \hat{\mathbf{b}}_i)^T \tilde{\mathbf{Z}}_i^T \tilde{\mathbf{Z}}_i (\mathbf{b}_i - \hat{\mathbf{b}}_i)\end{aligned}$$

donde

$$\hat{\mathbf{b}}_i = (\tilde{\mathbf{Z}}_i^T \tilde{\mathbf{Z}}_i)^{-1} \tilde{\mathbf{Z}}_i^T (\tilde{\mathbf{Y}}_i - \tilde{\mathbf{X}}_i\boldsymbol{\beta}).$$

Luego

$$\begin{aligned}&\int_{\mathbb{R}^q} \frac{\exp\{-\|\tilde{\mathbf{Y}}_i - \tilde{\mathbf{X}}_i\boldsymbol{\beta} - \tilde{\mathbf{Z}}_i\mathbf{b}_i\|^2/2\sigma^2\}}{(2\pi\sigma^2)^{q/2}} d\mathbf{b}_i \\ &= \exp\left\{-\frac{1}{2\sigma^2}\|\tilde{\mathbf{Y}}_i - \tilde{\mathbf{X}}_i\boldsymbol{\beta} - \tilde{\mathbf{Z}}_i\hat{\mathbf{b}}_i\|^2\right\} \int_{\mathbb{R}^q} \frac{\exp\{-(\mathbf{b}_i - \hat{\mathbf{b}}_i)^T \tilde{\mathbf{Z}}_i^T \tilde{\mathbf{Z}}_i (\mathbf{b}_i - \hat{\mathbf{b}}_i)/2\sigma^2\}}{(2\pi\sigma^2)^{q/2}} d\mathbf{b}_i\end{aligned}$$



Por otro lado

$$\begin{aligned} & \frac{|\tilde{\mathbf{Z}}_i^T \tilde{\mathbf{Z}}_i|^{1/2}}{|\tilde{\mathbf{Z}}_i^T \tilde{\mathbf{Z}}_i|^{1/2}} \int_{\mathbb{R}^q} \frac{\exp\{-(\mathbf{b}_i - \hat{\mathbf{b}}_i)^T \tilde{\mathbf{Z}}_i^T \tilde{\mathbf{Z}}_i (\mathbf{b}_i - \hat{\mathbf{b}}_i)/2\sigma^2\}}{(2\pi\sigma^2)^{q/2}} d\mathbf{b}_i \\ &= \frac{1}{|\tilde{\mathbf{Z}}_i^T \tilde{\mathbf{Z}}_i|^{1/2}} \int_{\mathbb{R}^q} \frac{\exp\{-(\mathbf{b}_i - \hat{\mathbf{b}}_i)^T \tilde{\mathbf{Z}}_i^T \tilde{\mathbf{Z}}_i (\mathbf{b}_i - \hat{\mathbf{b}}_i)/2\sigma^2\}}{(2\pi\sigma^2)^{q/2} / |\tilde{\mathbf{Z}}_i^T \tilde{\mathbf{Z}}_i|^{1/2}} d\mathbf{b}_i \\ &= \frac{1}{|\tilde{\mathbf{Z}}_i^T \tilde{\mathbf{Z}}_i|^{1/2}} = \frac{1}{|\mathbf{Z}_i^T \mathbf{Z}_i + \boldsymbol{\Delta}^T \boldsymbol{\Delta}|^{1/2}}. \end{aligned}$$

Combinando los resultados anteriores, tenemos que la integral requerida asume la forma

$$\begin{aligned} & \int_{\mathbb{R}^q} \frac{\exp\{-\|\tilde{\mathbf{Y}}_i - \tilde{\mathbf{X}}_i\boldsymbol{\beta} - \tilde{\mathbf{Z}}_i\mathbf{b}_i\|^2/2\sigma^2\}}{(2\pi\sigma^2)^{q/2}} d\mathbf{b}_i \\ &= |\tilde{\mathbf{Z}}_i^T \tilde{\mathbf{Z}}_i|^{-1/2} \exp\left\{-\frac{1}{2\sigma^2} \|\tilde{\mathbf{Y}}_i - \tilde{\mathbf{X}}_i\boldsymbol{\beta} - \tilde{\mathbf{Z}}_i\hat{\mathbf{b}}_i\|^2\right\}. \end{aligned}$$

Obteniendo finalmente

$$\begin{aligned} L(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2) &= (2\pi\sigma^2)^{-N/2} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n \|\tilde{\mathbf{Y}}_i - \tilde{\mathbf{X}}_i\boldsymbol{\beta} - \tilde{\mathbf{Z}}_i\hat{\mathbf{b}}_i\|^2\right\} \\ &\times \prod_{i=1}^n \frac{|\boldsymbol{\Delta}|}{|\mathbf{Z}_i^T \mathbf{Z}_i + \boldsymbol{\Delta}^T \boldsymbol{\Delta}|^{1/2}}. \end{aligned}$$

Por otro lado, para el **modelo marginal**:

$$\mathbf{Y}_i \sim N_{m_i}(\mathbf{X}_i\boldsymbol{\beta}, \mathbf{Z}_i\boldsymbol{\Psi}\mathbf{Z}_i^T + \mathbf{R}_i), \quad i = 1, \dots, n,$$

tenemos que la función de verosimilitud está dada por:

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2) = (2\pi\sigma^2)^{-N/2} \prod_{i=1}^n |\mathbf{V}_i|^{-1/2} \\ \times \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta})^T \mathbf{V}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}) \right\}.$$

Observación:

Cuando $\boldsymbol{\Psi}$ y \mathbf{R}_i son **conocidas** es bastante simple llevar a cabo la inferencia relativa a $\boldsymbol{\beta}$ y \mathbf{b}_i basado en la **verosimilitud marginal**.



Considere reunir la información para los n individuos

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \vdots \\ \mathbf{Y}_n \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_n \end{pmatrix},$$

donde $\mathbf{Y} \in \mathbb{R}^N$, $\mathbf{b} \in \mathbb{R}^{nq}$ y $\mathbf{X} \in \mathbb{R}^{N \times p}$. Además, sea

$$\tilde{\Psi} = \mathbf{I} \otimes \Psi, \quad \mathbf{Z} = \bigoplus_{i=1}^n \mathbf{Z}_i, \quad \mathbf{R} = \bigoplus_{i=1}^n \mathbf{R}_i.$$

De este modo,

$$\mathbf{V} = \bigoplus_{i=1}^n \mathbf{V}_i = \mathbf{Z} \tilde{\Psi} \mathbf{Z}^T + \mathbf{R},$$

con $\tilde{\Psi} \in \mathbb{R}^{nq \times nq}$, $\mathbf{Z} \in \mathbb{R}^{N \times nq}$, $\mathbf{R} \in \mathbb{R}^{N \times N}$ y $\mathbf{V} \in \mathbb{R}^{N \times N}$.



De este modo, escribimos el **modelo combinado** para los n individuos como

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\epsilon}.$$

Las suposiciones distribucionales anteriores pueden escribirse como

$$\mathbf{Y}|\mathbf{b} \sim N_N(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}, \mathbf{R}), \quad \mathbf{b} \sim N_{nq}(\mathbf{0}, \tilde{\Psi}),$$

con distribución marginal para la respuesta \mathbf{Y} ,

$$\mathbf{Y} \sim N_N(\mathbf{X}\boldsymbol{\beta}, \mathbf{V}).$$



Considere el problema de mínimos cuadrados mediante un enfoque de **pseudo-datos**, con vector de **respuestas aumentadas**

$$\begin{pmatrix} R^{-1/2}Y \\ \mathbf{0} \end{pmatrix},$$

y **función de regresión**

$$\begin{pmatrix} R^{-1/2}X & R^{-1/2}Z \\ \mathbf{0} & \tilde{\Psi}^{-1/2} \end{pmatrix} \begin{pmatrix} \beta \\ b \end{pmatrix} = \begin{pmatrix} R^{-1/2}(X\beta + Zb) \\ \tilde{\Psi}^{-1/2}b \end{pmatrix}.$$

La solución de mínimos cuadrados al problema de pseudo-datos está dada por las **ecuaciones de Henderson**

$$\begin{pmatrix} X^T R^{-1}X & X^T R^{-1}Z \\ Z^T R^{-1}X & Z^T R^{-1}Z + \tilde{\Psi}^{-1} \end{pmatrix} \begin{pmatrix} \hat{\beta} \\ \hat{b} \end{pmatrix} = \begin{pmatrix} X^T R^{-1}Y \\ Z^T R^{-1}Y \end{pmatrix}$$



Las soluciones de las ecuaciones de Henderson están dadas por:

$$\begin{aligned}\hat{\beta} &= (\mathbf{X}^T \mathbf{R}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{R}^{-1} (\mathbf{Y} - \mathbf{Z} \hat{\mathbf{b}}), \\ \hat{\mathbf{b}} &= (\mathbf{Z}^T \mathbf{R}^{-1} \mathbf{Z} + \tilde{\Psi}^{-1})^{-1} \mathbf{Z}^T \mathbf{R}^{-1} (\mathbf{Y} - \mathbf{X} \hat{\beta}).\end{aligned}$$

Si no se considera $\tilde{\Psi}$ en las ecuaciones de Henderson obtenemos el MLE para el modelo con \mathbf{b} considerado como [efectos fijos](#).

Usando la [fórmula de reducción](#)

$$\mathbf{R}^{-1} - \mathbf{R}^{-1} \mathbf{Z} (\mathbf{Z}^T \mathbf{R}^{-1} \mathbf{Z} + \tilde{\Psi}^{-1})^{-1} \mathbf{Z}^T \mathbf{R}^{-1} = (\mathbf{R} + \mathbf{Z} \tilde{\Psi} \mathbf{Z}^T)^{-1},$$

es posible mostrar (tarea) que $\hat{\beta}$ puede ser escrito como el estimador GLS

$$\begin{aligned}\hat{\beta} &= (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{Y} \\ &= \left(\sum_{i=1}^n \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i \right)^{-1} \sum_{i=1}^n \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{Y}_i.\end{aligned}$$



Análogamente se puede mostrar que (tarea)

$$\hat{\mathbf{b}} = \tilde{\Psi} \mathbf{Z}^T \mathbf{V}^{-1} (\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}}),$$

de este modo estimaciones individuales $\hat{\mathbf{b}}_i$ están dadas por

$$\hat{\mathbf{b}}_i = \Psi \mathbf{Z}_i^T \mathbf{V}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}), \quad i = 1, \dots, n.$$

Que es BLUP para \mathbf{b} , esto es, $\hat{\mathbf{b}}$ puede ser derivado como la media posterior $E(\mathbf{b} | \mathbf{Y}, \hat{\boldsymbol{\beta}}; \mathbf{R}, \tilde{\Psi})$, y por tanto tiene una interpretación Bayes empírica.



Referencias sobre estimación en LME



Bates, D. M., and Pinheiro, J. C. (1998).
Computational methods for multilevel modelling.
Technical Memorandum BL0112140-980226-01TM,
Bell Labs, Lucent Technologies, Murray Hill, NJ.



Pinheiro, J. C., and Bates, D. M. (1996).
Unconstrained parametrizations for variance-covariance matrices.
Statistics and Computing 6, 289-296.



Pinheiro, J. C., and Bates, D. M. (2000).
Mixed-Effects Models in S and S-Plus.
Springer, New York.



Descripción del modelo

Considere el modelo con efectos mixtos

$$\mathbf{Y}_i | \mathbf{b}_i \stackrel{\text{ind}}{\sim} N_{m_i}(\mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i, \sigma^2 \mathbf{I}_{m_i}), \quad \mathbf{b}_i \stackrel{\text{ind}}{\sim} N_q(\mathbf{0}, \boldsymbol{\Psi}),$$

para $i = 1, \dots, n$, donde $\boldsymbol{\Psi} = \boldsymbol{\Psi}(\boldsymbol{\theta})$ es una **matriz de covarianza** (i.e, simétrica y definida positiva).

La clave para expresar los cálculos requeridos en la estimación ML es expresar la matriz $\boldsymbol{\Psi}$, **relativa** a σ^2 .

Se define el **factor de precisión relativa**

$$\sigma^2 \boldsymbol{\Psi}^{-1} = \boldsymbol{\Delta}^T \boldsymbol{\Delta}, \quad \boldsymbol{\Delta} = \boldsymbol{\Delta}(\boldsymbol{\theta}).$$

Pinheiro y Bates (1996), describen algunas **parametrizaciones** para determinar el factor $\boldsymbol{\Delta}$ (Cholesky, log-Cholesky, Esférica, log-matricial, Givens).



Parametrización log-matricial

Por ejemplo, considere la parametrización **log-matricial**. Sea $\Psi = \sigma^2 e^{\mathbf{A}}$, donde \mathbf{A} es simétrica y $\boldsymbol{\theta}$ representa los elementos no-redundantes de \mathbf{A} . Entonces,

$$\Delta(\boldsymbol{\theta}) = e^{-\mathbf{A}(\boldsymbol{\theta})/2} \quad \Rightarrow \quad \sigma^2 \Psi^{-1} = \Delta^T \Delta.$$

Note que si, $\mathbf{A} = \mathbf{\Gamma} \mathbf{\Lambda} \mathbf{\Gamma}^T$ es la **descomposición espectral** de \mathbf{A} , con

$$\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_q),$$

y $\mathbf{\Gamma}$ ortogonal. Entonces una **función matricial**

$$f(\mathbf{A}) = \mathbf{\Gamma} f(\mathbf{\Lambda}) \mathbf{\Gamma}^T, \quad \mathbf{\Lambda} = \text{diag}(f(\lambda_1), \dots, f(\lambda_q)),$$

(ver Golub y van Loan, 1996).



Evaluando la log-verosimilitud

Sabemos que la verosimilitud puede ser escrita como:

$$\begin{aligned} L(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2) &= \prod_{i=1}^n \frac{|\boldsymbol{\Delta}|}{(2\pi\sigma^2)^{m_i/2}} \int_{\mathbb{R}^q} \frac{\exp\{-(\|\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i\|^2 + \|\boldsymbol{\Delta}\mathbf{b}_i\|^2)/2\sigma^2\}}{(2\pi\sigma^2)^{q/2}} d\mathbf{b}_i \\ &= \prod_{i=1}^n \frac{|\boldsymbol{\Delta}|}{(2\pi\sigma^2)^{m_i/2}} \int_{\mathbb{R}^q} \frac{\exp\{-\|\tilde{\mathbf{Y}}_i - \tilde{\mathbf{X}}_i\boldsymbol{\beta} - \tilde{\mathbf{Z}}_i\mathbf{b}_i\|^2/2\sigma^2\}}{(2\pi\sigma^2)^{q/2}} d\mathbf{b}_i \end{aligned} \quad (1)$$

con **pseudo-datos**:

$$\tilde{\mathbf{Y}}_i = \begin{pmatrix} \mathbf{Y}_i \\ \mathbf{0} \end{pmatrix}, \quad \tilde{\mathbf{X}}_i = \begin{pmatrix} \mathbf{X}_i \\ \mathbf{0} \end{pmatrix}, \quad \tilde{\mathbf{Z}}_i = \begin{pmatrix} \mathbf{Z}_i \\ \boldsymbol{\Delta} \end{pmatrix}.$$



Evaluando la log-verosimilitud

Ahora considere la descomposición QR de \tilde{Z}_i ,

$$\tilde{Z}_i = Q_{(i)} \begin{pmatrix} R_{11(i)} \\ \mathbf{0} \end{pmatrix},$$

donde $Q_{(i)} \in \mathbb{R}^{(n_i+q) \times (n_i+q)}$ y $R_{11(i)} \in \mathbb{R}^{q \times q}$ es triangular superior. Además, sea

$$Q_{(i)}^T \tilde{X}_i = \begin{pmatrix} R_{10(i)} \\ R_{00(i)} \end{pmatrix}, \quad Q_{(i)}^T \tilde{Y}_i = \begin{pmatrix} c_{1(i)} \\ c_{0(i)} \end{pmatrix},$$

o bien, expresado en forma más compacta

$$\begin{pmatrix} Z_i & X_i & Y_i \\ \Delta & \mathbf{0} & \mathbf{0} \end{pmatrix} = Q_{(i)} \begin{pmatrix} R_{11(i)} & R_{10(i)} & c_{1(i)} \\ \mathbf{0} & R_{00(i)} & c_{0(i)} \end{pmatrix}. \quad (2)$$



Luego,

$$\begin{aligned}\|\tilde{\mathbf{Y}}_i - \tilde{\mathbf{X}}_i\boldsymbol{\beta} - \tilde{\mathbf{Z}}_i\mathbf{b}_i\|^2 &= \\ &= \|\mathbf{Q}_{(i)}^T(\tilde{\mathbf{Y}}_i - \tilde{\mathbf{X}}_i\boldsymbol{\beta} - \tilde{\mathbf{Z}}_i\mathbf{b}_i)\|^2 \\ &= \|\mathbf{c}_{1(i)} - \mathbf{R}_{10(i)}\boldsymbol{\beta} - \mathbf{R}_{11(i)}\mathbf{b}_i\|^2 + \|\mathbf{c}_{0(i)} - \mathbf{R}_{00(i)}\boldsymbol{\beta}\|^2.\end{aligned}$$

El cálculo de la descomposición en (2) es sencillo, **eficiente** y **numéricamente estable**.

Rutinas están disponibles, por ejemplo, en bibliotecas **Linpack** (Dongarra et al., 1979)

o **LAPACK** (Anderson et al., 1994).



Volviendo a la integral en (1), tenemos

$$\begin{aligned} & \int_{\mathbb{R}^q} \frac{\exp\{-(\|\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i\|^2 + \|\boldsymbol{\Delta}\mathbf{b}_i\|^2)/2\sigma^2\}}{(2\pi\sigma^2)^{q/2}} d\mathbf{b}_i \\ &= \exp\left\{-\frac{1}{2\sigma^2}\|\mathbf{c}_{0(i)} - \mathbf{R}_{00(i)}\boldsymbol{\beta}\|^2\right\} \int_{\mathbb{R}^q} \frac{\exp\{-\|\mathbf{c}_{1(i)} - \mathbf{R}_{10(i)}\boldsymbol{\beta} - \mathbf{R}_{11(i)}\mathbf{b}_i\|^2/2\sigma^2\}}{(2\pi\sigma^2)^{q/2}} d\mathbf{b}_i. \end{aligned}$$

Considere el cambio de variables

$$\mathbf{g}_i = (\mathbf{c}_{1(i)} - \mathbf{R}_{10(i)}\boldsymbol{\beta} - \mathbf{R}_{11(i)}\mathbf{b}_i)/\sigma,$$

con diferencial $d\mathbf{g}_i = -\sigma^{-1}\mathbf{R}_{11(i)} d\mathbf{b}_i$ y Jacobiano de la transformación

$$|J(\mathbf{b}_i \rightarrow \mathbf{g}_i)| = \sigma^{-q/2} \text{abs} |\mathbf{R}_{11(i)}|.$$



Evaluando la log-verosimilitud

De este modo,

$$\begin{aligned} & \int_{\mathbb{R}^q} \frac{\exp\{-\|\mathbf{c}_{1(i)} - \mathbf{R}_{10(i)}\boldsymbol{\beta} - \mathbf{R}_{11(i)}\mathbf{b}_i\|^2/2\sigma^2\}}{(2\pi\sigma^2)^{q/2}} d\mathbf{b}_i \\ &= \frac{1}{\text{abs } |\mathbf{R}_{11(i)}|} \int_{\mathbb{R}^q} \frac{\exp(-\|\mathbf{g}_i\|^2/2)}{(2\pi)^{q/2}} d\mathbf{g}_i \\ &= \frac{1}{\text{abs } |\mathbf{R}_{11(i)}|}. \end{aligned}$$

Además,

$$\begin{aligned} |\tilde{\mathbf{Z}}_i^T \tilde{\mathbf{Z}}_i|^{1/2} &= |\mathbf{R}_{11(i)}^T \mathbf{R}_{11(i)}|^{1/2} = \sqrt{|\mathbf{R}_{11(i)}|^2} \\ &= \text{abs } |\mathbf{R}_{11(i)}| \end{aligned}$$



Evaluando la log-verosimilitud

La **función de log-verosimilitud** asume la forma,

$$\begin{aligned}\ell(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2) &= \log L(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2) \\ &= -\frac{N}{2} \log 2\pi\sigma^2 + \sum_{i=1}^n \log \text{abs} \left(\frac{|\Delta|}{|\mathbf{R}_{11(i)}|} \right) - \frac{1}{2\sigma^2} \sum_{i=1}^n \|\mathbf{c}_{0(i)} - \mathbf{R}_{00(i)}\boldsymbol{\beta}\|^2.\end{aligned}$$

Formando otra descomposición QR,

$$\begin{pmatrix} \mathbf{R}_{00(1)} & \mathbf{c}_{0(1)} \\ \vdots & \vdots \\ \mathbf{R}_{00(n)} & \mathbf{c}_{0(n)} \end{pmatrix} = \mathbf{Q}_0 \begin{pmatrix} \mathbf{R}_{00} & \mathbf{c}_0 \\ \mathbf{0} & \mathbf{c}_{-1} \end{pmatrix}.$$



Evaluando la log-verosimilitud

Podemos escribir la función de log-verosimilitud como

$$\ell(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2) = -\frac{N}{2} \log 2\pi\sigma^2 + \sum_{i=1}^n \log \left(\frac{|\boldsymbol{\Delta}|}{|\mathbf{R}_{11(i)}|} \right) - \frac{1}{2\sigma^2} (\|\mathbf{c}_{-1}\|^2 + \|\mathbf{c}_0 - \mathbf{R}_{00}\boldsymbol{\beta}\|^2).$$

Para $\boldsymbol{\theta}$ **fijado**, los valores de $\boldsymbol{\beta}$ y σ^2 que maximizan $\ell(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2)$ están dados por

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \mathbf{R}_{00}^{-1} \mathbf{c}_0 \quad \text{y} \quad \hat{\sigma}^2(\boldsymbol{\theta}) = \frac{1}{N} \|\mathbf{c}_{-1}\|^2,$$

con $N = \sum_{i=1}^n m_i$.



De este modo, para θ fijado, tenemos que la **log-verosimilitud perfilada**, asume la forma

$$\begin{aligned}\ell_*(\theta) &= \ell(\hat{\beta}(\theta), \theta, \hat{\sigma}^2(\theta)) \\ &= -\frac{N}{2} \log(2\pi\hat{\sigma}^2(\theta)) + \sum_{i=1}^n \log\left(\frac{|\Delta|}{|\mathbf{R}_{11(i)}|}\right) - \frac{N}{2} \\ &= k + N \log \|\mathbf{c}_{-1}\|^2 + \sum_{i=1}^n \log\left(\frac{|\Delta|}{|\mathbf{R}_{11(i)}|}\right),\end{aligned}$$

donde $k = N(\log N - \log(2\pi) - 1)/2$.

La log-verosimilitud perfilada es maximizada con relación a θ , obteniendo $\hat{\theta}$. Finalmente, los MLE $\hat{\beta}$ y $\hat{\sigma}^2$, son obtenidos tomando $\theta = \hat{\theta}$.



Desde las descomposiciones matriciales anteriores, tenemos que el **Mejor Predictor Lineal e Inssegado (BLUP)** para b_i , está dado por:

$$\hat{b}_i(\theta) = R_{11(i)}^{-1}(c_{1(i)} - R_{10(i)}\hat{\beta}(\theta)),$$

para $i = 1, \dots, n$.

En la práctica usamos el MLE $\hat{\theta}$, obteniendo BLUP **estimados**, $\hat{b}_i(\hat{\theta})$.



La verosimilitud restringida puede ser expresada como

$$L_{\text{REML}}(\boldsymbol{\tau}, \boldsymbol{\theta}) = \int L(\boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\theta}) d\boldsymbol{\beta}.$$

Usando la misma técnica anterior obtenemos la **log-verosimilitud restringida**

$$\begin{aligned} \ell_{\text{REML}}(\boldsymbol{\theta}, \sigma^2) = & -\frac{N-p}{2} \log 2\pi\sigma^2 - \frac{1}{2\sigma^2} \|\mathbf{c}_{-1}\|^2 \\ & - \log \text{abs } |\mathbf{R}_{00}| + \sum_{i=1}^n \log \left(\frac{|\boldsymbol{\Delta}|}{|\mathbf{R}_{11(i)}|} \right), \end{aligned}$$



De este modo, obtenemos el estimador condicional,

$$\hat{\sigma}_R^2(\boldsymbol{\theta}) = \frac{1}{N-p} \|\mathbf{c}_{-1}\|^2,$$

y perfilando, obtenemos

$$\begin{aligned}\ell_{\text{REML}}^*(\boldsymbol{\theta}) &= \ell_{\text{REML}}(\boldsymbol{\theta}, \hat{\sigma}_R^2(\boldsymbol{\theta})) \\ &= k^* - (N-p) \log \|\mathbf{c}_{-1}\|^2 - \log |\mathbf{R}_{00}| + \sum_{i=1}^n \log \left(\frac{|\Delta|}{|\mathbf{R}_{11(i)}|} \right),\end{aligned}$$

donde k^* es una constante.

