

Linear mixed models (LMMs)

Example (from prev. lecture)

i_j denotes group i ($i=1, \dots, m$),

subject j ($j=1, \dots, n_i$)

$$n = \sum_{i=1}^m n_i$$

Y_{ij} test score student j in program i

X_{ij} study hours — " —

Model :

$$Y_{ij} = \underbrace{(\beta_0 + \gamma_{0i})}_{\text{intercept}} + \underbrace{(\beta_1 + \gamma_{1i})}_{\text{slope}} X_{ij} + \epsilon_{ij}$$

$\beta_0 + \beta_1 X_{ij}$ fixed "population level"

$\gamma_{0i} + \gamma_{1i} X_{ij}$ random "group level"

ϵ_{ij} random "subject level"

For student ij : $\epsilon_{ij} \sim N(0, \sigma^2)$

For program i : $\vec{\gamma}_i = \begin{pmatrix} \gamma_{0i} \\ \gamma_{1i} \end{pmatrix} \sim \text{MVN}_2(\vec{0}, Q)$

$$Q = \begin{bmatrix} \tau_0^2 & \tau_{01} \\ \tau_{01} & \tau_1^2 \end{bmatrix}$$

Parameters

Coefficients: β_0, β_1

Variance components: $\sigma^2, \tau_0^2, \tau_1^2, \tau_{01}$

} to be estimated from data

Matrix formulation

Group i ($i = 1, \dots, m$)

$$\text{Model: } \vec{Y}_i = \underbrace{X_i \vec{\beta}}_{\text{fixed}} + \underbrace{U_i \vec{\gamma}_i}_{\text{random}} + \vec{\epsilon}_i$$

* X_i $n_i \times (k+1)$ design matrix fixed effects

example: $X_i = \begin{bmatrix} 1 & x_{i1} \\ \vdots & \vdots \\ 1 & x_{in_i} \end{bmatrix} \quad k=1$

$\nwarrow$ study hours

* $\vec{\beta}$ $(k+1)$ 'fixed effects'

* U_i $n_i \times (q+1)$ design matrix random effects

example: $U_i = X_i$

* $\vec{\gamma}_i$ $(q+1)$ 'random effects'

ex: $\vec{\gamma}_i = \begin{pmatrix} \gamma_{0i} \\ \gamma_{1i} \end{pmatrix}$

* $\vec{\epsilon}_i$ (n_i) 'random noise'

* $\vec{Y}_i$ (n_i) responses, group i

Assumptions:

$$\begin{aligned}\vec{y}_i &\sim \text{MVN}_{q+1}(\vec{0}, Q) \\ \vec{\varepsilon}_i &\sim \text{MVN}_{n_i}(\vec{0}, \sigma^2 I)\end{aligned} \quad \begin{array}{c} \updownarrow \\ \text{independent} \end{array}$$

Properties:

$$E(\vec{y}_i) = x_i \vec{\beta}$$

$$\text{Cov}(\vec{y}_i) = U_i Q U_i^T + \sigma^2 I$$

$$\vec{y}_i \sim \text{MVN}_{n_i}(x_i \vec{\beta}, \underbrace{U_i Q U_i^T + \sigma^2 I}_{\Sigma_i})$$

All groups

$$\vec{Y} = X \vec{\beta} + U \vec{\gamma} + \vec{\varepsilon}$$

$$\begin{aligned}\vec{Y} &= \begin{pmatrix} \vec{y}_1 \\ \vdots \\ \vec{y}_m \end{pmatrix} & X &= \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} & \vec{\beta} &_{k+1} \\ \uparrow & & & & \\ \text{length } n & & n \times k+1 & & \end{aligned}$$

$$U = \text{diag} (U_1, U_2, \dots, U_m)$$

$$n_i \times m(q+1)$$

$$\vec{\gamma} = \begin{pmatrix} \vec{\gamma}_1 \\ \vdots \\ \vec{\gamma}_m \end{pmatrix} \quad m(q+1) \quad \vec{\varepsilon} = \begin{pmatrix} \vec{\varepsilon}_1 \\ \vdots \\ \vec{\varepsilon}_m \end{pmatrix} \quad n$$

$$\vec{Y} \sim \text{MVN} \left(X\vec{\beta}, \underbrace{UQU^T + \sigma^2 I_{n \times n}}_{\Sigma} \right)$$

Estimation

Recall from GLS; known $\text{cov}(\vec{Y}) = \Sigma$

$$\hat{\beta} = (X^T \Sigma^{-1} X)^{-1} \Sigma^{-1} \vec{Y} \quad (\text{maximum likelihood})$$

$$\hat{\beta} \sim \text{MVN}_{k+1} (\vec{\beta}, (X^T \Sigma^{-1} X)^{-1})$$

Strategy: first estimate variance components in Σ , then estimate β by plugging in estimate of $\Sigma, \hat{\Sigma}$.

Step 1: $\hat{\Sigma}$

Consider $A\vec{Y}$ where $A_{(n-k-1) \times n}$ full rank,
 $AX = \vec{0}$ and $AA^T = I$.

Then, $E(A\vec{Y}) = AX\vec{\beta} = \vec{0}$ and

$$A\vec{Y} \sim MVN_{n-k-1}(\vec{0}, A\Sigma A^T) (*)$$

↑ *no β -s

* includes all variance components
* as much information as possible
retained by requiring A full rank.

A 'annihilates' $X\vec{\beta}$

rows of A make up orthonormal basis
for orthogonal complement of $\text{col}(X)$

Restricted maximum likelihood (REML):

Estimate variance components by

maximizing likelihood from (*):

$$\hat{\Sigma} \text{ from } L(\sigma^2, \tau_0^2, \dots) = f(A\vec{Y})$$

(often requiring numeric methods)

Step 2 : $\hat{\beta} = (X^T \hat{\Sigma}^{-1} X)^{-1} X^T \hat{\Sigma}^{-1} \vec{y}$
 $\uparrow$ approximate GLS estimator

Properties

$\hat{\Sigma}$: less biased than MLE,
 sometimes unbiased :

In LM : $\vec{y} = X\vec{\beta} + \vec{\varepsilon}$, $\vec{\varepsilon} \sim \text{MVN}(\vec{0}, \sigma^2 I)$

MLE $\hat{\sigma}^2 = \frac{1}{n} \vec{y}^T (I - H) \vec{y}$ biased

REML $\rightarrow \hat{\sigma}^2 = \frac{1}{n-k-1} \vec{y}^T (I - H) \vec{y}$ unbiased

$\hat{\beta}$ approx $\text{MVN}_{k+1}(\vec{\beta}, (X^T \hat{\Sigma}^{-1} X)^{-1})$

$\uparrow$ can be used to obtain
 approximate tests and intervals.

[R example]

Remarks LMMS

* Estimation of var. comp. better for large m ($m \geq 5$)

* Homoskedasticity (as in LM)

$$\vec{\varepsilon} \sim \text{MVN}(\vec{0}, \sigma^2 \mathbf{I})$$

not dependent on $\mathbf{x}$

* Exogeneity

$$\vec{\gamma} \sim \text{MVN}(\vec{0}, \mathbf{Q})$$

also not dependent on $\mathbf{x}$.

violated e.g. if

the best programs were also assigned more study hours.