Mixed Model Analysis

Basic model:

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

where

- X is a $n \times p$ model matrix of known constants
- β is a $p \times 1$ vector of "fixed" unknown parameter values
- Z is a $n \times q$ model matrix of known constants
- u is a $q \times 1$ random vector
- e is a $n \times 1$ vector of random errors

with

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

Then

$$\begin{aligned} E(\mathrm{Y}) &= E(X\beta + Z\mathrm{u} + \mathrm{e}) \\ &= X\beta + ZE(\mathrm{u}) + E(\mathrm{e}) \\ &= X\beta \\ Var(\mathrm{Y}) &= Var(X\beta + Z\mathrm{u} + \mathrm{e}) \\ &= Var(Z\mathrm{u}) + Var(\mathrm{e}) \\ &= ZGZ^T + R \\ &\equiv \Sigma \end{aligned}$$

Normal-theory mixed model

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

Then,

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

$$\uparrow$$

call this Σ

Example 10.1: Random Blocks

Comparison of four processes for producing penicillin

$$\left.\begin{matrix} Process\ A \\ Process\ B \\ Process\ C \\ Process\ D \end{matrix}\right\}$$ Levels of a "fixed" treatment factor

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

S-PLUS code: penclln.ssc

Model:

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

$$\beta_j \sim NID(0, \sigma_\beta^2)$$
$$e_{ij} \sim NID(0, \sigma_e^2)$$

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 S-PLUS 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

- $\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 = 0$
- $\mu = (\mu_1 + \mu_2 + \mu_3 + \mu_4)/4$
- $\alpha_i = \mu_i - \mu$

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:

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

Correlation among yields for runs on the same batch:

$$\begin{aligned} \rho &= \frac{Cov(Y_{ij}, Y_{kj})}{\sqrt{Var(Y_{ij}) Var(Y_{kj})}} \\ &= \frac{\sigma_\beta^2}{\sigma_\beta^2 + \sigma_e^2} \text{ for } i \neq k \end{aligned}$$

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:

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

$$= \sigma_\beta^2 \underset{\substack{\uparrow \\ \text{matrix} \\ \text{of} \\ \text{ones}}}{J} + \sigma_e^2 \underset{\substack{\uparrow \\ \text{identity} \\ \text{matrix}}}{I}$$

This special type of covariance structure is called

compound symmetry

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

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

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

Here

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

and

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

$$\begin{bmatrix} \sigma_\beta^2 J + \sigma_e^2 I & & & \\ & \sigma_\beta^2 J + \sigma_e^2 I & & \\ & & \ddots & \\ & & & \sigma_\beta^2 J + \sigma_e^2 I \end{bmatrix}$$

Example 10.2: 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

- avearge 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

<u>Data file</u>: pigment.dat

<u>SAS code</u>: pigment.sas

<u>S-PLUS code</u>: pigment.ssc

Write this model in the form:

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

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

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

where

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

Then

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

because $Z = [I_b \otimes 1_{sr} | I_{bs} \otimes 1_r]$

$$\underset{\substack{(sr) \times 1 \\ \text{vector} \\ \text{of ones}}}{\nearrow} \qquad \underset{\substack{r \times 1 \\ \text{vector} \\ \text{of ones}}}{\nearrow}$$

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

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

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):

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

and solutions have the form

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

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

$$\begin{aligned} Var(\mathrm{Y}) &= ZGZ^T + R \\ &\neq \sigma^2 I \end{aligned}$$

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

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

  $$\mathrm{C}^T\mathrm{b} = \mathrm{C}^T(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T\mathrm{Y}$$

  where

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

  is a solution to the normal equations.

- The OLS estimator $\mathrm{C}^T\mathrm{b}$ is a linear function of **Y**.
- $E(\mathrm{C}^T\mathrm{b}) = \mathrm{C}^T\beta$

- $Var(\mathrm{C}^T\mathrm{b}) =$
  $$\mathrm{C}^T(X^TX)^-X^T(ZGZ^T+R)$$
  $$X(X^TX)^-\mathrm{C}$$
- If $\mathrm{Y} \sim N(X\beta,\, ZGZ^T+R)$, then $\mathrm{C}^T\mathrm{b}$ has a normal distribution with mean
  $$C^T\beta$$
  and covariance matrix
  $$\mathrm{C}^T(X^TX)^-X^T(ZGZ^T+R)$$
  $$X(X^TX)^-\mathrm{C}$$

II. Generalized Least Squares (GLS) Estimation:

Suppose

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

and also suppose

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

is known. Then a GLS estimator for β is any b that minimizes

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

The estimating equations are:

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

and

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

is a solution.

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

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

with
$$Var(C^T\mathrm{b}_{GLS}) = C^T(X^T\Sigma^{-1}X)^-C$$

If $Y \sim N(X\beta,\, \Sigma)$, then

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

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

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

- $C^T\mathrm{b}^*_{GLS}$ is not a linear function of Y
- $C^T\mathrm{b}^*_{GLS}$ is not a best linear unbiased estimator (BLUE)
- See Kackar and Harville (1981, 1984) for conditions under which $C^T\mathrm{b}^*_{GLS}$ is an unbiased estimator for $C^T\beta$
- $C^T(X^T\hat{\Sigma}^{-1}X)^-C$ tends to "underestimate" $Var(C^T\mathrm{b}^*_{GLS})$ (see Eaton (1984))
- For "large" samples

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

Variance component estimation

- Estimation of parameters in G and R
- Crucial to the estimation of estimable functions of fixed effects (e.g. $E(\mathrm{Y}) = X\beta$)
- 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)

ANOVA method
(Method of Moments)

- Compute an ANOVA table
- Equate mean squares to their expected values
- Solve the resulting equations

Example 10.1 Penicillin production

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

where

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

and

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

Source of Variation	d.f.	Sums of Squares
Blocks	4	$a \sum_{j=1}^{b} (\bar{Y}_{.j} - \bar{Y}_{..})^2 = SS_{blocks}$
Processes	3	$b \sum_{i=1}^{a} (\bar{Y}_{1.} - \bar{Y}_{..})^2 = SS_{processes}$
error	12	$\sum_{i=1}^{a} \sum_{j=1}^{b} (Y_{ij} - \bar{Y}_{1.} - \bar{Y}_{.j} + \bar{Y}_{..})^2 = SSE$
C. total	19	$\sum_{i=1}^{a} \sum_{j=1}^{b} (Y_{ij} - \bar{Y}_{..})^2$

Start at the bottom:

$$\begin{aligned} MS_{error} &= \frac{SSE}{(a-1)(b-1)} \\ E(MS_{error}) &= \sigma^2_e \end{aligned}$$

Then an unbiased estimator for σ_e is

$$\hat{\sigma}^2_e = MS_{error}$$

Next, consider the mean square for the random block effects:

$$\begin{aligned} MS_{blocks} &= \frac{SS_{blocks}}{b-1} \\ E(MS_{blocks}) &= \sigma^2_e + \underset{\substack{\uparrow \\ \text{number of} \\ \text{observations} \\ \text{for each block}}}{a}\sigma^2_\beta \end{aligned}$$

Then,

$$\begin{aligned} \sigma^2_\beta &= \frac{E(MS_{blocks}) - \sigma^2_e}{a} \\ &= \frac{E(MS_{blocks}) - E(MS_{error})}{a} \end{aligned}$$

An unbiased estimator for σ_β^2 is

$$\hat{\sigma}_\beta^2 = \frac{MS_{blocks} - MS_{error}}{a}$$

For the penicillin data

$$\begin{aligned}
\hat{\sigma}_e^2 &= MS_{error} = 18.83 \\
\hat{\sigma}_\beta &= \frac{MS_{blocks} - MS_{error}}{4} \\
&= \frac{66.0 - 18.83}{4} = 11.79 \\
\widehat{Var(Y_{ij})} &= \hat{\sigma}_\beta^2 + \hat{\sigma}_e^2 \\
&= 11.79 + 18.83 = 30.62
\end{aligned}$$

```
/*  This is a program for analyzing the
    penicillan data from Box, Hunter, and
    Hunter.  It is posted in the file

             penclln.sas

    First enter the data  */

data set1;
  infile 'penclln.dat';
  input batch process $ yield;
  run;

/* Compute the ANOVA table, formulas for
   expectations of mean squares, process
   means and their standard errors */

proc glm data=set1;
  class batch process;
  model yield = batch process / e e3;
  random batch / q test;
  lsmeans process / stderr pdiff tdiff;
  output out=set2 r=resid p=yhat;
  run;
```

```
/* Compute a normal probability plot for
   the residuals and the Shapiro-Wilk test
   for normality  */

proc rank data=set2 normal=blom out=set2;
  var resid;  ranks q;
  run;

proc univariate data=set2 normal plot;
  var resid;
  run;

goptions cback=white colors=(black)
  target=win device=winprtc rotate=portrait;

  axis1 label=(h=2.5 r=0 a=90 f=swiss 'Residuals')
        value=(f=swiss h=2.0) w=3.0 length=5.0 in;

  axis2 label=(h=2.5 f=swiss 'Standard
                               Normal Quantiles')
        value=(f=swiss h=2.0) w=3.0 length=5.0 in;
```

```
  axis3 label=(h=2.5 f=swiss 'Production Process')
        value=(f=swiss h=2.0) w=3.0 length=5.0 in;

  symbol1 v=circle i=none h=2 w=3 c=black;

proc gplot data=set2;
  plot resid*q / vaxis=axis1 haxis=axis2;
  title h=3.5 f=swiss c=black
        'Normal Probability Plot';
  run;

proc gplot data=set2;
  plot resid*process / vaxis=axis1 haxis=axis3;
  title h=3.5 f=swiss c=black 'Residual Plot';
  run;
```

```
/* Fit the same model using PROC MIXED.  Compute
   REML estimates of variance components.  Note
   that PROC MIXED provides appropriate standard
   errors for process means.  When block effects
   are random. PROC GLM does not provide correct
   standard errors for process means  */



proc mixed data=set1;
  class process batch;
  model yield = process / ddfm=satterth solution;
  random batch / type=vc g solution cl alpha=.05;
  lsmeans process / pdiff tdiff;
run;
```

```
          General Linear Models Procedure
              Class Level Information

             Class    Levels    Values

             BATCH         5    1 2 3 4 5

             PROCESS       4    A B C D


      Number of observations in data set = 20



    General Form of Estimable Functions

    Effect        Coefficients

 INTERCEPT        L1

 BATCH    1       L2
          2       L3
          3       L4
          4       L5
          5       L1-L2-L3-L4-L5

 PROCESS  A       L7
          B       L8
          C       L9
          D       L1-L7-L8-L9
```

```
 Type III Estimable Functions for: BATCH

   Effect       Coefficients
INTERCEPT       0

BATCH    1     L2
         2     L3
         3     L4
         4     L5
         5     -L2-L3-L4-L5

PROCESS   A     0
          B     0
          C     0
          D     0


Type III Estimable Functions for: PROCESS

  Effect        Coefficients
INTERCEPT       0

BATCH    1     0
         2     0
         3     0
         4     0
         5     0

PROCESS  A     L7
         B     L8
         C     L9
         D     -L7-L8-L9
```

```
Dependent Variable: YIELD

                     Sum of       Mean
Source       DF     Squares     Square   F Value  Pr > F

Model         7      334.00      47.71     2.53    0.0754

Error        12      226.00      18.83

Cor. Total   19      560.00



     R-Square      C.V.      Root MSE      YIELD Mean

       0.596      5.046       4.3397           86.0



                   Type III    Mean
Source       DF       SS      Square    F Value   Pr > F

BATCH         4     264.0      66.0      3.50     0.0407
PROCESS       3      70.0      23.3      1.24     0.3387
```

```
        Quadratic Forms of Fixed Effects in the
                 Expected Mean Squares

Source: Type III Mean Square for PROCESS

                 PROCESS    PROCESS    PROCESS    PROCESS
                    A          B          C          D
PROCESS A         3.750     -1.250     -1.250     -1.250
PROCESS B        -1.250      3.750     -1.250     -1.250
PROCESS C        -1.250     -1.250      3.750     -1.250
PROCESS D        -1.250     -1.250     -1.250      3.750



  Source              Type III Expected Mean Square

BATCH                   Var(Error) + 4 Var(BATCH)

PROCESS                 Var(Error) + Q(PROCESS)
```

```
          Tests of Hypotheses for Mixed Model
                 Analysis of Variance

Dependent Variable: YIELD

Source: BATCH
Error: MS(Error)
                        Denominator
   DF    Type III MS     DF     MS      F Value    Pr > F
    4        66          12   18.83      3.5044    0.0407


Source: PROCESS
Error: MS(Error)
                        Denominator
   DF    Type III MS     DF     MS       F Value     Pr > F
    3       23.33        12   18.83      1.2389     0.3387


                      Least Squares Means

  YIELD    Std     Pr > |T|
 LSMEAN   Error                  t-tests / p-values

  A  84   1.941   0.0001   1      .     -0.364  -1.822  -0.729
                                         0.722   0.093   0.480
  B  85   1.941   0.0001   2    0.364    .      -1.457  -0.364
                                0.722            0.171   0.722
  C  89   1.941   0.0001   3    1.822  1.457     .       1.093
                                0.094  0.171             0.296
  D  86   1.941   0.0001   4    0.729  0.364  -1.093     .
                                0.480  0.723   0.296
```

```
           The MIXED Procedure

            Model Information

 Data Set                       WORK.SET1
 Dependent Variable             yield
 Covariance Structure           Variance Components
 Estimation Method              REML
 Residual Variance Method       Profile
 Fixed Effects SE Method        Model-Based
 Degrees of Freedom Method      Satterthwaite



            Class Level Information

            Class     Levels  Values

           PROCESS       4    A B C D
           BATCH         5    1 2 3 4 5
```

```
                  Iteration History

 Iteration    Eval    -2 Res Log Like      Criterion

         0       1       106.59285141
         1       1       103.82994387      0.00000000

             Convergence criteria met.


                Estimated G Matrix

Row  Effect  batch   Col1    Col2    Col3    Col4    Col5

  1  batch   1    11.7917
  2  batch   2           11.7917
  3  batch   3                  11.7917
  4  batch   4                         11.7917
  5  batch   5                                11.7917


            Covariance Parameter Estimates
               Cov Parm     Estimate

               batch         11.7917
               Residual      18.8333
```

```
                 Fit Statistics

    Res Log Likelihood                      -51.9
    Akaike's Information Criterion          -53.9
    Schwarz's Bayesian Criterion            -53.5
    -2 Res Log Likelihood                   103.8



               Solution for Fixed Effects

                           Standard
Effect   process  Estimate    Error    DF  t Value  Pr > |t|

Intercept         86.0000    2.4749  11.1    34.75    <.0001
process    A      -2.0000    2.7447    12    -0.73    0.4802
process    B      -1.0000    2.7447    12    -0.36    0.7219
process    C       3.0000    2.7447    12     1.09    0.2958
process    D            0         .     .        .         .
```

```
           Solution for Random Effects

                      Std Err
Effect batch  Estimate    Pred    DF       t   Pr > |t|

batch    1     4.2879   2.2473  5.29    1.91   0.1115
batch    2    -2.1439   2.2473  5.29   -0.95   0.3816
batch    3    -0.7146   2.2473  5.29   -0.32   0.7627
batch    4     1.4293   2.2473  5.29    0.64   0.5513
batch    5    -2.8586   2.2473  5.29   -1.27   0.2564



           Solution for Random Effects

    Effect    batch     Alpha        Lower       Upper

    batch     1          0.05      -1.3954      9.9712
    batch     2          0.05      -7.8273      3.5394
    batch     3          0.05      -6.3980      4.9687
    batch     4          0.05      -4.2540      7.1126
    batch     5          0.05      -8.5419      2.8247
```

```
              Type 3 Tests of Fixed Effects

                      Num     Den
         Effect        DF      DF    F Value     Pr > F

         process        3      12       1.24     0.3387

                 Least Squares Means

                          Standard
 Effect process   Est.      Error     DF  t Value  Pr > |t|

 process  A    84.0000    2.4749  11.1    33.94    <.0001
 process  B    85.0000    2.4749  11.1    34.35    <.0001
 process  C    89.0000    2.4749  11.1    35.96    <.0001
 process  D    86.0000    2.4749  11.1    34.75    <.0001

          Differences of Least Squares Means

                             Standard
 Effect   process  Estimate    Error   DF      t     Pr > |t|

 process   A  B   -1.0000    2.7447   12   -0.36    0.7219
 process   A  C   -5.0000    2.7447   12   -1.82    0.0935
 process   A  D   -2.0000    2.7447   12   -0.73    0.4802
 process   B  C   -4.0000    2.7447   12   -1.46    0.1707
 process   B  D   -1.0000    2.7447   12   -0.36    0.7219
 process   C  D    3.0000    2.7447   12    1.09    0.2958
```

Inferences about treatment means:

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

Consider the sample mean (one observation for each treatment in each block):

$$\bar{Y}_{i.} = \frac{1}{b} \sum_{j=1}^{b} Y_{ij}$$

$$E(\bar{Y}_{i.}) = \begin{cases} \mu + \alpha_i & \text{for random block effects } \beta_j \sim NID(0, \sigma^2_\beta) \\ \mu + \alpha_i + \frac{1}{b}\sum_{j=1}^{b} \beta_j & \text{for fixed block effects} \end{cases}$$

and

$$
\begin{aligned}
Var(\bar{Y}_{i.}) &= Var(\frac{1}{b}\sum_{j=1}^{b} Y_{ij}) \\
&= \frac{1}{b^2}\sum_{j=1}^{b} Var(Y_{ij}) \\
&= \begin{cases} \frac{1}{b}(\sigma_e^2+\sigma_b^2) & \text{random block effects} \\ \frac{1}{b}(\sigma_e^2) & \text{fixed block effects} \end{cases}
\end{aligned}
$$

Confidence Intervals

Fixed additive block effects:

$$S^2_{\bar{Y}_{i.}} = \frac{1}{b}\hat{\sigma}_e^2 = \frac{1}{b}MS_{error}$$

The standard error for $\bar{Y}_{i.}$ is

$$S_{\bar{Y}_{i.}} = \sqrt{\frac{1}{b}MS_{error}} = 1.941$$

A $(1-\alpha)\times 100\%$ confidence interval for

$$\mu + \alpha_i + \frac{1}{b}\sum_{j=1}^{b}\beta_j$$

is

$$\bar{Y}_{i.} \pm t_{(a-1)(b-1),\frac{\alpha}{2}}\sqrt{\frac{1}{b}MS_{error}}$$

t-tests:

Reject $H_0 : \mu + \alpha_i + \frac{1}{b}\Sigma_{j=1}^{b}\beta_j = d$ if

$$|t| = \frac{|\bar{Y}_{i.} - d|}{\sqrt{\frac{1}{b}MS_{error}}} > t_{(a-1)(b-1),\frac{\alpha}{2}}$$

- This is what is done by the LSMEANS option in the GLM procedure in SAS, even when you specify

  RANDOM BATCH;
- This is what is done by the MIXED procedure in SAS when batch effects are not random

Models with random additive block effects:

$$
\begin{aligned}
S^2_{\bar{Y}_{i.}} &= \frac{1}{b}(\hat{\sigma}_e^2 + \hat{\sigma}_\beta^2) \\
&= \frac{1}{b}\left(MS_{error} + \frac{MS_{blocks} - MS_{error}}{a}\right) \\
&= \frac{a-1}{ab}MS_{error} + \frac{1}{ab}MS_{blocks} \\
&= \frac{1}{ab(b-1)}\left[SS_{error} + SS_{blocks}\right] \\
& \qquad\qquad \nearrow \qquad\qquad\qquad \nwarrow \\
& \quad \sigma_e^2\chi^2_{(a-1)(b-1)} \qquad (\sigma_e^2 + a\sigma_\beta^2)\chi^2_{(b-1)}
\end{aligned}
$$

Hence, the distribution of $S^2_{\bar{Y}_{i.}}$ is not a multiple of a central chi-square random variable.

Standard error for $\bar{Y}_{i.}$ is

$$S_{\bar{Y}_{i.}} = \sqrt{\frac{a-1}{ab} MS_{error} + \frac{1}{ab} MS_{blocks}}$$
$$= 2.4749$$

An approximate $(1-\alpha) \times 100\%$ confidence interval for $\mu + \alpha_i$ is

$$\bar{Y}_{i.} \pm t_{\nu, \frac{\alpha}{2}} \sqrt{\frac{a-1}{ab} MS_{error} + \frac{1}{ab} MS_{blocks}}$$

where

$$v = \frac{\left[\frac{a-1}{ab} MS_{error} + \frac{1}{ab} MS_{blocks}\right]^2}{\frac{\left(\frac{a-1}{ab} MS_{error}\right)^2}{(a-1)(b-1)} + \frac{\left(\frac{1}{ab} MS_{blocks}\right)^2}{b-1}}$$

$$= \frac{\left[\frac{4-1}{(4)(5)} MS_{error} + \frac{1}{ab} MS_{blocks}\right]^2}{\frac{\left(\frac{a-1}{ab} MS_{error}\right)^2}{(a-1)(b-1)} + \frac{\left(\frac{1}{ab} MS_{blocks}\right)^2}{b-1}} = 11.075$$

Result 10.1: Cochran-
Satterthwaite approximation

Suppose $MS_1, MS_2, \cdots, MS_k$ are mean squares with

- independent distributions
- degrees of freedom $= df_i$
- $\frac{(df_i)MS_i}{E(MS_i)} \sim \chi^2_{df_i}$

Then, for positive constants

$$a_i > 0, \qquad i = 1, 2, \ldots, k$$

the distribution of

$$S^2 = a_1 MS_1 + a_2 MS_2 + \ldots + a_k MS_k$$

is approximated by

$$\frac{vS^2}{E(S^2)} \quad \dot{\sim} \quad \chi^2_v$$

where

$$v = \frac{[E(S^2)]^2}{\frac{[a_1 E(MS_1)]^2}{df_1} + \ldots + \frac{[a_k E(MS_k)]^2}{df_k}}$$

is the value for the degrees of freedom.

In practice, the degrees of freedom are evaluated as

$$V = \frac{[S^2]^2}{\frac{(a_1 MS_1)^2}{df_1} + \ldots + \frac{(a_k MS_k)^2}{df_k}}$$

These are called the Cochran-Satterthwaite degrees of freedom.

Cochran, W.G. (1951) Testing a Linear Relation among Variances, Biometrics 7, 17-32.

Difference between two means: $E(\bar{Y}_{i.} - \bar{Y}_{k.})$

$$\begin{aligned}
&= E\left(\frac{1}{b}\sum_{j=1}^{b} Y_{ij} - \frac{1}{b}\sum_{j=1}^{b} Y_{kj}\right) \\
&= E\left(\frac{1}{b}\sum_{j=1}^{b}(Y_{ij} - Y_{kj})\right) \\
&= E\left(\frac{1}{b}\sum_{j=1}^{b}(\mu + \alpha_i + \beta_j + \epsilon_{ij} \right. \\
&\qquad\qquad \left. -\mu - \alpha_k - \beta_j - \epsilon_{kj})\right) \\
&= \alpha_i - \alpha_k + \frac{1}{b}\sum_{j=1}^{b} \underline{E(\epsilon_{ij} - \epsilon_{kj})} \\
&\qquad\qquad\qquad\qquad \uparrow \text{this is zero} \\
&= \alpha_i - \alpha_k \\
&= (\mu + \alpha_i) - (\mu + \alpha_k)
\end{aligned}$$

whether block effects are fixed or random.

$$\begin{aligned}
Var(\bar{Y}_{i.} - \bar{Y}_{k.}) &= Var\left(\alpha_i - \alpha_k + \frac{1}{b}\sum_{j=1}^{b}(\epsilon_{ij} - \epsilon_{kj})\right) \\
&= \frac{1}{b^2}\sum_{j=1}^{b} Var(\epsilon_{ij} - \epsilon_{kj}) \\
&= \frac{2\sigma_e^2}{b}
\end{aligned}$$

The standard error for $\bar{Y}_{i.} - \bar{Y}_{k.}$ is

$$S_{\bar{Y}_{i.} - \bar{Y}_{k.}} = \sqrt{\frac{2MS_{error}}{b}}$$

A $(1-\alpha) \times 100\%$ confidence interval for $\alpha_i - \alpha_i$ is

$$(\bar{Y}_{i.} - \bar{Y}_{k.}) \pm t_{\underline{(a-1)(b-1)},\frac{\alpha}{2}}\sqrt{\frac{2MS_{error}}{b}}$$

$$\uparrow$$
d.f. for MS_{error}

<u>t-test</u>:

Reject $H_0 : \alpha_i - \alpha_k = 0$ if

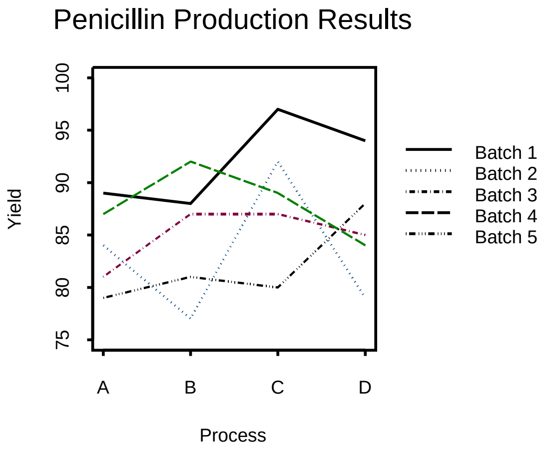
$$|t| = \frac{|\bar{Y}_{i.} - \bar{Y}_{j.}|}{\sqrt{\frac{2MS_{error}}{b}}} > t_{\underline{(a-1)(b-1)},\frac{\alpha}{2}}$$

$$\uparrow$$
d.f. for MS_{error}

```
# Analyze the penicillin data from Box,
# Hunter, and Hunter.  This code is
# posted as  penclln.ssc

# Enter the data into a data frame and
# change the Batch and Process variables
# into factors

> penclln <- read.table("penclln.dat",
+     col.names=c("Batch","Process","Yield"))
> penclln$Batch <- as.factor(penclln$Batch)
> penclln$Process <- as.factor(penclln$Process)
> penclln
```

```
   Batch Process Yield
1      1       1    89
2      1       2    88
3      1       3    97
4      1       4    94
5      2       1    84
6      2       2    77
7      2       3    92
8      2       4    79
9      3       1    81
10     3       2    87
11     3       3    87
12     3       4    85
13     4       1    87
14     4       2    92
15     4       3    89
16     4       4    84
17     5       1    79
18     5       2    81
19     5       3    80
20     5       4    88
```

```
# Construct a profile plot.  UNIX users
# should use the  motif( ) command to open
# a graphics window

> attach(penclln)
> means <- tapply(Yield,list(Process,Batch),mean)

> par(fin=c(6,7),cex=1.2,lwd=3,mex=1.5)
> x.axis <- unique(Process)
> matplot(c(1,4), c(75,100), type="n", xaxt="n",
+      xlab="Process",  ylab="Yield",
+      main= "Penicillin Production Results")
> axis(1, at=(1:4)*1, labels=c("A", "B", "C", "D"))
> matlines(x.axis,means,type=’l’,lty=1:5,lwd=3)

> legend(4.2,95, legend=c(’Batch 1’,’Batch 2’,
+  ’Batch 3’,’Batch 4’,’Batch 5’), lty=1:5,bty=’n’)
> detach( )
```



```
#  Use the  lme( )  function to fit a model
#  with additive batch (random) and process
#  (fixed) effects and create diagnostic plots.

> options(contrasts=c("contr.treatment",
+    "contr.poly"))

> penclln.lme <- lme(Yield ~ Process,
+    random= ~ 1|Batch, data=penclln,
+    method=c("REML"))

> summary(penclln.lme)

Linear mixed-effects model fit by REML
 Data: penclln
       AIC      BIC    logLik
  83.28607 87.92161 -35.64304
```

```
Random effects:
 Formula:  ~ 1 | Batch

        (Intercept) Residual
StdDev:    3.433899 4.339739


Fixed effects: Yield ~ Process
            Value Std.Error DF  t-value p-value
(Intercept)    84  2.474874 12 33.94113  <.0001
   Process2     1  2.744692 12  0.36434  0.7219
   Process3     5  2.744692 12  1.82170  0.0935
   Process4     2  2.744692 12  0.72868  0.4802

 Correlation:
         (Intr) Prcss2 Prcss3
Process2 -0.555
Process3 -0.555  0.500
Process4 -0.555  0.500  0.500

Standardized Within-Group Residuals:
       Min         Q1        Med        Q3     Max
 -1.415158 -0.5017351 -0.1643841 0.6829939 1.28365

Number of Observations: 20
Number of Groups: 5
```

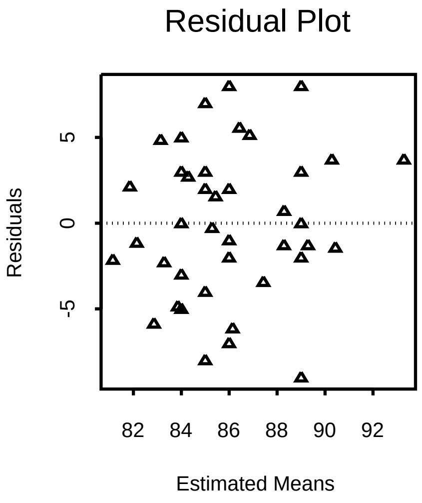
```
> names(penclln.lme)

 [1] "modelStruct"    "dims"         "contrasts"
 [4] "coefficients"   "varFix"       "sigma"
 [7] "apVar"          "logLik"       "numIter"
[10] "groups"         "call"         "method"
[13] "fitted"         "residuals"    "fixDF"


> #  Contruct ANOVA table for fixed effects
> anova(penclln.lme)

            numDF denDF  F-value p-value
(Intercept)     1    12 2241.213  <.0001
    Process     3    12    1.239  0.3387
```

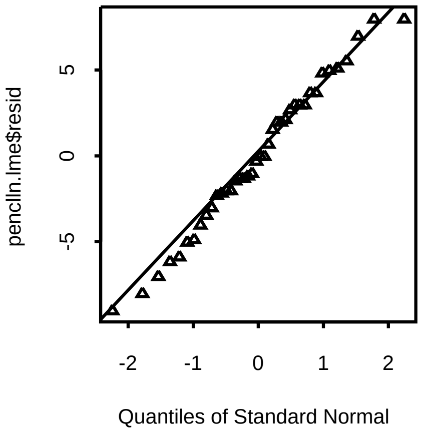
```
> #  Estimated parameters for fixed effects
> coef(penclln.lme)

  (Intercept) Process2 Process3 Process4
1    88.28788        1        5        2
2    81.85606        1        5        2
3    83.28535        1        5        2
4    85.42929        1        5        2
5    81.14141        1        5        2

> # BLUP's for random effects
> ranef(penclln.lme)

  (Intercept)
1   4.2878780
2  -2.1439390
3  -0.7146463
4   1.4292927
5  -2.8585854
```

```
> # Confidence intervals for fixed effects
> # and estimated standard deviations
> intervals(penclln.lme)

Approximate 95% confidence intervals

 Fixed effects:
                 lower est.    upper
(Intercept)  78.6077137   84 89.39229
   Process2  -4.9801701    1  6.98017
   Process3  -0.9801701    5 10.98017
   Process4  -3.9801701    2  7.98017

 Random Effects:
  Level: Batch
                   lower     est.    upper
sd((Intercept)) 0.8555882 3.433899 13.78193

 Within-group standard error:
    lower     est.   upper
 2.464606 4.339739 7.64152
```

```
> # Create a listing of the original data
> # residuals and predicted values

> data.frame(penclln$Process,penclln$Batch,
+           penclln$Yield,
+           Pred=penclln.lme$fitted,
+           Resid=round(penclln.lme$resid,3))


   X1 X2 X3 Pred.fixed Pred.Batch Resid.fixed Resid.Batch
 1  1  1 89         84   88.28788           5       0.712
 2  2  1 88         85   89.28788           3      -1.288
 3  3  1 97         89   93.28788           8       3.712
 4  4  1 94         86   90.28788           8       3.712
 5  1  2 84         84   81.85606           0       2.144
 6  2  2 77         85   82.85606          -8      -5.856
 7  3  2 92         89   86.85606           3       5.144
 8  4  2 79         86   83.85606          -7      -4.856
 9  1  3 81         84   83.28535          -3      -2.285
10  2  3 87         85   84.28535           2       2.715
11  3  3 87         89   88.28535          -2      -1.285
12  4  3 85         86   85.28535          -1      -0.285
13  1  4 87         84   85.42929           3       1.571
14  2  4 92         85   86.42929           7       5.571
15  3  4 89         89   90.42929           0      -1.429
16  4  4 84         86   87.42929          -2      -3.429
17  1  5 79         84   81.14141          -5      -2.141
18  2  5 81         85   82.14141          -4      -1.141
19  3  5 80         89   86.14141          -9      -6.141
20  4  5 88         86   83.14141           2       4.859
```

```
> # Create residual plots

> frame( )
> par(fin=c(7,7),cex=1.2,lwd=3,mex=1.5)
>   plot(penclln.lme$fitted, penclln.lme$resid,
+       xlab="Estimated Means",
+       ylab="Residuals",
+       main="Residual Plot")
>  abline(h=0, lty=2, lwd=3)

>  qqnorm(penclln.lme$resid)
>  qqline(penclln.lme$resid)
```





Example 10.2 Pigment production

In this example the main objective is the estimation of the variance components

Source of Variation	d.f.	MS	E(MS)
Batches	15-1=14	86.495	$\sigma_e^2 + 2\sigma_\delta^2 + 4\sigma_\beta^2$
Samples in Batches	15(2-1)=15	57.983	$\sigma_e^2 + 2\sigma_\delta^2$
Tests in Samples	(30)(2-1)=30	0.917	σ_e^2

Estimates of variance components:

$$\begin{aligned}\hat{\sigma}_e^2 &= MS_{tests} = 0.917 \\ \hat{\sigma}_\delta^2 &= \frac{MS_{samples} - MS_{tests}}{2} = 28.533 \\ \hat{\sigma}_\beta^2 &= \frac{MS_{batches} - MS_{samples}}{4} = 7.128\end{aligned}$$

```
/* This is a SAS program for analyzing
   data from a nested or heirarchical
   experiment.  This program is posted
   as
               pigment.sas

   The data are measurements of moisture
   content of a pigment taken from Box,
   Hunter and Hunter (page 574).*/

data set1;
  infile ’pigment.dat’;
  input batch sample test y;
  run;

proc print data=set1;
  run;
```

```
/* The  "random"  statement in the following
   GLM procedure prints of formulas for
   expectations of mean squares.  These results
   are used in variance component estimation  */

proc glm data=set1;
  class batch sample;
  model y = batch sample(batch) / e1;
  random batch sample(batch) / q test;
  run;
```

```
/*  Alternatively, REML estimates of variance
    components are  produced by the MIXED
    procedure in SAS.  Note that there are
    no terms on the rigth of the equal sign in
    the model statement because the only
    non-random effect is the intercept.        */

proc mixed data=set1;
  class batch sample test;
  model y = ;
  random batch sample(batch);
  run;


/*  Use the MIXED procedure in SAS to compute
    maximum likelihood estimates of variance
    components */

proc mixed data=set1 method=ml;
  class batch sample test;
  model y = ;
  random batch sample(batch);
  run;
```

```
OBS    BATCH    SAMPLE    TEST     Y

  1       1         1        1     40
  2       1         1        2     39
  3       1         2        1     30
  4       1         2        2     30
  5       2         1        1     26
  6       2         1        2     28
  7       2         2        1     25
  8       2         2        2     26
  9       3         1        1     29
 10       3         1        2     28
 11       3         2        1     14
 12       3         2        2     15
 13       4         1        1     30
 14       4         1        2     31
 15       4         2        1     24
 16       4         2        2     24
 17       5         1        1     19
 18       5         1        2     20
 19       5         2        1     17
 20       5         2        2     17
 21       6         1        1     33
 22       6         1        2     32
 23       6         2        1     26
 24       6         2        2     24
 25       7         1        1     23
 26       7         1        2     24
 27       7         2        1     32
 28       7         2        2     33
 29       8         1        1     34
 30       8         1        2     34
```

```
OBS     BATCH    SAMPLE   TEST     Y

 31       8        2        1     29
 32       8        2        2     29
 33       9        1        1     27
 34       9        1        2     27
 35       9        2        1     31
 36       9        2        2     31
 37      10        1        1     13
 38      10        1        2     16
 39      10        2        1     27
 40      10        2        2     24
 41      11        1        1     25
 42      11        1        2     23
 43      11        2        1     25
 44      11        2        2     27
 45      12        1        1     29
 46      12        1        2     29
 47      12        2        1     31
 48      12        2        2     32
 49      13        1        1     19
 50      13        1        2     20
 51      13        2        1     29
 52      13        2        2     30
 53      14        1        1     23
 54      14        1        2     24
 55      14        2        1     25
 56      14        2        2     25
 57      15        1        1     39
 58      15        1        2     37
 59      15        2        1     26
 60      15        2        2     28
```

```
           General Linear Models Procedure
              Class Level Information

   Class   Levels  Values

   BATCH     15    1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

   SAMPLE     2    1 2


        Number of observations in the data set = 60


Dependent Variable: Y

Source   DF  Sum of Squares  Mean Square  F Value   Pr > F

Model    29    2080.68         71.7477     78.27    0.0001

Error    30      27.5000        0.9167

C.Total  59    2108.18333333


Source         DF   Type I SS   Mean Square  F Value   Pr > F

BATCH          14    1210.93      86.4952     94.36    0.0001
SAMPLE(BATCH)  15     869.75      57.9833     63.25    0.0001
```

```
Source          Type I Expected Mean Square

BATCH           Var(Error) + 2 Var(SAMPLE(BATCH))
                   + 4 Var(BATCH)

SAMPLE(BATCH)   Var(Error) + 2 Var(SAMPLE(BATCH))


Dependent Variable: y

 Source       DF  Type I SS     MS       F    Pr>F

 batch        14   1210.933   86.495  1.49  0.2256

 Error:       15    869.750   57.983
 MS(sample(batch))


 Source         DF  Type I SS    MS       F     Pr>F

 sample(batch)  15   869.75   57.983  63.25  <.0001

 Error:
  MS(Error)     30    27.50    0.917
```

```
                The MIXED Procedure

              Class Level Information

   Class   Levels  Values

   BATCH    15     1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
   SAMPLE    2     1 2
   TEST      2     1 2



          REML Estimation Iteration History

      Iteration  Evaluations     Objective     Criterion

             0            1    274.08096606
             1            1    183.82758851     0.0000000

              Convergence criteria met.
```

```
          Covariance Parameter Estimates (REML)

 Cov Parm         Ratio  Estimate  Std Error   Z  Pr > |Z|

 BATCH           7.7760   7.1280    9.7373   0.73  0.4642
 SAMPLE(BATCH)  31.1273  28.5333   10.5869   2.70  0.0070
 Residual        1.0000   0.9167    0.2367   3.87  0.0001



             Model Fitting Information for Y

         Description                        Value

         Observations                     60.0000
         Variance Estimate                 0.9167
         Standard Deviation Estimate       0.9574
         REML Log Likelihood             -146.131
         Akaike's Information Criterion  -149.131
         Schwarz's Bayesian Criterion    -152.247
         -2 REML Log Likelihood          292.2623
```

```
                 The MIXED Procedure

               Class Level Information

 Class     Levels  Values

 BATCH      15     1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
 SAMPLE      2     1 2
 TEST        2     1 2



           ML Estimation Iteration History

 Iteration  Evaluations     Objective     Criterion

        0            1  273.55423884
        1            1  184.15844023    0.00000000

              Convergence criteria met.
```

```
              Covariance Parameter Estimates (MLE)

Cov Parm        Ratio    Estimate  Std Error   Z  Pr > |Z|

BATCH           6.2033    5.68639    9.07341  0.63  0.5309
SAMPLE(BATCH)  31.1273   28.53333   10.58692  2.70  0.0070
Residual        1.0000    0.91667    0.23668  3.87  0.0001


                 Model Fitting Information for Y

      Description                          Value

      Observations                       60.0000
      Variance Estimate                   0.9167
      Standard Deviation Estimate         0.9574
      Log Likelihood                    -147.216
      Akaike's Information Criterion    -150.216
      Schwarz's Bayesian Criterion      -153.357
      -2 Log Likelihood                  294.4311
```

Estimation of $\mu = E(Y_{ijk})$:

$$\hat{\mu} = \bar{Y}... = \frac{1}{bsr} \sum_{i=1}^{b} \sum_{j=1}^{s} \sum_{k=1}^{r} Y_{ijk}$$

$$E(\bar{Y}...) = \mu$$

$$Var(\bar{Y}...) = \frac{1}{bsr}(\sigma_e^2 + r\sigma_\delta^2 + Sr\sigma_\beta^2)$$

<u>Standard error</u>:

$$\begin{aligned} S_{\bar{Y}...} &= \sqrt{\frac{1}{bsr}\left(\hat{\sigma}_e^2 + r\hat{\sigma}_\delta^2 + sr\hat{\sigma}_\beta^2\right)} \\ &= \sqrt{\frac{1}{bsr}\left(MS_{Batches}\right)} \\ &= \sqrt{\frac{86.495}{60}} = 1.4416 \end{aligned}$$

A 95% confidence interval for μ

$$\bar{Y}... \pm t_{14,.025} S_{\bar{Y}...}$$

$\nearrow$

df for $MS_{Batches}$

Here, $t_{14,.025} = 2.510$ and the confidence interval is

$$26.783 \ \pm \ (2.510)(1.4416)$$

$$\Rightarrow (23.16, 30.40)$$

```
> # This file is stored as   pigment.spl

> pigment <- read.table("pigment.dat",
+     col.names=c("Batch","Sample","Test","Y"))
> pigment$Batch <- as.factor(pigment$Batch)
> pigment$Sample <- as.factor(pigment$Sample)
> pigment

   Batch Sample Test  Y
 1     1      1    1 40
 2     1      1    2 39
 3     1      2    1 30
 4     1      2    2 30
 5     2      1    1 26
 6     2      1    2 28
 7     2      2    1 25
 8     2      2    2 26
 9     3      1    1 29
10     3      1    2 28
11     3      2    1 14
12     3      2    2 15
 .     .      .    .  .
 .     .      .    .  .
```

```
> # The function raov() may be used for
> # balanced designs with only random effects,
> # and gives a conventional analysis including
> # the estimation of variance components.
> # The function varcomp() is more general. It
> # may be used to estimate variance components
> # for balanced or unbalanced mixed models.

> # raov(): Random Effects Analysis of Variance
>
> pigment.raov <- raov(Y ~ Batch/Sample,
+                            data=pigment)

> # or you could use
> # pigment.raov <- raov(Y ~ Batch +
> #           Batch:Sample, data=pigment)
> # pigment.raov <- raov(Y ~ Batch +
> #           Sample %in% Batch, data=pigment)
```

```
> summary(pigment.raov)

                  Df Sum of Sq  Mean Sq Est. Var.
            Batch 14  1210.933 86.49524   7.12798
Sample %in% Batch 15   869.750 57.98333  28.53333
        Residuals 30    27.500  0.91667   0.91667

> names(pigment.raov)
 [1] "coefficients"  "residuals"  "fitted.values"
 [4] "effects"       "R"             "rank"
 [7] "assign"        "df.residual"  "contrasts"
[10] "terms"         "call"          "model"
[13] "replications"  "ems.coef"

> pigment.raov$rep
 Batch Sample %in% Batch
     4                 2

> pigment.raov$ems.coef
                           Sample
                  Batch  %in% Batch   Residuals
            Batch     4           0          0
Sample %in% Batch     2           2          0
        Residuals     1           1          1
```

```
> # The same variance component estimates can be
> # found using varcomp(), but this allows mixed
> # models and we first must declare which factors
> # are random using the  is.random() function.
> # All factors in the data frame are established
> # as random effects by the following
>
> is.random(pigment) <- T
> is.random(pigment)
 Batch Sample
     T      T
>
> #  The possible estimation methods are
> # "minque0": minimum norm quadratic estimators
> #            (the default)
> # "reml"   : residual (or reduced or restricted)
> #            maximum likelihood.
> # "ml"     : maximum likelihood.
```

```
> varcomp(Y ~ Batch/Sample, data=pigment,
+                method="reml")$var

   Batch Sample %in% Batch Residuals
 7.12866          28.53469  0.916641


> varcomp(Y ~ Batch/Sample, data=pigment,
+                 method="ml")$var

   Batch Sample %in% Batch Residuals
 5.68638          28.53333 0.9166668
```

Properties of ANOVA methods for variance component estimation:

(i) Broad applicability
- easy to compute in balanced cases
- ANOVA is widely known
- not required to completely specify distributions for random effects

(ii) Unbiased estimators

(iii) Sampling distribution is not exactly known, even under the usual normality assumptions (except for $\hat{\sigma}_e^2 = MS_{error}$)

(iv) May produce negative estimates of variances

(v) REML estimates have the same values
- in simple balanced cases
- when ANOVA estimates of variance components are inside the parameter space

(vi) For unbalanced studies, there may be no "natural" way to choose

$$\hat{\sigma}^2 = \sum_{i=1}^{k} a_i MS_i$$

<u>Result 10.2</u>: If $MS_1, MS_2, \ldots, MS_k$ are distributed independently with

$$\frac{(df_i)MS_i}{E(MS_i)} \sim \chi^2_{df_i}$$

and constants $a_i > 0, \quad i = 1, 2, \ldots, k$ are selected so that

$$\hat{\sigma}^2 = \sum_{i=1}^{k} a_i MS_i$$

has expectation σ^2, then

$$Var(\hat{\sigma}^2) = 2 \sum_{i=1}^{k} \frac{a_i^2 [E(MS_i)]^2}{df_i}$$

and an unbiased estimator of this variance is

$$\widehat{Var}(\hat{\sigma}^2) = \frac{2a_i^2 MS_i^2}{(df_i + 2)}$$

<u>Proof</u>: Since $\frac{(df_i)MS_i}{E(MS_i)} \sim \chi^2_{df_i}$ and $E(\chi^2_{df_i}) = df_i$ and $Var(\chi^2_{df_i}) = 2df_i$, it follows that

$$\begin{aligned} Var(MS_i) &= Var(\frac{E(MS_i)\chi^2_{df_i}}{df_i}) \\ &= \frac{2[E(MS_i)]^2}{df_i} \end{aligned}$$

From the independence of the MS_i's, we have

$$\begin{aligned} Var(\hat{\sigma}^2) &= \sum_{i=1}^{k} a_i^2 Var(MS_i) \\ &= 2 \sum_{i=1}^{k} \frac{a_i^2 [E(MS_i)]^2}{df_i} \end{aligned}$$

Furthermore,

$$\begin{aligned} E(MS_i^2) &= Var(MS_i) + [E(MS_i)]^2 \\ &= \frac{2[E(MS_i)]^2}{df_i} + [E(MS_i)]^2 \\ &= \left(\frac{df_i + 2}{df_i}\right) [E(MS_i)]^2 \end{aligned}$$

Consequently,

$$E\left[2 \sum_{i=1}^{k} \frac{a_i^2 MS_i^2}{(df_i + 2)}\right] = Var(\hat{\sigma}^2)$$

A "standard error" for

$$\hat{\sigma}^2 = \sum_{i=1}^{k} a_i MS_i$$

could be reported as

$$S_{\hat{\sigma}^2} = \sqrt{2 \sum_{i=1}^{k} \frac{a_i^2 MS_i^2}{(df_i + 2)}}$$

Using the Cochran-Satterthwaite approximation (Result 10.1), an approximate $(1-\alpha) \times 100\%$ confidence interval for σ^2 could be constructed as:

$$\begin{aligned} 1 - \alpha &\doteq Pr\left\{\chi^2_{\nu,1-\alpha/2} \leq \frac{v\hat{\sigma}^2}{\sigma^2} \leq \chi^2_{\nu,\alpha/2}\right\} \\ &= Pr\left\{\frac{v\hat{\sigma}^2}{\chi^2_{\nu,\alpha/2}} \leq \sigma^2 \leq \frac{v\hat{\sigma}^2}{\chi_{\nu,1-\alpha/2}}\right\} \end{aligned}$$

where $\hat{\sigma}^2 = \sum_{i=1}^{k} a_i MS_i$ and

$$v = \frac{\left[\sum_{i=1}^{k} a_i MS_i\right]^2}{\sum_{i=1}^{k} \frac{[a_i MS_i]^2}{df_i}}$$

10.3 Likelihood-based methods:

Consider the mixed model

$$\mathrm{Y}_{n\times 1} = X\beta_{p\times 1} + Z\mathrm{u}_{q\times 1} + \mathrm{e}_{n\times 1}$$

where

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

Then,

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

where $\Sigma = ZGZ^T + R$

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

Maximum Likelihood Estimation

Multivariate normal likelihood:

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

The log-likelihood function is

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

Given the values of the observed responses, Y, find values β 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"

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

whose distribution does not depend on

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

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

$$\mathbf{e} = \begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix} = (I - P_1)\mathbf{Y} \sim N(0, \sigma^2(I - P_1))$$

The MLE $\hat{\sigma}^2 = \frac{1}{n} \Sigma_{j=1}^{n} e_j^2$ fails to acknowledge that $\mathbf{e}$ is restricted to an $(n-1)$-dimensional space, i.e., $\Sigma_{j=1}^{n} e_j = 0$.

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

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

Then

$$\mathbf{e} = \begin{bmatrix} Y_1 - \bar{Y} \\ Y_2 - \bar{Y} \\ Y_3 - \bar{Y} \\ Y_4 - \bar{Y} \end{bmatrix} = (I - P_1)\mathbf{Y}$$

$$\sim \quad N(0, \underset{\substack{\uparrow \\ \text{This covariance} \\ \text{matrix is singular.}}}{\underline{\sigma^2(I - P_1)}})$$

Here, $m = rank(I - P_1) = n - 1 = 3$.

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

$$\begin{aligned} \mathbf{r} &= \begin{bmatrix} r_1 \\ r_2 \\ r_3 \end{bmatrix} = \begin{bmatrix} Y_1 + Y_2 - Y_3 - Y_4 \\ Y_1 - Y_2 + Y_3 - Y_4 \\ Y_1 - Y_2 - Y_3 + Y_4 \end{bmatrix} \\ &= M(I - P_1)\mathbf{Y} \\ &\sim N(0, \underset{\substack{\uparrow \\ \text{call this } \sigma^2 W}}{\underline{\sigma^2 M(I - P_1)M^T}}) \end{aligned}$$

Restricted Likelihood function:

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

Restricted Log-likelihood:

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

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

(Restricted) likelihood equation:

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

Solution (REML estimator for σ^2):

$$\begin{aligned}\hat{\sigma}^2_{REML} \;&=\; \frac{1}{m}\mathrm{r}^T W^{-1} r \\ &=\; \frac{1}{m}\mathrm{Y}^T \underline{(I - P_1)^T M^T (M(I - P_1)M^T)^{-1} M(I - P_1)\mathrm{Y}}\end{aligned}$$

$\nearrow$

This is a projection of Y onto the column space of $M(I - P_1)$ which is the column space of $I - P_1$

$$\begin{aligned}&= \frac{1}{m}\mathrm{Y}^T (I - P_1)\mathrm{Y} \\ &= \frac{1}{n-1}\sum_{j=1}^{n} (Y_j - \bar{Y})^2 = S^2\end{aligned}$$

REML (Restricted Maximum Likelihood) estimation

- Estimate parameters in

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

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

- 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:

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

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

Then

$$\begin{aligned} L\mathrm{Y} &= L(X\beta + Z\mathrm{u} + \mathrm{e}) \\ &= LX\beta + LZ\mathrm{u} + L\mathrm{e} \end{aligned}$$

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

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

for some M.

(Here $P_X = X(X^T X)^- X^T$)

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

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

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

The "Restricted" likelihood is

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

The resulting log-likelihood is

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

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

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

as

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

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

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

Using the REML estimator for

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

an approximation is

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

and for “large” samples:

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

<u>Prediction of random effects</u>:

Given the observed responses Y, predict the value of u.

For our model,

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

Then (from result 4.1)

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

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

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

$\uparrow$

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

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

Then, the BLUP for u is

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

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 u is

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

For "large" samples, the distribution of $\hat{\mathrm{u}}$ is approximately multivariate normal with mean vector 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 $\hat{\mathrm{u}}$ provide a solution to the mixed model equations:

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

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 $\begin{bmatrix} \hat{\beta} \\ \hat{\mathrm{u}} \end{bmatrix}$

References:

Bates, D.M. and Pinheiro, J.C. (1998) "Computational methods for multilevel models" available in PostScript or PDF formats at http://franz.stat.wisc.edu/pub/NLME/

Davidian, M. and Giltinan, D.M. (1995) "Nonlinear Mixed Effects Models for Repeated Measurement Data", Chapman and Hall.

Hartley, H.O. and Rao, J.N.K. (1967) Maximum-likelihood estimation for the mixed analysis of variance model, Biometrika, 54, 93-108.

Harville, D.A. (1977) Maximum likelihood approaches to variance component estimation and to related problems, Journal of the American Statistical Association, 72, 320-338.

Jennrich, R.I. and Schluchter, M.D. (1986) Unbalanced repeated-measures models with structured covariance matrices, Biometrics, 42, 805-820.

Kackar, R.N. and Harville, D.A. (1984) Approximations for standard errors of estimators of fixed and random effects in mixed linear models. Journal of the American Statistical Association, 79, 853-862.

Laird, N.M. and Ware, J.H. (1982) Random-effects models for longitudinal data, Biometrics, 38, 963-974.

Lindstrom, M.J. and Bates, D.M. (1988) Newton-Raphson and EM algorithms for linear-mixed-effects models for repeated-measures data. Journal of the American Statistical Association, 83, 1014-1022.

Littel, R.C., Milliken, G.A., Stroup, W.W., and Wolfinger, R.D. (1997) "SAS Systems for Mixed Models", SAS Institute.

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, New York, Springer-Verlag.

Robinson, G.K. (1991) That BLUP is a good thing: the estimation of random effects, Statistical Science, 6, 15-51.

Searle, S.R., Casella, G. and McCulloch, C.E. (1992) Variance Components, New York, John Wiley & Sons.

Wolfinger, R.D., Tobias, R.D. and Sall, J. (1994) Computing Gaussian likelihoods and their derivatives for general linear mixed models, SIAM Journal on Scientific Computing, 15(6), 1294-1310.