Lecture 5 – Matrix Notation

We can write any mixed effect model as

$$\underline{Y} = \underline{\underline{X}}\,\underline{\beta} + \underline{\underline{Z}}\,\underline{b} + \underline{\varepsilon}$$

design matrix for random effects

vector of random errors

design matrix for the fixed effects

vector of fixed effects

vector of random effects

Example 1 $\quad Y_{ij} = \beta + b_i + \varepsilon_{ij} \qquad i = 1, \dots, 6 \quad j = 1, \dots, 3$

$$b_i \sim N(0, \sigma_b^2)$$

$$\varepsilon_{ij} \sim N(0, \sigma^2)$$

$$\underset{18\times 1}{\begin{pmatrix} Y_{11} \\ Y_{12} \\ Y_{13} \\ Y_{21} \\ \vdots \\ Y_{63} \end{pmatrix}} = \underset{18\times 1}{\begin{pmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}} \underset{1\times 1}{\beta} + \underset{18\times 6}{\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ & \vdots & & & & \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}} \underset{6\times 1}{\begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_6 \end{pmatrix}} + \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{13} \\ \varepsilon_{21} \\ \vdots \\ \varepsilon_{63} \end{pmatrix}$$

$$\underline{Y} = \underline{\underline{X}}\,\underline{\beta} + \underline{\underline{Z}}\,\underline{b} + \underline{\varepsilon}$$

So
$$\mathbb{E}\underline{Y} = \mathbb{E}(X\underline{\beta} + Z\underline{b} + \underline{\varepsilon})$$
$$= X\beta$$

$$\boxed{Var(\underline{b}) = \sigma_b^2 I_{6\times 6}}$$

$$Var(\underline{Y}) = Var(X\underline{\beta} + Z\underline{b} + \underline{\varepsilon})$$
$$= \cancel{Var(X\underline{\beta})}^{\,0} + Var(Z\underline{b}) + Var(\underline{\varepsilon})$$
$$= 0 + Z\,Var(\underline{b})\,Z^T + \sigma^2 I_{18\times 18}$$

18x18 identity matrix

$$= \sigma_b^2 ZZ^T + \sigma^2 I_{18} := V \text{ say.}$$

So $\underline{Y} \sim MVN(X\underline{\beta}, V)$

V is block diagonal for our example

$$Z = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & & & \vdots & & \\ 1 & & & & & \\ 0 & 1 & 0 & 0 & 0 & 0 \\ & & \vdots & & & \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$$ZZ^T = \begin{pmatrix} \begin{matrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{matrix} & & & 0 \\ & \begin{matrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{matrix} & & \\ 0 & & \ddots & \\ 0 & 0 & & \begin{matrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{matrix} \end{pmatrix}$$

$$= \begin{pmatrix} V_1 & & & 0 \\ & V_2 & & \\ & & \ddots & \\ 0 & & & V_6 \end{pmatrix}$$ – block diagonal structure

Where $V_i = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$

$$\text{So } V = \begin{pmatrix} \tilde{V}_1 & & & 0 \\ & \tilde{V}_2 & & \\ & & \ddots & \\ 0 & & & \tilde{V}_6 \end{pmatrix}$$

Where

$$V_i = \begin{pmatrix} \sigma_b^2 + \sigma^2 & \sigma_b^2 & \sigma_b^2 \\ \sigma_b^2 & \sigma_b^2 + \sigma^2 & \sigma_b^2 \\ \sigma_b^2 & \sigma_b^2 & \sigma_b^2 + \sigma^2 \end{pmatrix}$$

So $Var(Y_{ij}) = \sigma_b^2 + \sigma^2$

$Cov(Y_{ij}, Y_{ij'}) = \sigma_b^2$ – 2 observations on same subject (same i) different block (different j) are correlated

$Cov(Y_{ij}, Y_{i'j}) = 0$ – covariance between observations on different subjects (different i)

<u>Exercise</u> What are $X, \underline{\beta}, \underline{Z}, \underline{b}$ for the sleepstudy model?

Model fitting – 'Parameter estimation'

$$\underline{Y} = X\underline{\beta} + Z\underline{b} + \underline{\varepsilon}$$

$$\underline{\varepsilon} \sim N(0, \sigma^2 I) \qquad \underline{b} \sim N(0, \Sigma)$$

$$Y \sim N(X\underline{\beta}, \underbrace{Z\Sigma Z^T + \sigma^2 I}_{V})$$

Parameters $\underline{\beta}$ & $\underline{\theta} = \{\Sigma, \sigma^2\}$ – variance parameters

How should we estimate the unknown parameters?

- ↳ classical approaches
- ↳ maximum likelihood & Restricted maximum likelihood (REML)
- ↳ Bayesian inference

We need to use ML for hypothesis testing.

default approach in the lme4 R package

In practice, ML & REML ~~are~~ give very similar answers.

Maximum likelihood

$|V| = \det(V)$ $\qquad n = \#$ of observations

Likelihood $\mathcal{L}(\beta, \sigma^2, \Sigma \mid y) = \dfrac{|V|^{-\frac{1}{2}}}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}(y - X\underline{\beta})^T V^{-1}(y - X\underline{\beta})\right)$

So $-2\log \mathcal{L} \propto \log|V| + (y - X\underline{\beta})^T V^{-1}(y - X\underline{\beta})$

The Maximum likelihood estimates are found by maximizing $\mathcal{L}$ wrt β, σ^2, Σ or minimizing $-2\log \mathcal{L}$.

<u>Example</u> Lets consider fixed effects only

$$\underline{y} = X\underline{\beta} + \underline{\varepsilon} \qquad \underline{\varepsilon} \sim N(0, \sigma^2 \underline{I})$$

So $V = \sigma^2 \underline{I}$

fitted residuals

Then $\hat{\underline{\beta}} = (X^T X)^{-1} X^T \underline{y}$

$$\& \quad \hat{\sigma}^2 = \frac{1}{n} (\underline{y} - X\hat{\underline{\beta}})^T (\underline{y} - X\hat{\underline{\beta}}) = \frac{1}{n} \hat{\underline{\varepsilon}}^T \hat{\underline{\varepsilon}}$$

& $\mathbb{E}\, \hat{\underline{\beta}} = \underline{\beta}$ so $\hat{\underline{\beta}}$ is an unbiased estimator of $\underline{\beta}$.

but $\mathbb{E}\, \hat{\sigma}^2 = \frac{n-p}{n} \sigma^2$ where $p = \dim(\underline{\beta})$

So $\hat{\sigma}^2$ is a biased estimator of σ^2.

So instead use

$$s^2 = \frac{1}{n-p} (\underline{y} - X\hat{\underline{\beta}})^T (\underline{y} - X\hat{\underline{\beta}})$$ as an estimate of σ^2.

as $\mathbb{E}\, s^2 = \sigma^2$.

<u>Message</u> Maximum likelihood gives biased variance estimates.

$\hookrightarrow$ use REML instead.