LMM Summary

Sunday, September 24, 2023 17:13

When do we need LMMs? For clustered data!

- Longitudinal data → y_{ij} — i: subject, e.g. a student; j: repeated measurements, e.g. annual math score
- Multi-level data → y_{ij} — i: groups, e.g. each school; j: observations, e.g. student score

INCORRECT MODEL:

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \epsilon_{ij} \qquad \epsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$$

i: subject, j: week

IID does not hold b/c errors are not independent

↳ conflates WITHIN & BETWEEN subject variability

Cannot forecast individual subject's growth curve.

FIXED EFFECTS MODEL

$$y_{ij} = \beta_{0i} + \beta_1 x_{ij} + \epsilon_{ij} \qquad \epsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$$

β_{0i}: A subject-specific intercept. β_1: constant slope, same for every subject

now reasonable b/c β_{0i} removes subject-specific effect

↳ now only captures WITHIN subject variability

HOWEVER:

- This subject-specific intercept can't leverage population info → smaller sample size → ↓ precision
- Cannot forecast NEW subject's growth curve.

Fixed versus Random

Consider the model:

$$y_{ij} = \theta_i + \beta' x_{ij} + \epsilon_{ij},\ \epsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2) \theta_i : \text{either fixed or random}$$

- Fixed effects: we can treat θ_i as fixed. Note: to make comparable to RE, we can use the sum-to-zero constraint, $\sum_{i=1}^{m} \theta_i = 0$, and estimate the intercept.
- We can treat θ_i as random, $\theta_i \overset{iid}{\sim} N(0, \tau^2)$.

A useful paradigm: one person's covariance structure is another person's mean structure.

Random: Consider $E(y_{ij} - \beta' x_{ij})^2 = \sigma^2 + E\theta_i^2$. (model the variance)
Fixed: $E(y_{ij} - \theta_i - \beta' x_{ij})^2 = \sigma^2$. (model the mean structure)

Guidelines for choosing fixed vs random

- Are we interested in predicting subject effects?
  - RE leverages population info – lower prediction error if treat θ_i as random.
- If the experiment were repeated, would the same subjects (i.e., groups) be used?
  - If yes, suggests FE.
- Or are the subjects a random sample from a population of interest?
  - RE
- Are there enough subjects to estimate heterogeneity?
  - E.g., if two subjects, use FE.
- Are there enough repeated measurements to estimate FE?
  - E.g., two measurements for a subject, use RE
- Do some subjects have only 1 observation and/or is there different number of samples for each subject?
  - Consider RE to leverage subjects with more information.

MIXED MODEL: RANDOM INTERCEPT

In R: use lmer

β_{0i} → subject i's intercept → $\beta_{0i} \overset{iid}{\sim} N(\mu, \tau^2)$ if using β_{0i} to sub in for $\mu + \theta_i$, this is called a HIERARCHICAL model

$$y_{ij} = \underbrace{\mu + \theta_i}_{\beta_{0i}} + x_{ij}'\beta + \epsilon_{ij} \qquad \epsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$$

- μ: overall intercept / grand mean when all $x_{ij} = 0$ (FIXED)
- θ_i: subject-specific difference from μ aka "random effect" (RANDOM)
- x_{ij}: vector of predictors
- β: vector of coefficients that do not vary between groups (FIXED)
- σ^2: ↳ residual variance, measurement error, WITHIN-GROUP variability

$\theta_i \overset{iid}{\sim} N(0, \tau^2)$ — τ^2: BETWEEN group variability in the intercepts. Aka, random effect variance

$\theta_i \perp\!\!\!\perp \epsilon_{ij}$ (↳ independent)

PROPERTIES:

- $E[y_{ij}] = \mu + x_{ij}'\beta$ — overall/average trend
- $Var[y_{ij}] = \tau^2 + \sigma^2$ — Total variability around the overall trend
- $E[y_{ij} \mid \theta_i] = \mu + \theta_i + x'\beta$ — Conditional/group-specific trend
- $Var[y_{ij} \mid \theta_i] = \sigma^2$ — Conditional/within-group residual variance

INTERPRETATIONS: (slides 21-22)

```
> fit = lmer (arm ~ age + (1|id), data = nepal)

Linear mixed model fit by REML
Random effects:
 Groups   Name        Variance Std.Dev.
 id       (Intercept) 0.78073  0.88359
 Residual             0.24907  0.49906

Fixed effects:
            Estimate Std. Error t value
(Intercept) 12.753789  0.109867  116.30
age          0.031697  0.002357   13.45
```

$\hat{\tau}^2$; just $\sqrt{\ }$ of variance aka $\sqrt{0.78073}$

We found a 0.032 cm ($CI_{95\%}$ 0.027, 0.037) increase in arm circumference per month after controlling for a child's arm circumference at birth.

We also found evidence of heterogeneity in arm circumference at birth. The estimated population-average arm circumference at birth is 12.8 cm, and the standard deviation of the random effect is 0.88 cm.

Nepalese Children: Random Intercept Model Interpretation

Consider another model with an indicator for mother's literacy.

```
> fit2 = lmer(arm~age+lit+(1|id), data = nepal)

> summary (fit2)
Random effects:
 Groups   Name        Variance Std.Dev.
 id       (Intercept) 0.74712  0.86436
 Residual             0.24824  0.49823

Fixed effects:
            Estimate Std. Error t value
(Intercept) 12.710666  0.109304  116.29
age          0.031789  0.002338   13.60
lit          0.930247  0.316301    2.94
```

We found literacy to be significantly associated with arm circumference as a main effect. Also note that there is a small decrease in the degree of heterogeneity (from 0.78 to 0.75). Therefore mother's literacy may help explain some of the observed between-children variation in arm circumference at birth. Also the intercept estimate 12.71 now corresponds to the population-average arm circumference at birth from mothers who are illiterate.

IN MATRIX FORM: (Given N observations and n groups/subjects)

$$\underset{(N\times 1)}{y} = \underset{(N\times n)}{Z}\underset{(n\times 1)}{\theta} + \underset{(N\times p)}{X}\underset{(p\times 1)}{\beta} + \underset{(N\times 1)}{\epsilon}$$

- y: vector of observed response variable values
- Z: Design matrix for random effects
- θ: vector of random effect values, each value is the random intercept for a group
- X: Design matrix for fixed effects, includes a column of 1's for the intercept & a column for each predictor var
- β: vector of fixed effect coefficients, includes intercept
- ϵ: vector of residuals

$\epsilon \sim N(0, \sigma^2 I_{n\times n})$

$\theta \sim N(0, \tau^2 I_{n\times n})$

```
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  id       (Intercept) 0.0000 τ² 0.0000
##  Residual             0.1003 σ² 0.3167
## Number of obs: 62, groups:  id, 32
```

Example:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \end{bmatrix} = \begin{array}{c} \text{subject 1} \quad \text{subject 2} \quad \text{subject 3} \cdots \\ \begin{bmatrix} 1 & 0 & 0 & \cdots \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ \vdots & \vdots & \vdots \end{bmatrix} \end{array} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \vdots \end{bmatrix} + \begin{bmatrix} 1 & x_{111} & x_{112} & \cdots \\ 1 & x_{121} & x_{122} & \cdots \\ \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} \hat{\beta} \end{bmatrix} \cdots$$

ICC (ρ) → INTRA-CLASS CORRELATION

$$\rho = Corr(y_{ij}, y_{ij'}) = \frac{\tau^2}{\tau^2 + \sigma^2} \quad \text{for all } j \neq j'$$

Measures the degree of similarity among same-group observations COMPARED TO residual error (σ^2)

AKA reproducibility.

- < 0.40 POOR
- 0.40 - 0.59 FAIR
- 0.60 - 0.74 GOOD
- 0.75 - 1.00 EXCELLENT

- when $\tau^2 \to 0$, $\rho \to 0$ (same intercept)
- when $\tau^2 \to \infty$, $\rho \to 1$ (large separation in intercepts)
- when $\sigma^2 \to 0$, $\rho \to 1$ (no measurement error)
- when $\sigma^2 \to \infty$, $\rho \to 0$ (growing measurement error)

SHRINKAGE

$$\hat{\theta}_i = \frac{(\sigma^2/\tau^2)\mu + \sum_{j=1}^{r_i} y_{ij}}{r_i + \sigma^2/\tau^2}$$

τ^2 controls shrinkage & how much info borrowed across groups

- when $\tau^2 \to 0$, $\hat{\theta}_i \to \mu$ (aka shrinks random intercepts to μ)
- when $\tau^2 \to \infty$, $\hat{\theta}_i \to \bar{y}_i$ (no shrinkage, just using raw group/subject mean estimates)
- when $\sigma^2 \to 0$, $\hat{\theta}_i \to \bar{y}_i$ " "
- when $\sigma^2 \to \infty$, $\hat{\theta}_i \to \bar{y}_i$ " "

Can also be written as:

$$\hat{\theta}_i = \frac{\left(\frac{1}{\tau^2}\right)\mu + \left(\frac{r_i}{\sigma^2}\right)\bar{y}_i}{\frac{1}{\tau^2} + \frac{r_i}{\sigma^2}}$$

→ Random effects are weighted avg of $\bar{y}_i$ and μ with the weights as inverse variances

In ICC form:

$$\hat{\theta}_i = \frac{\rho^{-1}\mu + r_i(1-\rho)^{-1}\bar{y}_i}{\rho^{-1} + r_i(1-\rho)^{-1}}$$

when $\rho \to 1$, $\hat{\theta}_i \to \bar{y}_i$

RANDOM INTERCEPT AND RANDOM SLOPE MODEL

Treats both intercept & slope as group/subject specific

$$y_{ij} = \underbrace{\beta_0 + \theta_{0i}}_{\text{subject/group specific intercept}} + (\underbrace{\beta_1 + \theta_{1i}}_{\text{subject/group specific slope}}) x_{ij} + \epsilon_{ij} \qquad \epsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$$

$$\begin{bmatrix} \theta_{0i} \\ \theta_{1i} \end{bmatrix} \overset{iid}{\sim} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_1^2 & \rho\tau_1\tau_2 \\ \rho\tau_1\tau_2 & \tau_2^2 \end{bmatrix}\right) \longrightarrow \theta_i \perp\!\!\!\perp \epsilon_{ij}$$

- τ_1^2: between-subject/group variation in baseline outcome at x var = 0
- τ_2^2: between-subject/group variation in the linear effects of x var aka variation of random slope
- ρ: correlation between subject/group-specific intercept and slope → $Corr(\theta_{0i}, \theta_{1i}) = \rho$
  - $\frac{Cov(\theta_{0i}, \theta_{1i})}{\tau_1 \tau_2} = \rho$ → $Cov(\theta_{0i}, \theta_{1i}) = \rho\tau_1\tau_2$
- σ^2: within-subject/group variation around subject/group linear growth trend aka measurement error

- Can also do a fixed effects model where $\sum_{i=1}^{n} \theta_{0i} = 0$ so the random slopes don't leverage pop. info, and total effect for each subject becomes $\hat{\mu} + \hat{\theta}_i$

BLUPs → look at slide 41 of LMM handout

Best Linear Unbiased Prediction

For the random intercept model

$$y_{ij} = \theta_i + \epsilon_{ij}, \quad \theta_i \overset{iid}{\sim} N(\mu, \tau^2) \quad \epsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$$

we wish to estimate the unobserved random variable θ_i.

We can also derive the estimators using the MVN (multivariate normal) distribution. Assume τ^2 and σ^2 are known. Then

$$\begin{bmatrix} y_i \\ \theta_i \end{bmatrix} \sim N\left(\begin{bmatrix} \mu 1_{r_i} \\ \mu \end{bmatrix}, \begin{bmatrix} \tau^2 1_{r_i}1_{r_i}' + \sigma^2 I_{r_i \times r_i} & \tau^2 1_{r_i} \\ \tau^2 1_{r_i}' & \tau^2 \end{bmatrix}\right)$$

$n \times 1$, $n = 1$

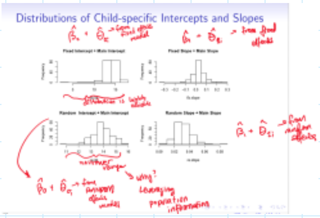
because $cov(y_{ij}, \theta_i) = cov(\theta_i + \epsilon_{ij}, \theta_i) = \tau^2$.

To make a prediction of θ_i given the data y_i, we can use the conditional aka measurement error

$\downarrow$

$Cov(\theta_{0i}, \theta_{1i}) = \rho \tau_1 \tau_2$

- Can also do a fixed effects model where $\sum_{i=1}^{n} \theta_{0i} = 0$ so the random slopes don't leverage pop. info, and total effect for each subject becomes $\hat{\beta}_1 + \hat{\theta}_{1i}$



COMPARING MODELS (with vs. without random slope)

AIC: Rule of thumb: Difference of 2 or more is substantially better

Likelihood Ratio Test: ANOVA (reduced vs. full)

] Both use MLE, so set REML = false

→ for nested models (one model contains a subset of parameters from other model)

→ However, χ_1^2 approximation is usually poor & p-value is too large bc $\tau_2^2 = 0$ is on the boundary of the parameter space

↳ prefers simpler models

In R: reduced/second model is first. anova(reduced, original)
"Acquisition has a significant effect on the z-transformed correlations when controlling for gender and scanner type."

p-values for likelihood ratio tests of random effects tend to be too large, and similarly, the AIC can select overly simple models, which is a consequence of the null hypothesis being on the boundary of the parameter space (i.e., variance = 0).

Hierarchical formulations: Check slides 76-81 in LMMs

BLUPs → look at slide 41 of LMM handout

We can also derive the estimators using the MVN distribution. (multi-variate normal) Assume τ^2 and σ^2 are known. Then

$$\begin{bmatrix} \mathbf{y}_i \\ \theta_i \end{bmatrix} \sim N\left(\begin{bmatrix} \mu \mathbf{1}_{n_i} \\ \mu \end{bmatrix}, \begin{bmatrix} \tau^2 \mathbf{1}_{n_i}\mathbf{1}_{n_i}' + \sigma^2 \mathbf{I}_{n_i \times n_i} & \tau^2 \mathbf{1}_{n_i} \\ \tau^2 \mathbf{1}_{n_i}' & \tau^2 \end{bmatrix} \right)$$

because $cov(y_{ij}, \theta_i) = cov(\theta_i + \epsilon_{ij}, \theta_i) = \tau^2$.

To make a prediction of θ_i given the data $\mathbf{y}_i$, we can use the conditional distribution of the multivariate normal density. Specifically our estimator will be

$$\hat{\theta}_i = E[\theta_i \mid \mathbf{y}_i].$$

REML

The MLE estimate of variances are biased. An alternative is restricted maximum likelihood (REML)

$$l(\sigma^2, \tau^2) - \frac{1}{2} \log |\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}|$$

to account for the degrees of freedom in the fixed effects (e.g., Ch. 6 in Searle et al. 1992, "Variance Components").

REML can be unbiased.
In the simple case of estimating σ^2 from $\mathbf{X}_i \overset{iid}{\sim} N(0, \sigma^2)$, we have

$$\hat{\sigma}^2_{MLE} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$\hat{\sigma}^2_{REML} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

Small samples: often prefer REML.
Likelihood ratio tests and AIC: use ML.

54/81 MD: Linear Mixed Models

$\epsilon_{ij}^* = \theta_i + \epsilon_{ij}$

$Cov(\epsilon_{ij}^*, \epsilon_{ij}^*) = \tau^2 + \sigma^2$

$Cov(\epsilon_{ij}^*, \epsilon_{ij'}^*) = Cov(\theta_i + \epsilon_{ij}, \theta_i + \epsilon_{ij'}) = Cov(\theta_i, \theta_i) = \tau^2$

same subject, diff measurement

$Cov(\epsilon_{ij}^*, \epsilon_{i'j}^*) = Cov(\theta_i + \epsilon_{ij}, \theta_{i'} + \epsilon_{i'j}) = 0$

diff subject, same measurement

new subject, so new realization of measurement error