STAT 8004 Lecture 6

Thursday, March 12, 2015 5:17 PM

- Linear models: $Y = X\beta + \varepsilon$
  - estimation
  - confidence set
  - hypothesis testing

- Linear mixed models

model error

$$Y_{n\times 1} = X_{n\times k}\beta_{k\times 1} + Z_{n\times q}u_{q\times 1} + \varepsilon_{n\times 1}$$

Z: known effect/predictor

u: random vector

* Zu : contribute as random effect
* $X\beta$: contribute as fixed effect

Assume:

- $E(u) = 0$, $E(\varepsilon) = 0$

- main objective of u is to capture dependence / covariance between variables

- Standard assumptions

$$E\begin{pmatrix} u \\ \varepsilon \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \qquad \text{Var}\begin{pmatrix} u \\ \varepsilon \end{pmatrix} = \begin{pmatrix} G & 0 \\ 0 & R \end{pmatrix}$$

i.e. u & ε are uncorrelated.

- Under these assumption:

$$E(Y) = E(X\beta + Zu + \varepsilon) = X\beta$$

$$\begin{aligned} \text{var}(Y) &= \text{var}(X\beta + Zu + \varepsilon) \\ &= \text{Var}(Zu) + \text{Var}(\varepsilon) \\ &= \underbrace{ZGZ^T + R}_{\text{call this } \Sigma} \end{aligned}$$

- Example:
  - for processes producing drugs, e.g. penicillin

process A
process B
process C
process D

$\leftarrow$ level of "fixed" effect $\leftarrow$ treatment

- Different batch of raw material: corn steep liquor
- random sample of 5 batches.
  - split each batch into 4 parts

- random sample of 5 batches.
  - split each batch into 4 parts
  - run each process on one part
  - randomize the order for the process running on the part from the batch.
- Some outcomes, % of component may be of interest.

* - same batch is used for all processes
- this will introduce between batch variations
- this is additional to between process variations
- covariance / dependence due to that the same batch of materials is used.

- $$\underbrace{y_{ij}}_{\substack{\text{outcome of the} \\ i\text{th process and} \\ j\text{th batch.}}} = \underbrace{\mu + \alpha_i}_{\substack{\text{mean outcome of the} \\ i\text{th process}}} + \underbrace{u_j}_{\substack{\text{random effect from} \\ \text{the } j\text{th batch}}} + \varepsilon_{ij}$$

$(i = 1, 2, 3, 4$

$j = 1, 2, 3, 4, 5)$

Assume $u_j \sim N(0, \sigma_u^2)$

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

independence everywhere

- $E(y_{ij}) = E(\mu + \alpha_i + u_j + \varepsilon_{ij}) = \mu + \alpha_i$

$\text{var}(y_{ij}) = \text{var}(\qquad\qquad) = \sigma_u^2 + \sigma_\varepsilon^2$

- covariance between runs on the batch of material

$$\text{cov}(y_{ij}, y_{kj}) = \text{cov}(\mu + \alpha_i + \underline{\underline{u_j}} + \varepsilon_{ij}, \ \mu + \alpha_k + \underline{\underline{u_j}} + \varepsilon_{kj})$$

$$= \text{var}(u_j) = \sigma_u^2$$

- correlation

$$\rho = \frac{\text{cov}(y_{ij}, y_{kj})}{\sqrt{\text{var}(y_{ij})\,\text{var}(y_{kj})}} = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_\varepsilon^2}$$

for $i \neq k$

- outcomes on different batches are uncorrelated.

- $$\text{var}\begin{pmatrix} y_{1j} \\ y_{2j} \\ y_{3j} \\ y_{4j} \end{pmatrix} = \begin{pmatrix} \sigma_u^2+\sigma_\varepsilon^2 & \sigma_u^2 & \sigma_u^2 & \sigma_u^2 \\ \sigma_u^2 & \sigma_u^2+\sigma_\varepsilon^2 & \sigma_u^2 & \sigma_u^2 \\ \sigma_u^2 & \sigma_u^2 & \sigma_u^2+\sigma_\varepsilon^2 & \sigma_u^2 \\ \sigma_u^2 & \sigma_u^2 & \sigma_u^2 & \sigma_u^2+\sigma_\varepsilon^2 \end{pmatrix} = B$$

  - This is the so called compound symmetric structure

$$\begin{pmatrix} 1 & & \rho \\ & 1 & \\ \rho & & 1 \\ & & & 1 \end{pmatrix}$$

—x—

- $Y = X\beta + Zu + \varepsilon$

—x—

- $Y = \underset{4\times5}{X\beta} + Zu + \varepsilon$

$$\begin{bmatrix} Y_{11}\\ Y_{21}\\ Y_{31}\\ Y_{41}\\ Y_{12}\\ Y_{22}\\ Y_{32}\\ Y_{42}\\ Y_{13}\\ Y_{23}\\ Y_{33}\\ Y_{43}\\ Y_{14}\\ Y_{24}\\ Y_{34}\\ Y_{44}\\ Y_{15}\\ Y_{25}\\ Y_{35}\\ Y_{45} \end{bmatrix} = \begin{bmatrix} 1&1&0&0&0\\ 1&0&1&0&0\\ 1&0&0&1&0\\ 1&0&0&0&1\\ 1&1&0&0&0\\ 1&0&1&0&0\\ 1&0&0&1&0\\ 1&0&0&0&1\\ 1&1&0&0&0\\ 1&0&1&0&0\\ 1&0&0&1&0\\ 1&0&0&0&1\\ 1&1&0&0&0\\ 1&0&1&0&0\\ 1&0&0&1&0\\ 1&0&0&0&1\\ 1&1&0&0&0\\ 1&0&1&0&0\\ 1&0&0&1&0\\ 1&0&0&0&1 \end{bmatrix} \begin{bmatrix} \mu\\ \alpha_1\\ \alpha_2\\ \alpha_3\\ \alpha_4 \end{bmatrix}$$

$$+ \begin{bmatrix} 1&0&0&0&0\\ 1&0&0&0&0\\ 1&0&0&0&0\\ 1&0&0&0&0\\ 0&1&0&0&0\\ 0&1&0&0&0\\ 0&1&0&0&0\\ 0&1&0&0&0\\ 0&0&1&0&0\\ 0&0&1&0&0\\ 0&0&1&0&0\\ 0&0&1&0&0\\ 0&0&0&1&0\\ 0&0&0&1&0\\ 0&0&0&1&0\\ 0&0&0&1&0\\ 0&0&0&0&1\\ 0&0&0&0&1\\ 0&0&0&0&1\\ 0&0&0&0&1 \end{bmatrix} \begin{pmatrix} u_1\\ u_2\\ u_3\\ u_4\\ u_5 \end{pmatrix} \begin{bmatrix} \beta_1\\ \beta_2\\ \beta_3\\ \beta_4\\ \beta_5 \end{bmatrix} + \begin{bmatrix} e_{11}\\ e_{21}\\ e_{31}\\ e_{41}\\ e_{12}\\ e_{22}\\ e_{32}\\ e_{42}\\ e_{13}\\ e_{23}\\ e_{33}\\ e_{43}\\ e_{14}\\ e_{24}\\ e_{34}\\ e_{44}\\ e_{15}\\ e_{25}\\ e_{35}\\ e_{45} \end{bmatrix}$$

- $R = var(\varepsilon) = \sigma_\varepsilon^2 I_n$

$$G = var(u) = var\begin{pmatrix} u_1\\ u_2\\ \vdots\\ u_5 \end{pmatrix} = \sigma_u^2 I_{5\times5}$$

$$Var(Y) = ZGZ^T + R$$

$$= \sigma_u^2 ZZ^T + \sigma_\varepsilon^2 I$$

$$= \begin{pmatrix} B & & & 0\\ & B & & \\ & & B & \\ 0 & & & B \end{pmatrix}$$

matrix of all ones

$$= \underline{\sigma_u^2 I_{4\times4} \otimes J_{5\times5}} + \underline{\sigma_\varepsilon^2 I_{n\times n}}$$

$\otimes$: Kronecker product.

$$\underset{=}{A_{k\times m} \otimes B_{n\times s}} = \begin{pmatrix} a_{11}\underline{\underline{B}} & a_{12}B & \cdots & a_{1m}B\\ \vdots & & & \vdots\\ a_{k1}\underline{\underline{B}} & \cdots & \cdots & a_{km}B \end{pmatrix}$$

example: $A = \begin{pmatrix} 1 & 0\\ 0 & 1 \end{pmatrix}$ $\underline{B = \begin{pmatrix} 1&1&1\\ 1&1&1\\ 1&1&1 \end{pmatrix}}$

$$A \otimes B = \begin{pmatrix} \begin{smallmatrix}1&1&1\\1&1&1\\1&1&1\end{smallmatrix} & 0\\ 0 & \begin{smallmatrix}1&1&1\\1&1&1\\1&1&1\end{smallmatrix} \end{pmatrix}.$$

—x—

- Hierarchical random effect model.

- Analysis of the sources of variations in a process of monitoring the production of pigment paste.

random

Analysis of the sources of variations in a process of monitoring the production of pigment paste.

- random Sample b barrels of pastes.
- s random samples taken from each barrel
- each sample is then treated and divided into r parts.
- send the samples to a lab for measuring the moisture content.

- $n = (b)(s)(r)$: total # of measurements

$$y_{ijk} = \mu + \beta_i + \delta_{ij} + \varepsilon_{ijk}$$

y_{ijk}: k part of j th sample in i th barrel.

β_i: barrel effect

δ_{ij}: sample effect in each barrel

ε_{ijk}: random error.

$$\beta_j \sim N(0, \sigma_\beta^2)$$
$$\delta_{ij} \sim N(0, \sigma_\delta^2) \quad \text{independence}$$
$$\varepsilon_{ijk} \sim N(0, \sigma_\varepsilon^2)$$

- $var(y_{ijk}) = \sigma_\beta^2 + \sigma_\delta^2 + \sigma_\varepsilon^2$

- Two parts from the same sample

$cov(y_{ijk}, y_{ijl})$

$= cov(\mu + \underline{\beta_i} + \underline{\delta_{ij}} + \underline{\underline{\varepsilon_{ijk}}},\ \mu + \underline{\beta_i} + \underline{\delta_{ij}} + \underline{\underline{\varepsilon_{ijl}}})$

$= \sigma_\beta^2 + \sigma_\delta^2$ for $k \neq l$

- Samples from the barrel.

$cov(y_{ijk}, y_{iml})$

$= \cdots\cdots = \sigma_\beta^2$

- let us take. $b = 15$, $s = 2$, $r = 2$

- $Y = X\beta + Zu + \varepsilon$

$$\begin{bmatrix} \mathbf{Y}_{111} \\ \mathbf{Y}_{112} \\ \mathbf{Y}_{121} \\ \mathbf{Y}_{122} \\ \mathbf{Y}_{211} \\ \mathbf{Y}_{212} \\ \mathbf{Y}_{221} \\ \mathbf{Y}_{222} \\ \vdots \\ \mathbf{Y}_{15,1,1} \\ \mathbf{Y}_{15,1,2} \\ \mathbf{Y}_{15,2,1} \\ \mathbf{Y}_{15,2,2} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ \vdots \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} [\underline{\underline{\mu}}] +$$

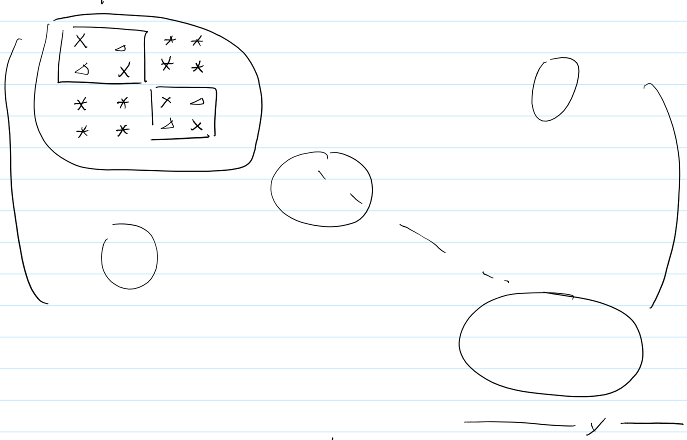
$$\begin{bmatrix} 1 & 0 & \dots & 0 & 1 & 0 & 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 1 & 0 & 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 & 1 & 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 & 1 & 0 & 0 & \dots & 0 & 0 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \end{bmatrix}$$

$$
\begin{bmatrix}
1 & 0 & \dots & 0 & 1 & 0 & 0 & 0 & \dots & 0 & 0 \\
1 & 0 & \dots & 0 & 1 & 0 & 0 & 0 & \dots & 0 & 0 \\
1 & 0 & \dots & 0 & 0 & 1 & 0 & 0 & \dots & 0 & 0 \\
1 & 0 & \dots & 0 & 0 & 1 & 0 & 0 & \dots & 0 & 0 \\
0 & 1 & \dots & 0 & 0 & 0 & 1 & 0 & \dots & 0 & 0 \\
0 & 1 & \dots & 0 & 0 & 0 & 1 & 0 & \dots & 0 & 0 \\
0 & 1 & \dots & 0 & 0 & 0 & 0 & 1 & \dots & 0 & 0 \\
0 & 1 & \dots & 0 & 0 & 0 & 0 & 1 & \dots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & \dots & 1 & 0 & 0 & 0 & 0 & \dots & 1 & 0 \\
0 & 0 & \dots & 1 & 0 & 0 & 0 & 0 & \dots & 1 & 0 \\
0 & 0 & \dots & 1 & 0 & 0 & 0 & 0 & \dots & 0 & 1 \\
0 & 0 & \dots & 1 & 0 & 0 & 0 & 0 & \dots & 0 & 1
\end{bmatrix}
\begin{bmatrix}
\beta_1 \\ \beta_2 \\ \vdots \\ \beta_{15} \\ \delta_{1,1} \\ \delta_{1,2} \\ \delta_{2,1} \\ \delta_{2,2} \\ \vdots \\ \delta_{15,1} \\ \delta_{15,2}
\end{bmatrix}
+ \mathbf{e}
$$

- $R = var(\varepsilon) = \sigma_\varepsilon^2 I_{n\times n}$
- $G = var(u) = \begin{pmatrix} \sigma_\beta^2 I_b & 0 \\ 0 & \sigma_\delta^2 I_{bs} \end{pmatrix}$
- $var(Y) = ZGZ^T + R$

$$= Z \begin{pmatrix} \sigma_\beta^2 I & \\ & \sigma_\delta^2 I \end{pmatrix} Z^T + \sigma_\varepsilon^2 I$$

$$= \sigma_\beta^2 I_b \otimes J_{sr} + \sigma_\delta^2 I_{bs} \otimes J_r + \sigma_\varepsilon^2 I_{bsr}$$



- Remark: to incorporate dependence / covariance.

- Analysis of Linear mixed models.
  - This is not the classical setting.

  Now, we have $E(Y) = X\beta$

  but $var(Y) = \underbrace{ZGZ^T + R}_{}$

  - Σ is called the variance component.
  - it contains unknown parameters e.g. σ_ε^2, σ_u^2

    $\sigma_\varepsilon^2 \quad \sigma_\delta^2 \quad \sigma_\beta^2$

we want to:

1) inference about the estimable fix effect
   - point estimation
   - confidence set
   - testing.

2) inference about the variance component

3) predicting the random effect

4) predicting future observations.

- To what extent does the OLS still work?
  - can still write $Y = X\beta + \underline{\varepsilon^*}$ $(\varepsilon^* = \underline{Zu + \varepsilon})$

  - OLS for estimable function $c^T\beta$

$$\widehat{c^T\beta} = c^T(X^TX)^- X^T Y$$

  it is a linear function of Y

$$\underline{E(\widehat{c^T\beta})} = c^T(X^TX)^- X^T E(Y)$$
$$= c^T(X^TX)^- X^T X\beta \qquad (c^T = a^T X)$$
$$= a^T X(X^TX)^- X^T X\beta$$
$$= a^T X\beta = \underline{\underline{c^T\beta}}$$

  - it is unbiased for $c^T\beta$.

STOP
  - efficiency. No!
  - Confidence set using previous formula. No!

    ↳ biased

  - GLS estimator for β (if Σ is known)

$$\underline{\widehat{c^T\beta}_{GLS}} = c^T(X^T\Sigma^{-1}X)^- X^T\Sigma^{-1}Y$$

  with Σ estimated by $\hat{\Sigma}$ REML.

$$\boxed{\widehat{c^T\beta}_{GLS} = c^T(X^T\hat{\Sigma}^{-1}X)^- X^T\hat{\Sigma}^{-1}Y} \leftarrow$$

  - This is for the fixed effects.

  - We are discussing now the MLE, & REML.

MLE: (Assume normality everywhere)

  then $Y \sim N(X\beta, \Sigma)$ $(\Sigma = ZGZ^T + R)$

  Now, let v be the vector containing unknown parameters in Σ. (e.g. $\sigma^2_\varepsilon, \sigma^2_\beta, \sigma^2_\delta$)

  Write Σ as $\Sigma(v)$.

Then, the likelihood function using Y

$L(X\beta, \Sigma(v))$

$$= (2\pi)^{-n/2} \left(\det \Sigma(v)\right)^{-1/2} \exp\left(-\frac{1}{2}(Y-X\beta)^T \Sigma^{-1}(v)(Y-X\beta)\right)$$

if we write then for v fixed

$$\widehat{X\beta(v)} = X(X^T\Sigma^{-1}(v)X)^{-} X^T \underbrace{\Sigma^{-1}(v)} Y$$

Now: we have a profile likelihood for v

$$- \underbrace{L^*(v)} = L(\widehat{X\beta}(v), \Sigma(v)).$$

$\uparrow$

only a function of v

- MLE is then obtained by maximizing $L^*(v)$.

- It is very computationally intensive, especially when Σ is large, and when v has many parameters.

- Another issue, is that MLE tends to "underestimate" the variance component.

recall, $\hat{\sigma}^2_{MLE} = \dfrac{SSE}{n}$

$\wedge$

$s^2 = \dfrac{SSE}{n - \text{rank}(X)}$ is unbiased.

- To fix this underestimation, one is to use the so-called restricted maximum likelihood (REML)

Rationale:

- for Y with 0 mean, $\boxed{\underline{\frac{1}{n}} \sum_{i=1}^n y_i^2}$ is the MLE & it is unbiased.

- but when it contains estimated mean, it will not be the case. because it still uses $\frac{1}{n}\sum_{i=1}^n (y_i - \hat{y}_i)^2$, even $\underline{\underline{\frac{1}{n}}} \sum_{i=1}^n (y_i - \bar{y})^2$ will not be unbiased

Example: $n = 4$, $Y \sim N(1\mu, \sigma^2 I)$.

$$1\mu = \begin{pmatrix} \mu \\ \mu \\ \mu \\ \mu \end{pmatrix}$$

$$\begin{pmatrix}\mu\\ \mu\\ \mu\\ \mu\end{pmatrix}$$

Then, $e = \begin{pmatrix} Y_1-\bar{Y} \\ Y_2-\bar{Y} \\ Y_3-\bar{Y} \\ Y_4-\bar{Y} \end{pmatrix} = (I-P_1)Y$

$$P_1 = \begin{pmatrix}1\\1\\1\\1\end{pmatrix}\left((1\,1\,1\,1)\begin{pmatrix}1\\1\\1\\1\end{pmatrix}\right)^{-1}(1\,1\,1\,1)$$

$$= \frac{1}{4}\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}.$$

- $\underset{4\times 1}{e} \sim N(0, \sigma^2(I-P_1))$

but $\text{rank}(I-P_1) = \underline{\underline{3}}$

- However, the singular matrix is "more appropriate" in the sense that 3 is what we want for the denominator.

- Define $r = \underbrace{\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}}_{M} \begin{pmatrix} Y_1-\bar{Y} \\ Y_2-\bar{Y} \\ Y_3-\bar{Y} \\ Y_4-\bar{Y} \end{pmatrix} \hat{=} M(I-P_1)Y$

$$\underline{\underline{r_{3\times 1}}} \sim N(0, \sigma^2 \underbrace{M(I-P_1)M^T}_{W})$$

- Adjust the data as aforementioned, and then run MLE on r. (this is called REML)

Now, work on the likelihood function of σ^2 based on r.

$$L(\sigma^2) = (2\pi)^{-3/2}\left(\det(\sigma^2 W)\right)^{-1/2}\exp\left(-\frac{1}{2\sigma^2} r^T \underline{\underline{W}}^{-1} r\right)$$

log likelihood

$$\ell(\sigma^2) = -\frac{3}{2}\log(2\pi) - \frac{3}{2}\log(\sigma^2) - \frac{1}{2}\log(W) - \frac{1}{2\sigma^2} r^T W^{-1} r$$

$$\frac{d\ell}{d\sigma^2} = -\frac{3}{2}\frac{1}{\sigma^2} + \frac{1}{2(\sigma^2)^2} r^T W^{-1} r$$

$$\Rightarrow \sigma^2_{REML} = \frac{1}{3} r^T W^{-1} r$$

$r = M(I-P_1)Y. \quad W = M(I-P_1)M^T$

$$= \frac{1}{3} Y^T \boxed{\underbrace{(I-P_1)}\, M^T (M(I-P_1)M^T)^{-1} M\underbrace{(I-P_1)}} Y$$

It is actually projecting Y onto the column space of $\underbrace{M(I-P_1)}$, which is the same as $\underbrace{(I-P_1)}$.

$$= \frac{1}{3} Y^T (I-P_1) Y \qquad ! \quad \text{done!}$$

$$= s^2$$

- Actually, the specific form of M is not used. That is, it is valid for any M of rank 3.

$$\text{e.g. } M = \begin{pmatrix} 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}$$

- Generalize the case to P_X from P_1 is conceptually the same. replace the quantities by $\boxed{M(I-P_X)}Y \triangleq BY$ with $\text{rank}(M) = n - \text{rank}(X)$.

- Typically, one use BY, with $BX = 0$. & $\text{rank}(B) = n - \text{rank}(X)$.

- REML, use BY with $BX=0$, to get the estimator for V in $\Sigma(V)$.

- To see the convenience

$$\boxed{BY} = B(X\beta + Zu + \varepsilon)$$

$$= BZu + B\varepsilon \quad \text{is invariant of } X.$$

Then, $r = BY = M(I-P_X)Y$

$$\sim N_{n-\text{rank}(X)}(0, B\Sigma B^T)$$

- To avoid efficiency loss, $\text{rank}(M) = n - \text{rank}(X)$ has to be the case.

— to avoid identifiability issue, $\text{rank}(M) = n - \text{rank}(X)$
has to be the case.

$$L_r(v) = (2\pi)^{-m/2} \left(\det B\Sigma(v)B^T\right)^{-1/2} \exp\left(-\frac{1}{2} r^T \left(B\Sigma(v)B^T\right)^{-1} r\right)$$

$m = n - \text{rank}(X)$

It is just a normal likelihood function, but it is using $BY = r$.

— Fact: — σ^2_{REML} is typically larger than σ^2_{MLE} for relevant variance components.

$$\left(\Sigma = Z G Z^T + R_{n \times n} \right)$$

Remark: if the # of parameter in V is larger than n, then identifiability issue occurs.

— σ^2_{REML} is invariant to the choice of M.

— REML will provide $\widehat{G}$ & $\widehat{R}$, so $\widehat{\Sigma} = Z\widehat{G}Z^T + \widehat{R}$.

— inference for $c^T\beta$, estimable fix effect

— predicting random effect u

— inference for variance component.

model

implementation (R)