Mixed Model Analysis

Basic model:

$$\mathbf{Y} = X\boldsymbol{\beta} + Z\mathbf{u} + \mathbf{e},$$

where

- X is a $n \times p$ model matrix of known constants
- $\boldsymbol{\beta}$ is a $p \times 1$ vector of *fixed* unknown parameter values
- Z is a $n \times q$ model matrix of known constants
- $\mathbf{u}$ is a $q \times 1$ random vector
- $\mathbf{e}$ is a $n \times 1$ vector of random errors

with

$$\begin{aligned} E(\mathbf{e}) &= \mathbf{0} \qquad\qquad Var(\mathbf{e}) = R \\ E(\mathbf{u}) &= \mathbf{0} \qquad\qquad Var(\mathbf{u}) = G \\ Cov(\mathbf{e}, \mathbf{u}) &= 0 \end{aligned}$$

Then

$$\begin{aligned} E(\mathbf{Y}) &= E(X\boldsymbol{\beta} + Z\mathbf{u} + \mathbf{e}) \\ &= X\boldsymbol{\beta} + ZE(\mathbf{u}) + E(\mathbf{e}) \\ &= X\boldsymbol{\beta} \end{aligned}$$

$$\begin{aligned} Var(\mathbf{Y}) &= Var(X\boldsymbol{\beta} + Z\mathbf{u} + \mathbf{e}) \\ &= Var(Z\mathbf{u}) + Var(\mathbf{e}) \\ &= ZGZ^T + R \end{aligned}$$

Normal-Theory Mixed Model

$$\begin{bmatrix} \mathbf{u} \\ \mathbf{e} \end{bmatrix} \sim N\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} G & 0 \\ 0 & R \end{bmatrix} \right)$$

Then,

$$\mathbf{Y} \sim N(X\boldsymbol{\beta}, \underline{ZGZ^T + R})$$

$$\uparrow$$

call this Σ

Example 9.1: Random Blocks

Comparison of four processes for producing penicillin

$$\left.\begin{array}{l} Process\ \ A \\ Process\ \ B \\ Process\ \ C \\ Process\ \ D \end{array}\right\} \qquad \begin{array}{c}\text{Levels of a ``fixed''} \\ \text{treatment factor}\end{array}$$

Blocks correspond to different batches of an important raw material, corn steep liquor

- Random sample of five batches
- Split each batch into four parts:
    - run each process on one part
    - randomize the order in which the processes are run within each batch

Here, batch effects are considered as *random* block effects:

- Batches are sampled from a population of many possible batches
- To repeat this experiment you would need to use a different set of batches of raw material

Data Source: Box, Hunter & Hunter (1978), *Statistics for Experimenters*. (Wiley & Sons, New York).

Data file: penclln.dat

SAS code: penclln.sas

R code: penclln.r

Model:

$$Y_{ij} = \mu + \alpha_i + \beta_j + e_{ij}$$

Y_{ij} $\uparrow$	α_i $\uparrow$	β_j $\uparrow$	e_{ij} $\uparrow$
Yield for the i-th process applied to the j-th batch	mean yield for the i-th process, averaging across the entire population of possible batches	random batch effect	random error

where

$$\beta_j \sim NID(0, \sigma^2_\beta), \quad e_{ij} \sim NID(0, \sigma^2_e)$$

and any e_{ij} is independent of any β_j.

Here

$$\begin{aligned} \mu_i = E(Y_{ij}) &= E(\mu + \alpha_i + \beta_j + e_{ij}) \\ &= \mu + \alpha_i + E(\beta_j) + E(e_{ij}) \\ &= \mu + \alpha_i \qquad i = 1, 2, 3, 4 \end{aligned}$$

represents the mean yield for the *i*-th process, averaging across all possible batches.

PROC GLM and PROC MIXED in SAS fit a restricted model with $\alpha_4 = 0$. Then

- $\mu = \mu_4$ is the mean yield for process D
- $\alpha_i = \mu_i - \mu_4 \qquad i = 1, 2, 3, 4.$

In R you could use the *treatment* constraints where $\alpha_1 = 0$. Then

- $\mu = \mu_1$ is the mean yield for process A
- $\alpha_i = \mu_i - \mu_1 \qquad i = 1, 2, 3, 4.$

Alternatively, you could choose the solution to the normal equations given by *sum* constraints

- $\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 = 0$
- $\mu = (\mu_1 + \mu_2 + \mu_3 + \mu_4)/4$
- $\alpha_i = \mu_i - \mu$ is the difference between the mean yield for the i-th process and the overall mean yield.

Variance-covariance structure:

$$
\begin{aligned}
Var(Y_{ij}) &= Var(\mu + \alpha_i + \beta_j + e_{ij}) \\
&= Var(\beta_j + e_{ij}) \\
&= Var(\beta_j) + Var(e_{ij}) \\
&= \sigma_\beta^2 + \sigma_e^2 \qquad \text{for all } (i,j)
\end{aligned}
$$

Different runs on the same batch:

$$
\begin{aligned}
Cov(Y_{ij}, Y_{kj}) &= Cov(\mu + \alpha_i + \beta_j + e_{ij},\ \mu + \alpha_k + \beta_j + e_{kj}) \\
&= Cov(\beta_j + e_{ij},\ \beta_j + e_{kj}) \\
&= Cov(\beta_j, \beta_j) + Cov(\beta_j,\ e_{kj}) + Cov(e_{ij}, \beta_j) + Cov(e_{ij},\ e_{kj}) \\
&= Var(\beta_j) \\
&= \sigma_\beta^2 \qquad \text{for all } i \neq k
\end{aligned}
$$

Correlation among yields for runs on the same batch:

$$\rho = \frac{Cov(Y_{ij}, Y_{kj})}{\sqrt{Var(Y_{ij})Var(Y_{kj})}}$$

$$= \frac{\sigma_\beta^2}{\sigma_\beta^2 + \sigma_e^2} \text{ for } i \neq k$$

Results for runs on different batches are uncorrelated (independent):

$$Cov(Y_{ij}, Y_{k\ell}) = 0 \quad \text{for j} \neq \ell$$

Results from the four runs on a single batch:

$$
Var \begin{bmatrix} Y_{1j} \\ Y_{2j} \\ Y_{3j} \\ Y_{4j} \end{bmatrix} = \begin{bmatrix} \sigma_\beta^2 + \sigma_e^2 & \sigma_\beta^2 & \sigma_\beta^2 & \sigma_\beta^2 \\ \sigma_\beta^2 & \sigma_\beta^2 + \sigma_e^2 & \sigma_\beta^2 & \sigma_\beta^2 \\ \sigma_\beta^2 & \sigma_\beta^2 & \sigma_\beta^2 + \sigma_e^2 & \sigma_\beta^2 \\ \sigma_\beta^2 & \sigma_\beta^2 & \sigma_\beta^2 & \sigma_\beta^2 + \sigma_e^2 \end{bmatrix}
$$

$$
= \sigma_\beta^2 J + \sigma_e^2 I
$$

This special type of covariance structure is called *compound symmetry*.

Write this model as $\mathbf{Y} = X\boldsymbol{\beta} + Z\mathbf{u} + \mathbf{e}$

$$
\begin{bmatrix} Y_{11} \\ Y_{21} \\ Y_{31} \\ Y_{41} \\ Y_{12} \\ Y_{22} \\ Y_{32} \\ Y_{42} \\ Y_{13} \\ Y_{23} \\ Y_{33} \\ Y_{43} \\ Y_{14} \\ Y_{24} \\ Y_{34} \\ Y_{44} \\ Y_{15} \\ Y_{25} \\ Y_{35} \\ Y_{45} \end{bmatrix}
=
\begin{bmatrix}
1 & 1 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & 1 \\
1 & 1 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & 1 \\
1 & 1 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & 1 \\
1 & 1 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & 1 \\
1 & 1 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & 1
\end{bmatrix}
\begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{bmatrix}
+
\begin{bmatrix}
1 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 1
\end{bmatrix}
\begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{bmatrix}
+
\begin{bmatrix} e_{11} \\ e_{21} \\ e_{31} \\ e_{41} \\ e_{12} \\ e_{22} \\ e_{32} \\ e_{42} \\ e_{13} \\ e_{23} \\ e_{33} \\ e_{43} \\ e_{14} \\ e_{24} \\ e_{34} \\ e_{44} \\ e_{15} \\ e_{25} \\ e_{35} \\ e_{45} \end{bmatrix}
$$

Here

$$
\begin{aligned}
G &= Var(\mathbf{u}) = \sigma_B^2 I_{5\times 5} \\
R &= Var(\mathbf{e}) = \sigma_e^2 I_{n\times n}
\end{aligned}
$$

and

$$
\begin{aligned}
Var(\mathbf{Y}) &= Var(X\beta + Z\mathbf{u} + \mathbf{e}) \\
&= Var(Z\mathbf{u}) + Var(\mathbf{e}) \\
&= ZGZ^T + R \\
&= \sigma_\beta^2 ZZ^T + \sigma_e^2 I \\
&= \begin{bmatrix} \sigma_\beta^2 J + \sigma_e^2 I & & & \\ & \sigma_\beta^2 J + \sigma_e^2 I & & \\ & & \ddots & \\ & & & \sigma_\beta^2 J + \sigma_e^2 I \end{bmatrix}
\end{aligned}
$$

Example 9.2: Hierarchical Random Effects Model

Analysis of sources of variation in a process used to monitor the production of a pigment paste.

Current Procedure:

- Sample barrels of pigment paste
- One sample from each barrel
- Send the sample to a lab for determination of moisture content

Measured Response: (Y) moisture content of the pigment paste (units of one tenth of 1%).

Problem: Variation in moisture content is too large

- average moisture content is approximately 25 (or 2.5%)
- standard deviation of about 6

Examine sources of variation:

Data Collection: Hierarchical (or nested) Study Design

- Sample b barrels of pigment paste
- s samples are taken from the content of each barrel
- Each sample is mixed and divided into r parts. Each part is sent to the lab.

There are $n = (b)(s)(r)$ observations.

Model:

$$Y_{ijk} \; = \; \mu \; + \; \beta_i \; + \; \delta_{ij} \; + \; e_{ijk}$$

where

Y_{ijk} is the moisture content determination for the k-th part of the j-th sample from the i-th barrel

μ is the mean moisture content

β_i is a random barrel effect:

$$\beta_i \sim NID(0, \sigma^2_\beta)$$

δ_{ij} is a random sample effect:

$$\delta_{ij} \sim NID(0, \sigma^2_\delta)$$

e_{ijk} corresponds to random measurement error:

$$e_{ijk} \sim NID(0, \sigma^2_e)$$

Covariance Structure

Homogeneous variances:

$$\begin{aligned} Var(Y_{ijk}) &= Var(\mu + \beta_i + \delta_{ij} + e_{ijk}) \\ &= Var(\beta_i) + Var(\delta_{ij}) + Var(e_{ijk}) \\ &= \sigma^2_\beta + \sigma^2_\delta + \sigma^2_e \end{aligned}$$

Two parts of one sample:

$$\begin{aligned} & Cov(Y_{ijk}, Y_{ij\ell}) \\ = \; & Cov(\mu + \beta_i + \delta_{ij} + e_{ijk}, \mu + \beta_i + \delta_{ij} + e_{ij\ell}) \\ = \; & Cov(\beta_i, \beta_i) + Cov(\delta_{ij}, \delta_{ij}) \\ = \; & \sigma^2_\beta + \sigma^2_\delta \qquad \text{for } k \neq \ell \end{aligned}$$

Observations on different samples taken from the same barrel:

$$
\begin{aligned}
Cov(Y_{ijk}, Y_{im\ell}) &= Cov(\mu + \beta_i + \delta_{ij} + e_{ijk}, \mu + \beta_i + \delta_{im} + e_{im\ell}) \\
&= Cov(\beta_i, \beta_i) \\
&= \sigma^2_\beta \quad j \neq m
\end{aligned}
$$

Observations from different barrels:

$$
Cov(Y_{ijk}, Y_{cm\ell}) = 0, \quad i \neq c
$$

In this study

$b = 15$ barrels were sampled

$s = 2$ samples were taken from each barrel

$r = 2$ sub-samples were analyzed from each sample taken from each barrel

Data file: pigment.dat

SAS code: pigment.sas

R code: pigment.r

Write this model in the form:

$$\mathbf{Y} = X\beta + Z\mathbf{u} + \mathbf{e}$$

$$\begin{bmatrix} Y_{111} \\ Y_{112} \\ Y_{121} \\ Y_{122} \\ Y_{211} \\ Y_{212} \\ Y_{221} \\ Y_{222} \\ \vdots \\ Y_{15,1,1} \\ Y_{15,1,2} \\ Y_{15,2,1} \\ Y_{15,2,2} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ \vdots \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} \mu \end{bmatrix} + \begin{bmatrix} 1 & 0 & \dots & 0 & 1 & 0 & 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 1 & 0 & 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 & 1 & 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 & 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 & 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 & 0 & 0 & 1 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 & 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 & 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & \dots & 1 & 0 & 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & \dots & 1 & 0 & 0 & 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & \dots & 1 & 0 & 0 & 0 & 0 & \dots & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{15} \\ \delta_{1,1} \\ \delta_{1,2} \\ \delta_{2,1} \\ \delta_{2,2} \\ \vdots \\ \delta_{15,1} \\ \delta_{15,2} \end{bmatrix} + \mathbf{e}$$

where

$$R = Var(\mathbf{e}) = \sigma_e^2 I, \quad G = Var(\mathbf{u}) = \begin{bmatrix} \sigma_\beta^2 I & 0 \\ 0 & \sigma_\delta^2 I \end{bmatrix}$$

Then

$$\begin{aligned}
E(\mathbf{Y}) &= X\beta = \mathbf{1}\boldsymbol{\mu} \\
Var(\mathbf{Y}) &= \Sigma \;=\; ZGZ^T + R \;=\; Z \begin{bmatrix} \sigma_\beta^2 I_b & 0 \\ 0 & \sigma_\delta^2 I_{bs} \end{bmatrix} Z^T + \sigma_e^2 I_{bsr} \\
&= \sigma_\beta^2 (I_b \otimes J_{sr}) + \sigma_\delta^2 (I_{bs} \otimes J_r) + \sigma_e^2 I_{bsr}
\end{aligned}$$

because $$Z = \left[I_b \otimes \underbrace{\mathbf{1}_{sr}}_{\substack{(sr)\times 1 \\ \text{vector} \\ \text{of ones}}} \middle| I_{bs} \otimes \underbrace{\mathbf{1}_r}_{\substack{r \times 1 \\ \text{vector} \\ \text{of ones}}} \right]$$

Example 9.3 A split-plot experiment with whole plots arranged in blocks.

Blocks: $r = 4$ fields (or locations).

Whole plots: Each field is divided into $a = 2$ whole plots.

Whole plot factor: two cultivars of grasses (A, B)

- within each block, cultivar A is grown in one whole plot, cultivar B is grown in the other
- separate random assignments of cultivars to whole plots is done in each block

Sub-plot factor:

$b = 3$ bacterial innoculation treatments:

CON	for control (no innoculation)
DEA	for dead
LIV	for live

Each whole plot is split into three sub-plots and independent random assignments of innoculation treatments to sub-plots are done within whole plots.

Measured response:

Dry weight yield

Source:

Littel, R.C. Freund, R.J. and Spector, P.C. (1991) SAS Systems for Linear Models, 3rd edition, SAS Institute, Cary, NC

Data: grass.dat

SAS code: grass.sas

S-PLUS code: grass.r

Block 1

Cultivar B	CON	DEA	LIV
Cultivar A	LIV	CON	DEA

Block 2

DEA	LIV
LIV	CON
CON	DEA
Cultivar A	Cultivar B

Block 3

DEA	CON
CON	DEA
LIV	LIV
Cultivar B	Cultivar A

Block 4

LIV	DEA	CON	Cultivar B
CON	DEA	LIV	Cultivar A

Model with random block effects:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \eta_{ij} + \gamma_k + \delta_{ik} + e_{ijk},$$

$$i = 1, \ldots, a, \quad j = 1, \ldots, r, \quad k = 1, \ldots, b,$$

$Y_{ijk} \Rightarrow$ observed yield for the k-th innoculant applied to the i-th cultivar in the j-th field

$\alpha_i \Rightarrow$ fixed cultivar effect

$\gamma_k \Rightarrow$ fixed innoculant effect

$\delta_{ik} \Rightarrow$ cultivar*innoculant interaction

The following random effects are independent of each other:

$\beta_j \sim NID(0, \sigma_\beta^2) \Rightarrow$ random block effects

$\eta_{ij} \sim NID(0, \sigma_w^2) \Rightarrow$ random whole plot effects

$e_{ijk} \sim NID(0, \sigma_e^2) \Rightarrow$ random errors

Example 9.4: Repeated Measures

In an exercise therapy study, subjects were assigned to one of three weightlifting programs

(i=1) The number of repetitions of weightlifting was increased as subjects became stronger (RI)

(i=2) The amount of weight was increased as subjects became stronger (WI)

(i=3) Subjects did not participate in weightlifting (XCont)

Measurements of strength (Y) were taken on days 2, 4, 6, 8, 10, 12 and 14 for each subject.

Source: Littel, Freund, and Spector (1991) SAS System for Linear Models

Data: weight2.dat

SAS code: weight2.sas

R code: weight2.r

Mixed model

$$Y_{ijk} = \mu + \alpha_i + S_{ij} + \tau_k + \gamma_{ik} + e_{ijk}$$

Y_{ijk} strength measurement at the k-th time point for the j-th subject in the i-th program

α_i *fixed* program effect

S_{ij} random subject effect

τ_k *fixed* time effect

e_{ijk} random error

where the random effects are all independent and

$$S_{ij} \sim NID(0, \sigma_S^2), \quad \epsilon_{ijk} \sim NID(0, \sigma_\epsilon^2)$$

Average strength after $2k$ days on the i-th program is

$$\begin{aligned}
\mu_{ik} &= E(Y_{ijk}) \\
&= E(\mu + \alpha_i + S_{ij} + \tau_k + \gamma_{ik} + e_{ijk}) \\
&= \mu + \alpha_i + E(S_{ij}) + \tau_k + \gamma_{ik} + E(e_{ijk}) \\
&= \mu + \alpha_i + \tau_k + \gamma_{ik}
\end{aligned}$$

for $i = 1, 2, 3$ and $k = 1, 2, \ldots, 7$. The variance of any single observation is

$$\begin{aligned}
Var(Y_{ijk}) &= Var(\mu + \alpha_i + S_{ij} + \tau_k + \alpha_{ik} + e_{ijk}) \\
&= Var(S_{ij} + e_{ijk}) \\
&= Var(S_{ij}) + Var(e_{ijk}) \\
&= \sigma_S^2 + \sigma_e^2
\end{aligned}$$

Correlation between observations taken on the same subject:

$$
\begin{aligned}
Cov(Y_{ijk}, Y_{ij\ell}) &= Cov(\mu + \alpha_i + S_{ij} + \tau_k + \gamma_{ik} + e_{ijk}, \\
&\qquad \mu + \alpha_i + S_{ij} + \tau_\ell + \gamma_{i\ell} + e_{ij\ell}) \\
&= Cov(S_{ij} + e_{ijk}, S_{ij} + e_{ij\ell}) \\
&= Cov(S_{ij}, S_{ij}) + Cov(S_{ij}, e_{ij\ell}) \\
&\quad + Cov(e_{ijk}, S_{ij}) + Cov(e_{ijk}, e_{ij\ell}) \\
&= Var(S_{ij}) \\
&= \sigma_S^2 \quad \text{for } k \neq \ell.
\end{aligned}
$$

The correlation between Y_{ijk} and $Y_{ij\ell}$ is

$$\frac{\sigma_S^2}{\sigma_S^2 + \sigma_e^2} \equiv \rho$$

Observations taken on different subjects are uncorrelated.

For the set of observations taken on a single subject, we have

$$
Var \begin{bmatrix} Y_{ij1} \\ Y_{ij2} \\ \vdots \\ Y_{ij7} \end{bmatrix} = \begin{bmatrix} \sigma_e^2 + \sigma_S^2 & \sigma_S^2 & \cdots & \sigma_S^2 \\ \sigma_S^2 & \sigma_e^2 + \sigma_S^2 & \cdots & \sigma_S^2 \\ \vdots & \vdots & \ddots & \sigma_S^2 \\ \sigma_S^2 & \sigma_S^2 & \sigma_S^2 & \sigma_e^2 + \sigma_S^2 \end{bmatrix}
$$
$$
= \sigma_e^2 I + \sigma_S^2 J
$$

This covariance structure is called compound symmetry.

Write this model in the form

$$\mathbf{Y} = X\beta + Z\mathbf{u} + \mathbf{e}$$

$$
\begin{bmatrix} Y_{111}\\ Y_{112}\\ \vdots \\ Y_{117}\\ Y_{121}\\ Y_{122}\\ \vdots\\ Y_{127}\\ \vdots\\ Y_{211}\\ Y_{212}\\ \vdots\\ Y_{217}\\ \vdots\\ Y_{3,n_3 1}\\ Y_{3,n_3 2}\\ \vdots\\ Y_{3,n_3 7}\end{bmatrix}
= \left[\begin{array}{cccc|c|l}
1&1&0&0&1000000& \\
1&1&0&0&0100000& \\
&\vdots&&&\ddots& 21\\
1&1&0&0&0000001& \text{columns}\\
1&1&0&0&1000000& \text{for}\\
1&1&0&0&0100000& \text{interaction}\\
&\vdots&&&\ddots& \\
1&1&0&0&0000001& \\
\vdots&&&&& \\
1&0&1&0&1000000& \\
1&0&1&0&0100000& \\
&&\vdots&&\ddots& \\
1&0&1&0&0000001& \\
\vdots&&&&& \\
1&0&0&1&1000000& \\
1&0&0&1&0100000& \\
\vdots&&\vdots&&\ddots& \\
1&0&0&1&0000001&
\end{array}\right]
\begin{bmatrix}\mu\\ \alpha_1\\ \alpha_2\\ \alpha_3\\ \tau_1\\ \tau_2\\ \tau_3\\ \tau_4\\ \tau_5\\ \tau_6\\ \tau_7\\ \gamma_{11}\\ \gamma_{12}\\ \vdots\\ \gamma_{37}\end{bmatrix}
$$

$$
+\begin{bmatrix}
1 & 0 & \cdots & 0 \\
1 & 0 & \cdots & 0 \\
\vdots & \vdots & & \vdots \\
1 & 0 & \cdots & 0 \\
0 & 1 & \cdots & 0 \\
0 & 1 & \cdots & 0 \\
\vdots & \vdots & & \vdots \\
0 & 1 & \cdots & 0 \\
\vdots & \vdots & & \vdots \\
\vdots & \vdots & & \vdots \\
\vdots & \vdots & & \vdots \\
0 & 0 & \cdots & 1 \\
0 & 0 & \cdots & 1 \\
\vdots & \vdots & & \vdots \\
0 & 0 & \cdots & 1
\end{bmatrix}
\begin{bmatrix}
S_{11} \\
S_{12} \\
\vdots \\
\vdots \\
\vdots \\
S_{3,n_3}
\end{bmatrix}
+
\begin{bmatrix}
e_{111} \\
e_{112} \\
\vdots \\
e_{117} \\
e_{121} \\
e_{122} \\
\vdots \\
e_{127} \\
\vdots \\
e_{211} \\
e_{212} \\
\vdots \\
e_{217} \\
\vdots \\
e_{3,n_3 1} \\
e_{3,n_3 2} \\
\vdots \\
e_{3,n_3 7}
\end{bmatrix}
$$

In this case:

$$R = Var(\mathbf{e}) \;=\; \sigma_e^2 I_{(7r)\times(7r)},$$

$$G = Var(\mathbf{u}) \;=\; \sigma_S^2 I_{r\times r},$$

where r is the number of subjects

$$\Sigma = Var(\mathbf{Y}) \;=\; ZGZ^T + R$$

is a block diagonal matrix with one block of the form

$$(\sigma_e^2 I_{7\times 7} + \sigma_S^2 J_{7\times 7})$$

for each subject

Analysis of Mixed Linear Models

$$\mathbf{Y} = X\boldsymbol{\beta} + Z\mathbf{u} + \mathbf{e}$$

where $X_{n\times p}$ and $Z_{n\times q}$ are known model matrices and

$$\begin{bmatrix} \mathbf{u} \\ \mathbf{e} \end{bmatrix} \sim N\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} G & 0 \\ 0 & R \end{bmatrix} \right)$$

Then

$$Y \sim N(X\boldsymbol{\beta}, \Sigma)$$

where

$$\Sigma = ZGZ^T + R$$

Some objectives:

(i) Inferences about estimable functions of fixed effects

- Point estimates
- Confidence intervals
- Tests of hypotheses

(ii) Estimation of variance components (elements of G and R)

(iii) Predictions of random effects (blup)

(iv) Predictions of future observations

Methods of Estimation

I. Ordinary Least Squares Estimation:

Normal equations (estimating equations):

$$(X^T X)\mathbf{b} = X^T \mathbf{Y}$$

and solutions have the form

$$\mathbf{b} = (X^T X)^- X^T \mathbf{Y}$$

The Gauss-Markov Theorem cannot be applied because it requires uncorrelated responses. In these models

$$Var(\mathbf{Y}) = ZGZ^T + R \neq \sigma^2 I$$

Hence, the OLS estimator of an estimable function $\mathbf{C}^T \boldsymbol{\beta}$ is not necessarily a best linear unbiased estimator (b.l.u.e.).

- The OLS estimator for $\mathbf{C}^T\beta$ is

$$\mathbf{C}^T\mathbf{b} = \mathbf{C}^T(X^TX)^{-}X^T\mathbf{Y}$$

  where

$$\mathbf{b} = (X^TX)^{-}X^T\mathbf{Y}$$

  is a solution to the normal equations.

- The OLS estimator $\mathbf{C}^T\mathbf{b}$ is a linear function of Y.
- $E(\mathbf{C}^T\mathbf{b}) = \mathbf{C}^T\beta$
- $Var(\mathbf{C}^T\mathbf{b}) = \mathbf{C}^T(X^TX)^{-}X^T(ZGZ^T + R)X(X^TX)^{-}\mathbf{C}$
- If $\mathbf{Y} \sim N(X\beta, ZGZ^T + R)$, then $\mathbf{C}^T\mathbf{b}$ has a normal distribution with mean $\mathbf{c}^T\beta$ and covariance matrix

$$\mathbf{C}^T(X^TX)^{-}X^T(ZGZ^T + R)X(X^TX)^{-}\mathbf{C}$$

II. Generalized Least Squares (GLS) Estimation:

Suppose

$$E(\mathbf{Y}) = X\beta$$

and also suppose

$$\Sigma = Var(\mathbf{Y}) = ZGZ^T + R$$

is known. Then a GLS estimator for β is any $\mathbf{b}$ that minimizes

$$Q(\mathbf{b}) = (\mathbf{Y} - X\mathbf{b})^T\Sigma^{-1}(\mathbf{Y} - X\mathbf{b})$$

The estimating equations are:

$$(X^T\Sigma^{-1}X)\mathbf{b} = X^T\Sigma^{-1}\mathbf{Y}$$

and

$$\mathbf{b}_{GLS} = (X^T\Sigma^{-1}X)^{-}(X^T\Sigma^{-1}\mathbf{Y})$$

is a solution.

For any estimable function $C^T\boldsymbol{\beta}$, the unique b.l.u.e. is

$$C^T\mathbf{b}_{GLS} = C^T(X^T\Sigma^{-1}X)^{-}X^T\Sigma^{-1}\mathbf{Y}$$

with $Var(C^T\mathbf{b}_{GLS}) = C^T(X^T\Sigma^{-1}X)^{-}C$. If $Y \sim N(X\boldsymbol{\beta}, \Sigma)$, then

$$C^T\mathbf{b}_{GLS} \sim N\left(C^T\boldsymbol{\beta},\ C^T(X^T\Sigma^{-1}X)^{-}C\right).$$

When G and/or R contain unknown parameters, you could obtain an *approximate BLUE* by replacing the unknown parameters with consistent estimators to obtain

$$\hat{\Sigma} = Z\hat{G}Z^T + \hat{R}$$

and

$$C^T\mathbf{b}^*_{GLS} = C^T(X^T\hat{\Sigma}^{-1}X)^{-}\hat{\Sigma}^{-1}\mathbf{Y}$$

- $C^T \mathbf{b}^*_{GLS}$ is not a linear function of $\mathbf{Y}$

- $C^T \mathbf{b}^*_{GLS}$ is not a best linear unbiased estimator (BLUE)

- See Kackar and Harville (1981, 1984) for conditions under which $C^T \mathbf{b}^*_{GLS}$ is an unbiased estimator for $C^T \boldsymbol{\beta}$

- $C^T (X^T \hat{\Sigma}^{-1} X)^- C$ tends to *underestimate* $Var(C^T \mathbf{b}^*_{GLS})$ (see Eaton (1984))

- For *large* samples

$$C^T \mathbf{b}^*_{GLS} \dot{\sim} N(C^T \boldsymbol{\beta}, C^T (X^T \Sigma^{-1} X)^- C)$$

Variance component estimation

- Estimation of parameters in G and R
- Crucial to the estimation of estimable functions of fixed effects (e.g. $E(\mathbf{Y}) = X\beta$)
- Of interest in its own right (sources of variation in the pigment paste production example)

Basic Approaches

(i) ANOVA methods (method of moments): Set observed values of mean squares equal to their expectations and solve the resulting equations.

(ii) Maximum likelihood estimation (ML)

(iii) Restricted maximum likelihood estimation (REML)

I. ANOVA method (Method of Moments)

- Compute an ANOVA table
- Equate mean squares to their expected values
- Solve the resulting equations
- will be discussed later in the examples

Likelihood-based methods:

Consider the mixed model

$$\mathbf{Y}_{n\times 1} = X\boldsymbol{\beta}_{p\times 1} + Z\mathbf{u}_{q\times 1} + \mathbf{e}_{n\times 1}$$

where

$$\begin{bmatrix} \mathbf{u} \\ \mathbf{e} \end{bmatrix} \sim N\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} G & 0 \\ 0 & R \end{bmatrix} \right)$$

Then,

$$\mathbf{Y}_{n\times 1} \sim N(X\boldsymbol{\beta}, \Sigma)$$

where $\Sigma = ZGZ^T + R$

- Maximum Likelihood Estimation
- Restricted Maximum Likelihood Estimation (REML)

Maximum Likelihood Estimation

Multivariate normal likelihood:

$$L(\boldsymbol{\beta}, \Sigma; \mathbf{Y}) = (2\pi)^{-n/2}|\Sigma|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{Y} - X\boldsymbol{\beta})^T \Sigma^{-1}(\mathbf{Y} - X\boldsymbol{\beta})\right\}$$

The log-likelihood function is

$$\begin{aligned} \ell(\boldsymbol{\beta}, \Sigma; \mathbf{Y}) &= -\frac{n}{2} log(2\pi) - \frac{1}{2} log(|\Sigma|) \\ &\quad -\frac{1}{2}(\mathbf{Y} - X\boldsymbol{\beta})^T \Sigma^{-1}(\mathbf{Y} - X\boldsymbol{\beta}) \end{aligned}$$

Given the values of the observed responses, $\mathbf{Y}$, find values $\boldsymbol{\beta}$ and Σ that maximize the log-likelihood function.

This is a difficult computational problem:

- no analytic solution (except in some balanced cases)
- use iterative numerical methods
  - Need starting values (initial guesses at the values of $\hat{\beta}$ and $\hat{\Sigma} = Z\hat{G}Z^T + \hat{R}$).
  - local or global maxima?
  - what if $\hat{\Sigma}$ becomes singular or is not positive definite?

- Constrained optimization
  - estimates of variances cannot be negative
  - estimated correlations between -1 and 1
  - $\hat{\Sigma}$, $\hat{G}$, and $\hat{R}$ are positive definite (or non-negative definite)

- Large sample distributional properties of estimators
  - consistency
  - normality
  - efficiency*

*not guaranteed for ANOVA methods

- Estimates of variance components tend to be too small

  Consider a sample $Y_1, \ldots, Y_n$ from a $N(\mu, \sigma^2)$ distribution. An unbiased estimator for σ^2 is

$$S^2 = \frac{1}{n-1} \sum_{j=1}^{n} (Y_j - \bar{Y})^2$$

  The MLE for σ^2 is

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{j=1}^{n} (Y_j - \bar{Y})^2$$

  with

$$E(\hat{\sigma}^2) = \left(\frac{n-1}{n} \right) \sigma^2 < \sigma^2$$

Note that S^2 and $\hat{\sigma}^2$ are based on *error contrasts*

$$\begin{aligned} e_1 = Y_1 - \bar{Y} &= \left(\tfrac{n-1}{n}, -\tfrac{1}{n}, \ldots, -\tfrac{1}{n}\right) \mathbf{Y} \\ \vdots \quad & \\ e_n = Y_n - \bar{Y} &= \left(-\tfrac{1}{n}, -\tfrac{1}{n}, \ldots, -\tfrac{1}{n}, \tfrac{n-1}{n}\right) \mathbf{Y} \end{aligned}$$

whose distribution does not depend on

$$\mu = E(Y_j)\ .$$

When $\mathbf{Y} \sim N(\mu\mathbf{1}, \sigma^2 I)$,

$$\mathbf{e} = \begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix} = (I - P_{\mathbf{1}})\mathbf{Y} \sim N\left[\mathbf{0}, \sigma^2(I - P_{\mathbf{1}})\right]$$

- The MLE $\hat{\sigma}^2 = \frac{1}{n}\sum_{j=1}^{n} e_j^2$ fails to acknowledge that $\mathbf{e}$ is restricted to an $(n-1)$-dimensional space, i.e., $\sum_{j=1}^{n} e_j = 0$.
- The MLE fails to make the appropriate adjustment in *degrees of freedom* needed to obtain an unbiased estimator for σ^2.

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

Then

$$\mathbf{e} = \begin{bmatrix} Y_1 - \bar{Y} \\ Y_2 - \bar{Y} \\ Y_3 - \bar{Y} \\ Y_4 - \bar{Y} \end{bmatrix} = (I - P_{\mathbf{1}})\mathbf{Y} \sim N\left[0, \underline{\sigma^2(I - P_{\mathbf{1}})}\right]$$

$\uparrow$

This covariance matrix is singular.

Here, $m = rank(I - P_{\mathbf{1}}) = n - 1 = 3$.

Define

$$\mathbf{r} = M\mathbf{e} = M(I - P_X)\mathbf{Y}$$

where

$$M = \begin{bmatrix} 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}$$

has row rank equal to

$$m = rank(I - P_X).$$

Then

$$\begin{aligned} \mathbf{r} &= \begin{bmatrix} r_1 \\ r_2 \\ r_3 \end{bmatrix} = \begin{bmatrix} Y_1 + Y_2 - Y_3 - Y_4 \\ Y_1 - Y_2 + Y_3 - Y_4 \\ Y_1 - Y_2 - Y_3 + Y_4 \end{bmatrix} \\ &= M(I - P_{\mathbf{1}})\mathbf{Y} \sim N(\mathbf{0}, \underline{\sigma^2 M(I - P_{\mathbf{1}})M^T}) \end{aligned}$$

$\uparrow$

call this $\sigma^2 W$

Restricted Likelihood function:

$$L(\sigma^2; \mathbf{r}) = \frac{1}{(2\pi)^{M/2}|\sigma^2 W|^{1/2}} \ e^{-\frac{1}{2\sigma^2}\mathbf{r}^T W^{-1}\mathbf{r}}$$

Restricted Log-likelihood:

$$\begin{aligned}\ell(\sigma^2; \mathbf{r}) &= \frac{-m}{2} log(2\pi) - \frac{m}{2} log(\sigma^2) \\ &\quad -\frac{1}{2} log|W| - \frac{1}{2\sigma^2}\mathbf{r}^T W^{-1}\mathbf{r}\end{aligned}$$

(Note that $|\sigma^2 W| = (\sigma^2)^m |W|$)

(Restricted) likelihood equation:

$$0 = \frac{\partial \ell(\sigma^2; \mathbf{r})}{\partial \sigma^2} = \frac{-m}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\mathbf{r}^T W^{-1}\mathbf{r}$$

Solution (REML estimator for σ^2):

$$\begin{aligned}\hat{\sigma}^2_{REML} &= \frac{1}{m}\mathbf{r}^T W^{-1}\mathbf{r} \\ &= \frac{1}{m}\mathbf{Y}^T \underline{(I - P_{\mathbf{1}})^T M^T (M(I - P_{\mathbf{1}})M^T)^{-1} M(I - P_{\mathbf{1}})\mathbf{Y}}\end{aligned}$$

This is a projection of $\mathbf{Y}$ onto the column space of $M(I - P_{\mathbf{1}})$ which is the column space of $I - P_{\mathbf{1}}$

$$\begin{aligned} &= \frac{1}{m}\mathbf{Y}^T (I - P_{\mathbf{1}})\mathbf{Y} \\ &= \frac{1}{n-1}\sum_{j=1}^{n}(Y_j - \bar{Y})^2 = S^2\end{aligned}$$

REML (Restricted Maximum Likelihood) estimation

- Estimate parameters in

$$\Sigma = ZGZ^T + R$$

  by maximizing the part of the likelihood that does not depend on $E(\mathbf{Y}) = X\beta$

- Maximize a likelihood function for *error contrasts*
  - linear combinations of observations that do not depend on $X\beta$
  - Find a set of

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

    linearly independent *error contrasts*

Mixed (normal-theory) model:

$$\mathbf{Y} = X\boldsymbol{\beta} + Z\mathbf{u} + \mathbf{e}$$

where $\begin{bmatrix} \mathbf{u} \\ \mathbf{e} \end{bmatrix} \sim N\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} G & 0 \\ 0 & R \end{bmatrix} \right)$

Then

$$L\mathbf{Y} = L(X\boldsymbol{\beta} + Z\mathbf{u} + \mathbf{e}) = LX\boldsymbol{\beta} + LZ\mathbf{u} + L\mathbf{e}$$

is invariant to $X\boldsymbol{\beta}$ if and only if $LX = 0$. But $LX = 0$ if and only if

$$L = M(I - P_X)$$

for some M. (Here $P_X = X(X^TX)^{-}X^T$)

To avoid losing information we must have

$$\begin{aligned}\text{row rank}(M) &= n - \text{rank}(X) \\ &= n - p\end{aligned}$$

Then a set of $n - p$ error contrasts is

$$\begin{aligned}\mathbf{r} &= M(I - P_X)\mathbf{Y} \\ &\sim N_{n-p}(\mathbf{0}, \underline{M(I - P_X)\Sigma^{-1}(I - P_X)M^T})\end{aligned}$$

call this W,
then $\text{rank}(W) = n - p$
and W^{-1} exists.

The *Restricted* likelihood is

$$L(\Sigma; \mathbf{r}) = \frac{1}{(2\pi)^{(n-p)/2}|W|^{1/2}} e^{-\frac{1}{2}\mathbf{r}^T W^{-1}\mathbf{r}}$$

The resulting log-likelihood is

$$\begin{aligned} \ell(\Sigma; \mathbf{r}) &= \frac{-(n-p)}{2} log(2\pi) - \frac{1}{2} log|W| \\ & -\frac{1}{2}\mathbf{r}^T W^{-1}\mathbf{r} \end{aligned}$$

For any $M_{(n-p)\times n}$ with row rank equal to

$$n - p = n - rank(X)$$

the log-likelihood can be expressed in terms of

$$\mathbf{e} = (I - X(X\Sigma^{-1}X^T)^{-}X^T\Sigma^{-1})\mathbf{Y}$$

as

$$\begin{aligned}\ell(\Sigma; \mathbf{e}) &= \text{constant} - \frac{1}{2}log(|\Sigma|) \\ &\quad -\frac{1}{2}log(|X_*^T\Sigma^{-1}X_*|) - \frac{1}{2}\mathbf{e}^T\Sigma^{-1}\mathbf{e}\end{aligned}$$

where X_* is any set of p =rank(X) linearly independent columns of X. Denote the resulting REML estimators as

$$\hat{G}, \quad \hat{R} \quad \text{and} \quad \hat{\Sigma} = Z\hat{G}Z^T + \hat{R}$$

Estimation of fixed effects

For any estimable function $C\beta$, the **blue** is the generalized least squares estimator

$$C\mathbf{b}_{GLS} = C(X^T\Sigma^{-1}X)^{-}X^T\Sigma^{-1}\mathbf{Y}$$

Using the REML estimator for

$$\Sigma = ZGZ^T + R$$

an approximation is

$$C\hat{\beta} = C(X^T\hat{\Sigma}^{-1}X)^{-}X^T\hat{\Sigma}^{-1}\mathbf{Y}$$

and for *large* samples:

$$C\hat{\beta} \dot{\sim} N(C\beta, C(X^T\Sigma^{-1}X)^{-}C^T)$$

Prediction of random effects

Given the observed responses $\mathbf{Y}$, predict the value of $\mathbf{u}$.

For our model,

$$\begin{bmatrix} \mathbf{u} \\ \mathbf{e} \end{bmatrix} \sim N\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} G & 0 \\ 0 & R \end{bmatrix} \right).$$

Then (from result 4.1)

$$\begin{aligned} \begin{bmatrix} \mathbf{u} \\ \mathbf{Y} \end{bmatrix} &= \begin{bmatrix} \mathbf{u} \\ X\boldsymbol{\beta} + Z\mathbf{u} + \mathbf{e} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{0} \\ X\boldsymbol{\beta} \end{bmatrix} + \begin{bmatrix} I & 0 \\ Z & I \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \mathbf{e} \end{bmatrix} \\ &\sim N\left(\begin{bmatrix} \mathbf{0} \\ X\boldsymbol{\beta} \end{bmatrix}, \begin{bmatrix} G & GZ^T \\ ZG & ZGZ^T + R \end{bmatrix} \right) \end{aligned}$$

The Best Linear Unbiased Predictor (BLUP) is the b.l.u.e. for

$$\begin{aligned} E(\mathbf{u}|\mathbf{Y}) &= E(\mathbf{u}) + (GZ^T)(ZGZ^T + R)^{-1}(Y - E(\mathbf{Y})) \\ &= \mathbf{0} + GZ^T(ZGZ^T + R)^{-1}(\mathbf{Y} - X\boldsymbol{\beta}) \end{aligned}$$

$\uparrow$

substitute the b.l.u.e. for $X\boldsymbol{\beta}$

$$X\mathbf{b}_{GLS} = X(X^T\Sigma^{-1}X)^{-}X^T\Sigma^{-1}\mathbf{Y}$$

Then, the BLUP for $\mathbf{u}$ is

$$\begin{aligned} BLUP(\mathbf{u}) =& GZ^T\Sigma^{-1}(\mathbf{Y} - X\mathbf{b}_{GLS}) \\ =& GZ^T\Sigma^{-1}(I - X(X^T\Sigma^{-1}X)^{-}X^T\Sigma^{-1})\mathbf{Y}, \end{aligned}$$

when G and $\Sigma = ZGZ^T + R$ are known.

Substituting REML estimators $\hat{G}$ and $\hat{R}$ for G and R, an approximate BLUP for $\mathbf{u}$ is

$$\begin{aligned}\hat{\mathbf{u}} &= \hat{G}Z^T\hat{\Sigma}^{-1}(I - X(X^T\hat{\Sigma}^{-1}X)^{-}X^T\hat{\Sigma}^{-1})\mathbf{Y} \\ &= \hat{G}Z^T\hat{\Sigma}^{-1}(\mathbf{Y} - \underline{X\hat{\beta}})\end{aligned}$$

For *large* samples, the distribution of $\hat{\mathbf{u}}$ is approximately multivariate normal with mean vector $\mathbf{0}$ and covariance matrix

$$GZ^T\Sigma^{-1}(I - P)\Sigma(I - P)\Sigma^{-1}ZG$$

where

$$P = X(X^T\Sigma^{-1}X)^{-}X^T\Sigma^{-1}$$

Given estimates $\hat{G}, \hat{R}$ and $\hat{\Sigma} = Z\hat{G}Z^T + \hat{R}, \hat{\beta}$ and $\hat{\mathbf{u}}$ provide a solution to the mixed model equations:

$$\begin{bmatrix} X^T\hat{R}^{-1}X & X^T\hat{R}^{-1}Z \\ Z^T\hat{R}^{-1} & Z^T\hat{R}^{-1}Z + \hat{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} X^T\hat{R}^{-1}\mathbf{Y} \\ Z^T\hat{R}^{-1}\mathbf{Y} \end{bmatrix}$$

A generalized inverse of

$$\begin{bmatrix} X^T\hat{R}^{-1}X & X^T\hat{R}^{-1}Z \\ Z^T\hat{R}^{-1} & Z^T\hat{R}^{-1}Z + \hat{G}^{-1} \end{bmatrix}$$

is used to approximate the covariance matrix for $\begin{bmatrix} \hat{\beta} \\ \hat{\mathbf{u}} \end{bmatrix}$