MAT-468: Modelos lineales con efectos mixtos

Felipe Osorio

fosorios.mat.utfsm.cl

Departamento de Matemática, UTFSM

Modelos lineales con efectos mixtos

- 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.

Etapa 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 $\boldsymbol{Y}_i$ para el i-ésimo individuo

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

- Modelo para la etapa 1:

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

  donde $\boldsymbol{Z}_i$ 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:

$$\boldsymbol{\epsilon}_i \sim \mathsf{N}_{m_i}(\boldsymbol{0}, \boldsymbol{R}_i),$$

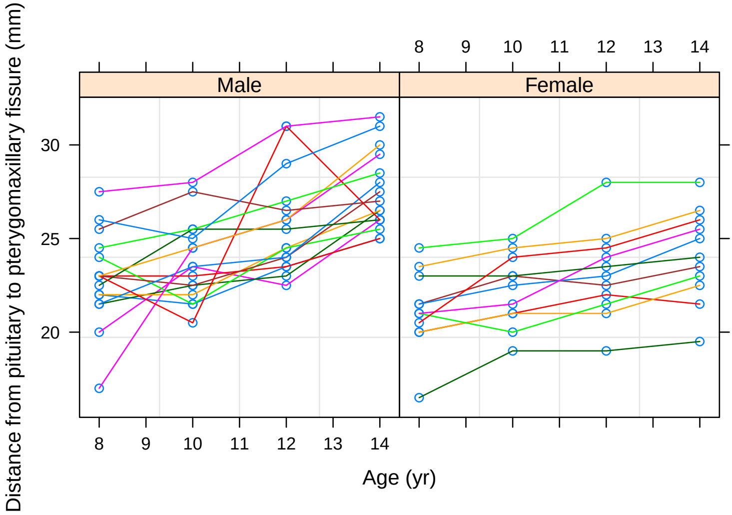
  con $\boldsymbol{R}_i = \boldsymbol{R}_i(\boldsymbol{\tau})$. Frecuentemente $\boldsymbol{R}_i = \sigma^2 \boldsymbol{I}_{m_i}$.

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

Etapa 2:

- Variabilidad **entre-individuos** puede ser estudiada relacionando $\boldsymbol{\beta}_i$ a covariables.
- Modelo para la etapa 2:
$$\boldsymbol{\beta}_i = \boldsymbol{K}_i\boldsymbol{\beta} + \boldsymbol{b}_i,$$
donde $\boldsymbol{K}_i$ es matriz de diseño $q \times p$.
- $\boldsymbol{\beta}$ es vector (común) de parámetros de regresión desconocidos.
- Supuesto distribucional:
$$\boldsymbol{b}_i \sim \mathsf{N}_q(\mathbf{0}, \boldsymbol{\Psi}).$$

Estudio de Ortodoncia (Potthoff y Roy, 1964)



Estudio de Ortodoncia: Modelo para la etapa 1

- Observaciones tomadas en tiempos $t = 8, 10, 12, 14$.
- Modelo para la Etapa 1:

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

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

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

- Notación matricial

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

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

$$\beta_{0i} = \beta_0 + b_{1i}, \qquad \beta_{1i} = \beta_1 + b_{1i}$$
$$\beta_{2i} = \beta_2 + b_{2i}, \qquad \beta_{3i} = \beta_3 + b_{2i},$$

donde

$$\boldsymbol{b}_i = \begin{pmatrix} b_{1i} \\ b_{2i} \end{pmatrix} \sim \mathsf{N}_2(\mathbf{0}, \boldsymbol{\Psi}), \quad \boldsymbol{\Psi} = \begin{pmatrix} \psi_{11} & \psi_{12} \\ \psi_{21} & \psi_{22} \end{pmatrix}$$

- Notación matricial:

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{b}_i.$$

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

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

$$\begin{aligned}\boldsymbol{Y}_i &= \boldsymbol{Z}_i\boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i,\\ \boldsymbol{\beta}_i &= \boldsymbol{K}_i\boldsymbol{\beta} + \boldsymbol{b}_i,\end{aligned}$$

esto es,

$$\begin{aligned}\boldsymbol{Y}_i &= \boldsymbol{Z}_i\boldsymbol{K}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i\\ &= \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i,\end{aligned}$$

con $\boldsymbol{X}_i = \boldsymbol{Z}_i\boldsymbol{K}_i$ matriz de diseño $m_i \times p$.

Modelos lineales con efectos mixtos (Laird y Ware, 1982)[1]

Definición 1: (Modelo lineal con efectos mixtos)

Considere

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

donde

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

con $\boldsymbol{X}_i$, $\boldsymbol{Z}_i$ matrices de diseño $m_i \times p$ y $m_i \times q$, respectivamente.

Terminología:

- El parámetro **común**, $\boldsymbol{\beta}$ es llamado vector de **efectos fijos**,
- "parámetros" **sujeto-específicos** $\boldsymbol{b}_i$ se denominan **efectos aleatorios**,
- elementos en $\boldsymbol{\Psi}$ y $\boldsymbol{R}_i$ son conocidos como **componentes de varianza**.

[1] Biometrics 38, 963-974.

Modelos lineales con efectos mixtos

El modelo lineal con efectos mixtos (lme)

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

con

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

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

$$\boldsymbol{Y}_i | \boldsymbol{b}_i \overset{\text{ind}}{\sim} \mathsf{N}_{m_i}(\boldsymbol{X}_i \boldsymbol{\beta} + \boldsymbol{Z}_i \boldsymbol{b}_i, \boldsymbol{R}_i), \qquad \boldsymbol{b}_i \overset{\text{ind}}{\sim} \mathsf{N}_q(\boldsymbol{0}, \boldsymbol{\Psi}),$$

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

El modelo lme también es conocido como:

- jerárquico,
- multinivel,
- en dos etapas.

Modelos lineales con efectos mixtos

El modelo lme puede ser re-expresado como:

$$\begin{aligned}\boldsymbol{Y}_i &= \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i \\ &= \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{u}_i, \qquad \boldsymbol{u}_i = \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i,\end{aligned}$$

como $\boldsymbol{b}_i \perp \boldsymbol{\epsilon}_i$, tenemos

$$\boldsymbol{u}_i \sim \mathsf{N}_{m_i}(\boldsymbol{0}, \boldsymbol{V}_i), \qquad \boldsymbol{V}_i = \boldsymbol{Z}_i\boldsymbol{\Psi}\boldsymbol{Z}_i^T + \boldsymbol{R}_i.$$

Finalmente, obtenemos el **modelo marginal**:

$$\boldsymbol{Y}_i \sim \mathsf{N}_{m_i}(\boldsymbol{X}_i\boldsymbol{\beta}, \boldsymbol{Z}_i\boldsymbol{\Psi}\boldsymbol{Z}_i^T + \boldsymbol{R}_i).$$

- De ahí que, supuestos muy específicos se realizan sobre la **media y covarianza marginal**.
  - Media implicada: $\boldsymbol{X}_i\boldsymbol{\beta}$,
  - Covarianza implicada: $\boldsymbol{V}_i = \boldsymbol{Z}_i\boldsymbol{\Psi}\boldsymbol{Z}_i^T + \boldsymbol{R}_i$.
- Note que el modelo jerárquico implica el marginal y **NO** viceversa.

Considere el modelo jerárquico:

$$\boldsymbol{Y}_i|\boldsymbol{b}_i \overset{\text{ind}}{\sim} \mathsf{N}_{m_i}(\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i, \sigma^2\boldsymbol{I}_{m_i}), \quad \boldsymbol{b}_i \overset{\text{ind}}{\sim} \mathsf{N}_q(\boldsymbol{0}, \boldsymbol{\Psi}), \quad \boldsymbol{\Psi} = \boldsymbol{\Psi}(\boldsymbol{\theta}),$$

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

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

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

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

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

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

Tenemos

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

y considere el factor de precisión relativa

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

esto permite escribir:

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

De este modo, obtenemos la verosimilitud

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

donde

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

son conocidos como pseudo-datos.

Note que

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

donde

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

Luego

$$\int_{\mathbb{R}^q} \frac{\exp\{-\|\tilde{\boldsymbol{Y}}_i - \tilde{\boldsymbol{X}}_i\boldsymbol{\beta} - \tilde{\boldsymbol{Z}}_i\boldsymbol{b}_i\|^2/2\sigma^2\}}{(2\pi\sigma^2)^{q/2}}\,\mathrm{d}\boldsymbol{b}_i$$

$$= \exp\{-\frac{1}{2\sigma^2}\|\tilde{\boldsymbol{Y}}_i - \tilde{\boldsymbol{X}}_i\boldsymbol{\beta} - \tilde{\boldsymbol{Z}}_i\widehat{\boldsymbol{b}}_i\|^2\}\int_{\mathbb{R}^q} \frac{\exp\{-(\boldsymbol{b}_i - \widehat{\boldsymbol{b}}_i)^T\tilde{\boldsymbol{Z}}_i^T\tilde{\boldsymbol{Z}}_i(\boldsymbol{b}_i - \widehat{\boldsymbol{b}}_i)/2\sigma^2\}}{(2\pi\sigma^2)^{q/2}}\,\mathrm{d}\boldsymbol{b}_i$$

Por otro lado

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

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

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

Obteniendo finalmente

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

Por otro lado, para el modelo marginal:

$$\boldsymbol{Y}_i \sim \mathsf{N}_{m_i}(\boldsymbol{X}_i\boldsymbol{\beta}, \boldsymbol{Z}_i\boldsymbol{\Psi}\boldsymbol{Z}_i^T + \boldsymbol{R}_i), \qquad i = 1, \dots, n,$$

tenemos que la función de verosimilitud está dada por:

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

Observación:

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

Considere reunir la información para los n individuos

$$\boldsymbol{Y} = \begin{pmatrix} \boldsymbol{Y}_1 \\ \vdots \\ \boldsymbol{Y}_n \end{pmatrix}, \qquad \boldsymbol{b} = \begin{pmatrix} \boldsymbol{b}_1 \\ \vdots \\ \boldsymbol{b}_n \end{pmatrix}, \qquad \boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1 \\ \vdots \\ \boldsymbol{X}_n \end{pmatrix},$$

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

$$\tilde{\boldsymbol{\Psi}} = \boldsymbol{I} \otimes \boldsymbol{\Psi}, \qquad \boldsymbol{Z} = \oplus_{i=1}^{n} \boldsymbol{Z}_i, \qquad \boldsymbol{R} = \oplus_{i=1}^{n} \boldsymbol{R}_i.$$

De este modo,

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

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

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

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

Las suposiciones distribucionales anteriores pueden escribirse como

$$\boldsymbol{Y}|\boldsymbol{b} \sim \mathsf{N}_N(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{b}, \boldsymbol{R}), \qquad \boldsymbol{b} \sim \mathsf{N}_{nq}(\boldsymbol{0}, \tilde{\boldsymbol{\Psi}}),$$

con distribución marginal para la respuesta $\boldsymbol{Y}$,

$$\boldsymbol{Y} \sim \mathsf{N}_N(\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{V}).$$

Pseudo-datos

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

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

y función de regresión

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

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

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

Las soluciones de las ecuaciones de Henderson están dadas por:

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{R}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{R}^{-1}(\boldsymbol{Y} - \boldsymbol{Z}\widehat{\boldsymbol{b}}),$$
$$\widehat{\boldsymbol{b}} = (\boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{Z} + \tilde{\boldsymbol{\Psi}}^{-1})^{-1}\boldsymbol{Z}^T\boldsymbol{R}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}).$$

Si no se considera $\tilde{\boldsymbol{\Psi}}$ en las ecuaciones de Henderson obtenemos el MLE para el modelo con $\boldsymbol{b}$ considerado como efectos fijos.

Usando la fórmula de reducción

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

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

$$\begin{aligned}\widehat{\boldsymbol{\beta}} &= (\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{Y} \\ &= \Big(\sum_{i=1}^n \boldsymbol{X}_i^T\boldsymbol{V}_i^{-1}\boldsymbol{X}_i\Big)^{-1}\sum_{i=1}^n \boldsymbol{X}_i^T\boldsymbol{V}_i^{-1}\boldsymbol{Y}_i.\end{aligned}$$

Análogamente se puede mostrar que (tarea)

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

de este modo estimaciones individuales $\widehat{\boldsymbol{b}}_i$ están dadas por

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

Que es BLUP para $\boldsymbol{b}$, esto es, $\widehat{\boldsymbol{b}}$ puede ser derivado como la media posterior $\mathsf{E}(\boldsymbol{b}|\boldsymbol{Y}, \widehat{\boldsymbol{\beta}}; \boldsymbol{R}, \tilde{\boldsymbol{\Psi}})$, 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

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

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

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

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 $\boldsymbol{\Psi} = \sigma^2 e^{\boldsymbol{A}}$, donde $\boldsymbol{A}$ es simétrica y $\boldsymbol{\theta}$ representa los elementos no-redundantes de $\boldsymbol{A}$. Entonces,

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

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

$$\boldsymbol{\Lambda} = \operatorname{diag}(\lambda_1, \dots, \lambda_q),$$

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

$$f(\boldsymbol{A}) = \boldsymbol{\Gamma} f(\boldsymbol{\Lambda}) \boldsymbol{\Gamma}^T, \qquad \boldsymbol{\Lambda} = \operatorname{diag}(f(\lambda_1), \dots, f(\lambda_q)),$$

(ver Golub y van Loan, 1996).

Evaluando la log-verosimilitud

Sabemos que la verosimilitud puede ser escrita como:

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

con pseudo-datos:

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

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

$$\tilde{\boldsymbol{Z}}_i = \boldsymbol{Q}_{(i)} \begin{pmatrix} \boldsymbol{R}_{11(i)} \\ \boldsymbol{0} \end{pmatrix},$$

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

$$\boldsymbol{Q}_{(i)}^T \tilde{\boldsymbol{X}}_i = \begin{pmatrix} \boldsymbol{R}_{10(i)} \\ \boldsymbol{R}_{00(i)} \end{pmatrix}, \qquad \boldsymbol{Q}_{(i)}^T \tilde{\boldsymbol{Y}}_i = \begin{pmatrix} \boldsymbol{c}_{1(i)} \\ \boldsymbol{c}_{0(i)} \end{pmatrix},$$

o bien, expresado en forma más compacta

$$\begin{pmatrix} \boldsymbol{Z}_i & \boldsymbol{X}_i & \boldsymbol{Y}_i \\ \boldsymbol{\Delta} & \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} = \boldsymbol{Q}_{(i)} \begin{pmatrix} \boldsymbol{R}_{11(i)} & \boldsymbol{R}_{10(i)} & \boldsymbol{c}_{1(i)} \\ \boldsymbol{0} & \boldsymbol{R}_{00(i)} & \boldsymbol{c}_{0(i)} \end{pmatrix}. \tag{2}$$

Evaluando la log-verosimilitud

Luego,

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

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

$$\int_{\mathbb{R}^q} \frac{\exp\{-(\|\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta} - \boldsymbol{Z}_i\boldsymbol{b}_i\|^2 + \|\boldsymbol{\Delta}\boldsymbol{b}_i\|^2)/2\sigma^2\}}{(2\pi\sigma^2)^{q/2}}\,\mathrm{d}\boldsymbol{b}_i$$
$$= \exp\Big\{-\frac{1}{2\sigma^2}\|\boldsymbol{c}_{0(i)} - \boldsymbol{R}_{00(i)}\boldsymbol{\beta}\|^2\Big\}\int_{\mathbb{R}^q} \frac{\exp\{-\|\boldsymbol{c}_{1(i)} - \boldsymbol{R}_{10(i)}\boldsymbol{\beta} - \boldsymbol{R}_{11(i)}\boldsymbol{b}_i\|^2/2\sigma^2\}}{(2\pi\sigma^2)^{q/2}}\,\mathrm{d}\boldsymbol{b}_i.$$

Considere el cambio de variables

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

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

$$|J(\boldsymbol{b}_i \to \boldsymbol{g}_i)| = \sigma^{-q/2}\,\mathrm{abs}\,|\boldsymbol{R}_{11(i)}|.$$

Evaluando la log-verosimilitud

De este modo,

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

Además,

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

Evaluando la log-verosimilitud

La función de log-verosimilitud asume la forma,

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

Formando otra descomposición QR,

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

Evaluando la log-verosimilitud

Podemos escribir la función de log-verosimilitud como

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

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

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

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

log-verosimilitud perfilada

De este modo, para $\boldsymbol{\theta}$ fijado, tenemos que la **log-verosimilitud perfilada**, asume la forma

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

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

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

BLUP

Desde las descomposiciones matriciales anteriores, tenemos que el Mejor Predictor Lineal e Insesgado (BLUP) para $\boldsymbol{b}_i$, está dado por:

$$\widehat{\boldsymbol{b}}_i(\boldsymbol{\theta}) = \boldsymbol{R}_{11(i)}^{-1}(\boldsymbol{c}_{1(i)} - \boldsymbol{R}_{10(i)}\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta})),$$

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

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

log-verosimilitud REML

La verosimilitud restricta puede ser expresada como

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

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

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

log-verosimilitud REML perfilada

De este modo, obtenemos el estimador condicional,

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

y perfilando, obtenemos

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

donde k^* es una constante.

Considere el modelo con efectos mixtos

$$\boldsymbol{Y}_i|\boldsymbol{b}_i \overset{\text{ind}}{\sim} \mathsf{N}_{m_i}(\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i, \sigma^2\boldsymbol{I}_{m_i}), \quad \boldsymbol{b}_i \overset{\text{ind}}{\sim} \mathsf{N}_q(\mathbf{0}, \boldsymbol{\Psi}), \tag{3}$$

para $i = 1, \dots, n$, donde $\boldsymbol{\Psi} = \boldsymbol{\Psi}(\boldsymbol{\theta})$.

Problema de datos incompletos:

Sea $\boldsymbol{Y}_{\text{com}} = (\boldsymbol{Y}^\top, \boldsymbol{b}^\top)^\top$, donde $\boldsymbol{Y} = (\boldsymbol{Y}_1^\top, \dots, \boldsymbol{Y}_n^\top)^\top$ es el vector de datos observados, mientras que $\boldsymbol{b} = (\boldsymbol{b}_1^\top, \dots, \boldsymbol{b}_n^\top)^\top$ es considerado missing.

La parte relevante de la log-verosimilitud de datos completos, asume la forma

$$\begin{aligned}\ell_{\mathsf{c}}(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2; \boldsymbol{Y}_{\text{com}}) &= \frac{N}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{n} \|\boldsymbol{Y}_i - \boldsymbol{X}_i \boldsymbol{\beta} - \boldsymbol{Z}_i \boldsymbol{b}_i\|^2 \\ &\quad - \frac{n}{2} \log |\boldsymbol{\Psi}| - \frac{1}{2} \sum_{i=1}^{n} \boldsymbol{b}_i^T \boldsymbol{\Psi}^{-1} \boldsymbol{b}_i.\end{aligned}$$

usando el factor de precisión relativa $\sigma^2 \boldsymbol{\Psi}^{-1} = \boldsymbol{\Delta}^T \boldsymbol{\Delta}$,

$$\begin{aligned}\ell_{\mathsf{c}}(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2; \boldsymbol{Y}_{\text{com}}) &= -\frac{N + nq}{2} \log \sigma^2 + \frac{n}{2} \log |\boldsymbol{\Delta}^T \boldsymbol{\Delta}| \\ &\quad - \frac{1}{2\sigma^2} \sum_{i=1}^{n} \{\|\boldsymbol{Y}_i - \boldsymbol{X}_i \boldsymbol{\beta} - \boldsymbol{Z}_i \boldsymbol{b}_i\|^2 + \|\boldsymbol{\Delta} \boldsymbol{b}_i\|^2\}.\end{aligned}$$

Usando la formulación jerárquica en (3) tenemos que:

$$\boldsymbol{b}_i|\boldsymbol{Y}_i \overset{\text{ind}}{\sim} \mathsf{N}_q(\widehat{\boldsymbol{b}}_i(\boldsymbol{\eta}), \widehat{\boldsymbol{\Omega}}_i(\boldsymbol{\eta})), \qquad i = 1, \dots, n,$$

donde

$$\begin{aligned}\widehat{\boldsymbol{b}}_i(\boldsymbol{\eta}) &= \boldsymbol{\Psi}\boldsymbol{Z}_i^T\boldsymbol{V}_i^{-1}(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}) \\ &= (\boldsymbol{\Delta}^T\boldsymbol{\Delta} + \boldsymbol{Z}_i^T\boldsymbol{Z}_i)^{-1}\boldsymbol{Z}_i^T(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}),\end{aligned}$$

y

$$\widehat{\boldsymbol{\Omega}}_i(\boldsymbol{\eta}) = \boldsymbol{\Psi} - \boldsymbol{\Psi}\boldsymbol{Z}_i^T\boldsymbol{V}_i^{-1}\boldsymbol{Z}_i\boldsymbol{\Psi} = \sigma^2(\boldsymbol{\Delta}^T\boldsymbol{\Delta} + \boldsymbol{Z}_i^T\boldsymbol{Z}_i)^{-1},$$

con $\boldsymbol{V}_i = \boldsymbol{Z}_i\boldsymbol{\Psi}\boldsymbol{Z}_i^T + \sigma^2\boldsymbol{I}$.

La esperanza condicional de la log-verosimilitud de datos completos, asume la forma

$$Q(\boldsymbol{\eta}|\widehat{\boldsymbol{\eta}}^{(r)}) = -\frac{N+nq}{2}\log\sigma^2 + \frac{n}{2}\log|\boldsymbol{\Delta}^T\boldsymbol{\Delta}| \\ -\frac{1}{2\sigma^2}\sum_{i=1}^n\big\{\|\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta} - \boldsymbol{Z}_i\widehat{\boldsymbol{b}}_i^{(r)}\|^2 + \|\boldsymbol{\Delta}\widehat{\boldsymbol{b}}_i^{(r)}\|^2 + \operatorname{tr}\widehat{\boldsymbol{\Omega}}_i^{(r)}(\boldsymbol{\Delta}^T\boldsymbol{\Delta} + \boldsymbol{Z}_i^T\boldsymbol{Z}_i)\big\}$$

La etapa-M en el algoritmo EM, está dada por:

$$\widehat{\boldsymbol{\beta}}^{(r+1)} = \Big(\sum_{i=1}^n \boldsymbol{X}_i^T\boldsymbol{X}_i\Big)^{-1}\sum_{i=1}^n \boldsymbol{X}_i^T(\boldsymbol{Y}_i - \boldsymbol{Z}_i\widehat{\boldsymbol{b}}_i^{(r)}),$$

$$\widehat{\sigma}^{2(r+1)} = \frac{1}{N+nq}\Big\{nq\widehat{\sigma}^{2(r)} + \sum_{i=1}^n(\|\boldsymbol{Y}_i - \boldsymbol{X}_i\widehat{\boldsymbol{\beta}}^{(r)} - \boldsymbol{Z}_i\widehat{\boldsymbol{b}}_i^{(r)}\|^2 + \|\widehat{\boldsymbol{\Delta}}^{(r)}\widehat{\boldsymbol{b}}_i^{(r)}\|^2)\Big\}.$$

Podemos obtener $\widehat{\boldsymbol{\theta}}^{(r+1)}$ perfilando la función de $Q(\boldsymbol{\eta}|\widehat{\boldsymbol{\eta}}^{(r)})$ y de este modo debemos maximizar la función

$$Q_*(\boldsymbol{\theta}|\widehat{\boldsymbol{\eta}}^{(r)}) = n \log|\boldsymbol{\Delta}| - \frac{1}{2\widehat{\sigma}^{(r)}} \operatorname{tr} \boldsymbol{\Delta} \Big\{ \sum_{i=1}^{n} (\widehat{\boldsymbol{\Omega}}_i^{(r)} + \widehat{\boldsymbol{b}}_i^{(r)} \widehat{\boldsymbol{b}}_i^{(r)^T}) \Big\} \boldsymbol{\Delta}^T.$$

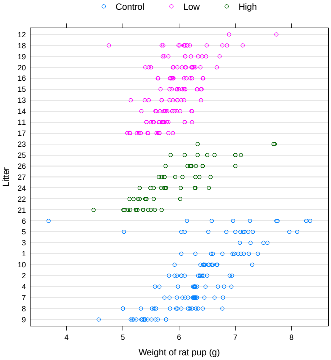
Note que, para $\boldsymbol{\Psi}$ no estructurada, tenemos

$$\widehat{\boldsymbol{\Psi}}^{(r+1)} = \frac{1}{n} \sum_{i=1}^{n} (\widehat{\boldsymbol{\Omega}}_i^{(r)} + \widehat{\boldsymbol{b}}_i^{(r)} \widehat{\boldsymbol{b}}_i^{(r)^T}).$$

Estudio reproductivo en roedores (Dempster et al., 1984)

Peso al nacer de las crías de 27 hembras de roedor sometidas a un compuesto toxicológico experimental que afecta el desempeño reproductivo. Se desea:

- Investigar si el nivel de la dosis afecta el desempeño reproductivo.
- Estudiar diferencias significativas entre los tratamientos.



Modelo:

$$Y_{ijkl} = \beta_j + b_i + \gamma L_{ij} + \delta S_k + \epsilon_{ijkl}, \quad i = 1, \ldots, 27,$$
$$j = 1, 2, 3, \quad k = 0, 1, \quad \text{y} \quad l = 1, \ldots, l_i,$$

donde Y_{ijkl} representa el peso al nascer de la l-ésima cria en la i-ésima nidada para el j-ésimo tratamiento con sexo k.

β_j es el efecto fijo del tratamento j; b_i denota el efecto aleatorio de nidada; L_{ij} es el tamaño de la nidada i asociada al j-ésimo tratamiento; S_k representa el sexo de la cría.

El modelo puede ser escrito como:

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i b_i + \boldsymbol{\epsilon}_i, \quad i = 1, \dots, 27,$$

donde $\boldsymbol{Y}_i$ es un vetor m_i-dimensional de pesos observados para las crías en la i-ésima nidada, $\boldsymbol{X}_i$ es matriz de diseño, $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \gamma, \delta)^\top$ y $\boldsymbol{Z}_i = \mathbf{1}_{m_i}$.

Tenemos la estructura jerárquica

$$\boldsymbol{Y}_i | b_i \overset{\text{ind}}{\sim} \mathsf{N}_{m_i}(\boldsymbol{X}_i\boldsymbol{\beta} + \mathbf{1}_{m_i} b_i, \sigma^2 \boldsymbol{I}_{m_i}), \quad b_i \overset{\text{ind}}{\sim} \mathsf{N}(0, \psi), \quad i = 1, \dots, 27.$$

Realizando el ajuste, obtenemos

$$\begin{aligned} \widehat{\boldsymbol{\beta}} &= (7.8525, 7.4297, 6.9690, -0.1189, 0.2971)^\top, \\ \widehat{\sigma}^2 &= 0.1676, \qquad \widehat{\psi} = 0.0591. \end{aligned}$$

y, por ejemplo,

$$\begin{aligned} \widehat{b}_5 &= 0.324, & \widehat{b}_6 &= -0.012, & \widehat{b}_8 &= -0.066 \\ \widehat{b}_9 &= -0.595, & \widehat{b}_{18} &= 0.369, & \widehat{b}_{23} &= 0.270 \end{aligned}$$