STAT 714 fa 2023 Lec 06

Variance component estimation in random and mixed effects models

Karl B. Gregory

University of South Carolina

These slides are an instructional aid; their sole purpose is to display, during the lecture, definitions, plots, results, etc. which take too much time to write by hand on the blackboard. They are not intended to explain or expound on any material.

Mixed model setup

A *linear mixed model* has the form

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{u} + \mathbf{e},$$

where

- $\mathbf{X}$ is an $n \times p$ design matrix
- $\mathbf{b}$ is a $p \times 1$ vector of parameters describing *fixed effects*
- $\mathbf{Z}$ is an $n \times q$ design matrix
- $\mathbf{u} \sim \text{Normal}(\mathbf{0}, \mathbf{G})$ is a $q \times 1$ vector of *random effects*
- $\mathbf{e} \sim \text{Normal}(\mathbf{0}, \mathbf{R})$ is an $n \times 1$ vector of error terms independent of $\mathbf{u}$.

Letting $\mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}^T + \mathbf{R}$, we have $\mathbf{y} \sim \text{Normal}(\mathbf{X}\mathbf{b}, \mathbf{V})$.

Typically $\mathbf{V} = \mathbf{V}(\boldsymbol{\theta})$ depends on some parameters $\boldsymbol{\theta}$ which we wish to estimate.

One-way random effects model

For responses Y_{ij}, assume

$$Y_{ij} = \mu + A_i + \varepsilon_{ij}, \quad i = 1, \dots, a, \quad j = 1, \dots, n_i$$

where

- ε_{ij} are independent Normal$(0, \sigma_\varepsilon^2)$
- A_i are independent Normal$(0, \sigma_A^2)$
- A_i and ε_{ij} are independent

Goals:

1. Estimate μ, σ_A^2, and σ_ε^2.
2. "Predict" the realized values of the random effects $A_1, \dots, A_a$.
3. Test H_0: $\sigma_A^2 = 0$ versus H_1: $\sigma_A^2 > 0$.

Exercise: Put equations in matrix form $\mathbf{Y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{u} + \mathbf{e}$.

Two-way mixed effects model (Randomized complete block design)

For responses Y_{ijk}, assume

$$Y_{ijk} = \mu + \alpha_i + B_j + (\alpha B)_{ij} + \varepsilon_{ijk}, \quad i = 1, \dots, a, \quad j = 1, \dots, b, \quad k = 1, \dots, n_{ij}$$

where

- μ is a mean
- α_i are treatment effects
- ε_{ijk} are independent Normal$(0, \sigma_\varepsilon^2)$
- B_i are independent Normal$(0, \sigma_B^2)$
- $(\alpha B)_{ij}$ are independent Normal$(0, \sigma_{AB}^2)$
- B_i, $(AB)_{ij}$, and ε_{ij} are independent

Exercise: Put equations in matrix form $\mathbf{Y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{u} + \mathbf{e}$.

Maximum likelihood estimation of variance components

Let $\mathbf{y} \sim \text{Normal}(\mathbf{X}\mathbf{b}, \mathbf{V})$, where $\mathbf{V} = \mathbf{V}(\boldsymbol{\theta})$ for parameters $(\mathbf{b}, \boldsymbol{\theta}) \in \mathbb{R}^p \times \Theta$. Then

$$(\hat{\mathbf{b}}_{\text{mle}}, \hat{\boldsymbol{\theta}}_{\text{mle}}) = \underset{(\mathbf{b},\boldsymbol{\theta}) \in \mathbb{R}^p \times \Theta}{\text{argmax}} \ \ell(\mathbf{b}, \boldsymbol{\theta}; \mathbf{y}),$$

where

$$\ell(\mathbf{b}, \boldsymbol{\theta}; \mathbf{y}) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|\mathbf{V}| - \frac{1}{2}(\mathbf{y} - \mathbf{X}\mathbf{b})^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\mathbf{b})$$

is the log-likelihood function for $(\mathbf{b}, \boldsymbol{\theta})$ based on $\mathbf{y}$.

Exercise: Show that the MLEs can be found in two steps:

1. Find $\hat{\boldsymbol{\theta}}_{\text{mle}}$ by maximizing (in many cases numerically) the profile likelihood

$$\ell(\boldsymbol{\theta}; \mathbf{y}) = \text{const.} - \frac{1}{2}\log|\mathbf{V}| - \frac{1}{2}\mathbf{y}^T(\mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}^T\mathbf{V}^{-1})\mathbf{y}.$$

2. Find $\hat{\mathbf{b}}_{\text{mle}}$ satisfying $(\mathbf{X}^T\hat{\mathbf{V}}^{-1}\mathbf{X})\hat{\mathbf{b}}_{\text{mle}} = \mathbf{X}^T\hat{\mathbf{V}}^{-1}\mathbf{y}$, where $\hat{\mathbf{V}} = \mathbf{V}(\hat{\boldsymbol{\theta}}_{\text{mle}})$.

Exercise: Let $Y_{ij} = \mu + A_i + \varepsilon_{ij}$, $i = 1, \dots, a$, $j = 1, \dots, n$, with

$$A_i \overset{\text{ind}}{\sim} \text{Normal}(0, \sigma_A^2), \quad \varepsilon_{ij} \overset{\text{ind}}{\sim} \text{Normal}(0, \sigma_\varepsilon^2).$$

Show that the MLE of $(\mu, \sigma_A^2, \sigma_\varepsilon^2)$ is given by

$$(\hat{\mu}, \hat{\sigma}_A^2, \hat{\sigma}_\varepsilon^2) = \begin{cases} \left(\bar{y}_{..}, \dot{\sigma}_A^2, \dfrac{\text{SSE}}{a(n-1)}\right), & \text{if } \dot{\sigma}_A^2 > 0 \\ \left(\bar{y}_{..}, 0, \dfrac{\text{SSE} + \text{SSA}}{an}\right), & \text{if } \dot{\sigma}_A^2 \leq 0 \end{cases}$$

where $\text{SSA} = n\sum_{i=1}^a (\bar{y}_{i.} - \bar{y}_{..})^2$ and $\text{SSE} = \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{i.})^2$ and

$$\dot{\sigma}_A^2 = \frac{1}{n}\left[\frac{\text{SSA}}{a} - \frac{\text{SSE}}{a(n-1)}\right].$$

Useful: $(a\mathbf{I}_n + b\mathbf{J}_n)^{-1} = \frac{1}{a}(\mathbf{I}_n - \frac{b}{a+nb}\mathbf{J}_n)$ and $|a\mathbf{I}_n + b\mathbf{J}_n| = a^{n-1}(a + nb)$.

Restricted maximum likelihood REML principle

The *REML* approach to estimating variance comps. is to first remove fixed effects.

That is, for $\mathbf{y} \sim \text{Normal}(\mathbf{Xb}, \mathbf{V}(\boldsymbol{\theta}))$, estimate $\boldsymbol{\theta}$ based on $(\mathbf{I} - \mathbf{P_X})\mathbf{y}$.

More precisely:

1. Reparameterize $\mathbf{X}$ so that it has full-column rank (simplifies our exposition).
2. Then let $(\mathbf{I} - \mathbf{P_X}) = \mathbf{A}\mathbf{A}^T$ with $\mathbf{A}^T\mathbf{A} = \mathbf{I}_{n-p}$, with $p = \text{rank}\,\mathbf{X}$.
3. Then the REML approach estimates $\boldsymbol{\theta}$ based on $\mathbf{w} = \mathbf{A}^T\mathbf{y}$.

Exercise:

1. Obtain the "restricted" log-likelihood $\ell_R(\boldsymbol{\theta}; \mathbf{w})$ for $\boldsymbol{\theta}$ based on $\mathbf{w}$.
2. Show that $\ell_R(\boldsymbol{\theta}; \mathbf{w})$ is equal to

$$\ell_R(\boldsymbol{\theta}; \mathbf{y}) = \text{const.} - \frac{1}{2}\log|\mathbf{V}| - \frac{1}{2}\log|\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}| - \frac{1}{2}(\mathbf{y} - \mathbf{X}\mathbf{b}_{\text{gls}})^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\mathbf{b}_{\text{gls}}),$$

where $\mathbf{b}_{\text{gls}} = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{y}$.

See pgs. 192,193 of Monahan (2008).

Summary of REML approach

For $\mathbf{y} \sim \text{Normal}(\mathbf{X}\mathbf{b}, \mathbf{V}(\boldsymbol{\theta}))$, estimate $\boldsymbol{\theta}$ and $\mathbf{b}$ in two steps:

1. Obtain $\hat{\boldsymbol{\theta}}_R$ by maximizing the restricted log-likelihood

$$\ell_R(\boldsymbol{\theta}; \mathbf{y}) = \text{const.} - \frac{1}{2}\log|\mathbf{V}| - \frac{1}{2}\log|\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}| - \frac{1}{2}(\mathbf{y} - \mathbf{X}\mathbf{b}_{\text{gls}})^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\mathbf{b}_{\text{gls}})$$

2. Set $\hat{\mathbf{V}} = \mathbf{V}(\hat{\boldsymbol{\theta}}_R)$ and compute $\hat{\mathbf{b}}_R = (\mathbf{X}^T\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\hat{\mathbf{V}}^{-1}\mathbf{y}$.

Can be useful to write $\ell_R(\boldsymbol{\theta}; \mathbf{y})$ (scaling and ignoring the constant) as

$$\ell_R(\boldsymbol{\theta}; \mathbf{y}) = -\log|\mathbf{V}| - \log|\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}| - \mathbf{y}^T(\mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}^T\mathbf{V}^{-1})\mathbf{y}.$$

Exercise: Let $Y_{ij} = \mu + A_i + \varepsilon_{ij}$, $i = 1, \dots, a$, $j = 1, \dots, n$, with

$$A_i \overset{\text{ind}}{\sim} \text{Normal}(0, \sigma_A^2), \quad \varepsilon_{ij} \overset{\text{ind}}{\sim} \text{Normal}(0, \sigma_\varepsilon^2).$$

Show that the REML estimators of $(\sigma_A^2, \sigma_\varepsilon^2)$ are given by

$$(\hat{\sigma}_A^2, \hat{\sigma}_\varepsilon^2) = \begin{cases} \left(\dot{\sigma}_A^2, \dfrac{\text{SSE}}{a(n-1)}\right), & \text{if } \dot{\sigma}_A^2 > 0 \\ \left(0, \dfrac{\text{SSE} + \text{SSA}}{an - 1}\right), & \text{if } \dot{\sigma}_A^2 \leq 0, \end{cases}$$

where $\text{SSA} = n \sum_{i=1}^a (\bar{y}_{i.} - \bar{y}_{..})^2$ and $\text{SSE} = \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{i.})^2$ and

$$\dot{\sigma}_A^2 = \frac{1}{n}\left[\frac{\text{SSA}}{a-1} - \frac{\text{SSE}}{a(n-1)}\right].$$

Let $\mathbf{c}^T\mathbf{b}$ be an estimable contrast in the model $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{u} + \mathbf{e}$.

Consider predicting the realized value of $v = \mathbf{c}^T\mathbf{b} + \mathbf{d}^T\mathbf{u}$.

Best linear unbiased predictor (BLUP)

1. A predictor $\tilde{v}$ of v is an *unbiased predictor* if $\mathbb{E}\tilde{v} = \mathbb{E}v$.
2. A *linear predictor* of v has the form $\tilde{v} = a_0 + \mathbf{a}^T\mathbf{y}$.
3. A linear predictor $\tilde{v}$ is called the *best linear unbiased predictor* of v if

$$\mathbb{E}(v - \tilde{v})^2 \leq \mathbb{E}(v - \tilde{v}^*)^2$$

for all linear unbiased predictors $\tilde{v}^*$ of v.

Exercise: In the one-way random effects model, give $\mathbf{c}$ and $\mathbf{d}$ such that the corresponding v are the realized treatment means.

Result (BLUP for the mixed model)

In the mixed model setup $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{u} + \mathbf{e}$, *where*

$$\mathbf{u} \stackrel{\text{ind}}{\sim} \text{Normal}(\mathbf{0}, \mathbf{G}) \quad \textit{and} \quad \mathbf{e} \stackrel{\text{ind}}{\sim} \text{Normal}(\mathbf{0}, \mathbf{R}),$$

the BLUP for $v = \mathbf{c}^T\mathbf{b} + \mathbf{d}^T\mathbf{u}$, *where* $\mathbf{c}^T\mathbf{b}$ *is estimable, is given by*

$$\tilde{v} = \mathbf{c}^T\hat{\mathbf{b}}_{\text{gls}} + \mathbf{d}^T\mathbf{G}\mathbf{Z}^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}_{\text{gls}}).$$

Note: $\tilde{v}$ is infeasible (need $\mathbf{V}$). In practice replace $\mathbf{V}$ with $\hat{\mathbf{V}}$. No longer BLUP 🤷‍♂️.

Prove the above in two steps:

1. Show that $\mathbb{E}[v|\mathbf{y}] = \mathbf{c}^T\mathbf{b} + \mathbf{d}^T\mathbf{G}\mathbf{Z}^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\mathbf{b})$.
2. Minimize $\mathbb{E}(v - (a_0 + \mathbf{a}^T\mathbf{y}))^2$, ensuring that $a_0 + \mathbf{a}^T\mathbf{y}$ is unbiased.

Exercise: Let $Y_{ij} = \mu + A_i + \varepsilon_{ij}$, $i = 1, \dots, a$, $j = 1, \dots, n$, with

$$A_i \overset{\text{ind}}{\sim} \text{Normal}(0, \sigma_A^2), \quad \varepsilon_{ij} \overset{\text{ind}}{\sim} \text{Normal}(0, \sigma_\varepsilon^2).$$

Show that the BLUP for $v_i = \mu + A_i$ is given by

$$\tilde{v}_i = \left(\frac{n\sigma_A^2}{\sigma_\varepsilon^2 + n\sigma_A^2} \right) \bar{y}_{i.} + \left(\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + n\sigma_A^2} \right) \bar{y}_{..}, \quad i = 1, \dots, a.$$

Exercise: Let $Y_{ij} = \alpha + \beta x_{ij} + A_i + B_i x_{ij} + \varepsilon_{ij}$, $i = 1, \dots, n$, $j = 1, \dots, r$, where

$$A_i \overset{\text{ind}}{\sim} \text{Normal}(0, \sigma_A^2), \quad B_i \overset{\text{ind}}{\sim} \text{Normal}(0, \sigma_B^2), \quad \varepsilon_{ij} \overset{\text{ind}}{\sim} \text{Normal}(0, \sigma_\varepsilon^2).$$

With the `sleepstudy` data in the R package `lme4`, obtain the following:

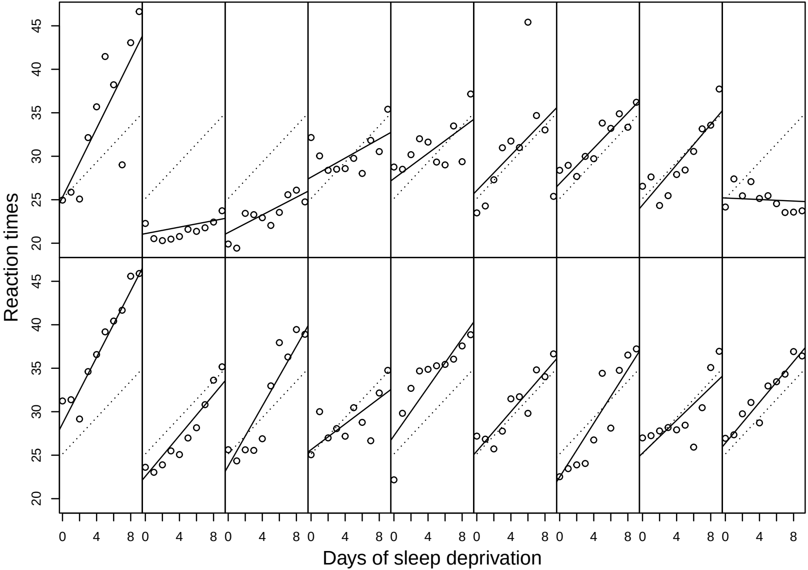
1. The REML estimates for σ_A^2, σ_B^2, and σ_ε^2.
2. Estimates of the fixed effects parameters α and β.
3. Predictions for the realized values of the random effects.

```r
# pull data from lme4 package
library(lme4)
y <- sleepstudy$Reaction/10 # rescale for numerical stability with optim()
x <- sleepstudy$Days

# construct design matrices X and Z
n <- 18; r <- 10
X <- cbind(rep(1,n*r),x)
Z <- matrix(0,n*r,2*n)
for(i in 1:n){

  ind <- ((i-1)*r + 1):(i*r)
  Z[ind,i] <- 1
  Z[ind,n + i] <- x[ind]

}
```

```
# define a function which is the negative restricted log-likelihood
negllR_slp <- function(th,y,X,Z){

  sgA <- th[1]; sgB <- th[2]; sge <- th[3]

  # build V
  G <- cbind(diag(rep(sgA^2,n)),matrix(0,n,n))
  G <- rbind(G,cbind(matrix(0,n,n),diag(rep(sgB^2,n))))
  V <- Z %*% G %*% t(Z) + diag(rep(sge^2,n*r))

  # compute restricted log-likelihood
  Vin <- solve(V); A <- t(X) %*% Vin %*% X
  d1 <- det(V); d2 <- det(A)
  quad <- t(y) %*% ( Vin - Vin %*% X %*% solve(A) %*% t(X) %*% Vin) %*% y
  return(log(d1) + log(d2) + quad)

}
```

```
# use optim() to find the REML estimators (this is a clumsy way)
optim_out <- optim(par = c(1,1,1), fn = negllR_slp, y = y, X = X, Z = Z)
th_hat <- optim_out$par
sgA_hat <- th_hat[1]; sgB_hat <- th_hat[2]; sge_hat <- th_hat[3]

# construct V_hat
G_hat <- cbind(diag(rep(sgA_hat^2,n)),matrix(0,n,n))
G_hat <- rbind(G_hat,cbind(matrix(0,n,n),diag(rep(sgB_hat^2,n))))
V_hat <- Z %*% G_hat %*% t(Z) + diag(rep(sge_hat^2,n*r))

# compute estimated fixed effects
b_hat <- solve(t(X) %*% solve(V_hat) %*% X) %*% t(X) %*% solve(V_hat)%*% y
A_hat <- b_hat[1]
B_hat <- b_hat[2]

# obtain predicted values of A and B (which is u)
u_pr <- G_hat %*% t(Z) %*% solve(V_hat) %*% ( y - X %*% b_hat)
A_pr <- u_pr[1:n]
B_pr <- u_pr[(n+1):(2*n)]
```

```
# compare output to that of the lmer function from the lme4 package
summary(lmer(Reaction/10 ~ Days + (Days | Subject), sleepstudy))

# make plots
par(mfrow = c(2,9),mar = c(0,0,0,0), oma = c(4,4,1,1))
xlims <- range(x); ylims <- range(y)
for( i in 1:n){

  ind <- ((i-1)*r + 1):(i*r)
  plot(y[ind]~sleepstudy$Days[ind],xlim = xlims, ylim = ylims,
       xaxt = "n",yaxt = "n")

  if(i %in% c(1,10)) axis(side = 2)
  if(i %in% c(10:18)) axis(side = 1)

  abline(A_hat + A_pr[i], B_hat + B_pr[i])
  abline(A_hat,B_hat,lty = 3)

}

mtext(side = 1, outer = TRUE, text = "Days of sleep deprivation",line=2.5)
mtext(side = 2, outer = TRUE, text = "Reaction times", line = 2.5)
```


Reaction times
Days of sleep deprivation
20
25
30
35
40
45
0
4
8

Exercise: Let $Y_{ij} = \mu + A_i + \varepsilon_{ij}$, $i = 1, \dots, a$, $j = 1, \dots, n$, with

$$A_i \overset{\text{ind}}{\sim} \text{Normal}(0, \sigma_A^2), \quad \varepsilon_{ij} \overset{\text{ind}}{\sim} \text{Normal}(0, \sigma_\varepsilon^2).$$

Define the sums of squares

$$\text{SST} = \textstyle\sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{..})^2, \; \text{SSA} = n \sum_{i=1}^a (\bar{y}_{i.} - \bar{y}_{..})^2, \; \text{SSE} = n \sum_{i=1}^a (y_{ij} - \bar{y}_{i.})^2$$

as well as the mean squares $\text{MSA} = \text{SSA}/(a-1)$ and $\text{MSE} = \text{SSE}/(a(n-1))$.

1. Show that $\text{SSA}/(\sigma_\varepsilon^2 + n\sigma_A^2) \sim \chi^2_{a-1}$
2. Show that $\text{SSE}/\sigma_\varepsilon^2 \sim \chi^2_{a(n-1)}$
3. Give $\mathbb{E}\,\text{MSA}$ and $\mathbb{E}\,\text{MSE}$.
4. Give ANOVA (method of moments) estimators of σ_A^2 and σ_ε^2.
5. Give the distribution of MSA / MSE.
6. Give a size-α test for H_0: $\sigma_A^2 = 0$ based on MSA / MSE. Give power function.

Monahan, J. F. (2008). *A primer on linear models*. CRC Press.