Linear Mixed Effects Models

Ziwen Tang, Shanshan Song, Lucas Rohleder

Universität Bonn

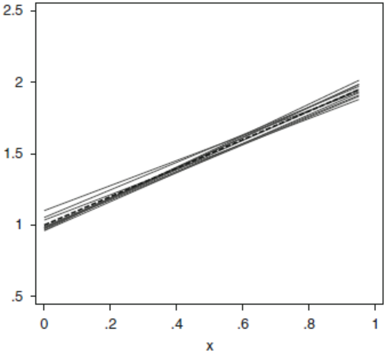
January 23, 2019

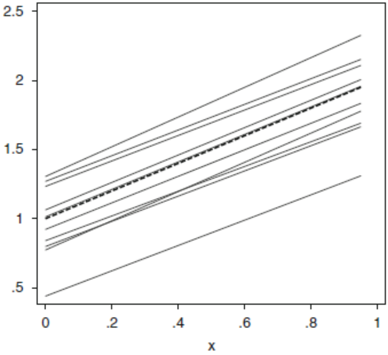
Overview

Random Effects

- Standard assumption for OLS: i.i.d errors
- Problematic with certain data structures
- Take longitudinal data: repeated measurements of same individual
  $\rightarrow$ Errors of an individual probably correlate
- Why "Random Effect"?
  $\rightarrow$ often belong to individuals who have been selected randomly from the population
- RE useful in many cases of grouped data
- Why not fixed effects?

Let's have a look





Random Intercept Model

To model this type of individual-specific heterogeneity we introduce individual-specific parameters γ_{0i}:

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \gamma_{0i} + \epsilon_{ij} \tag{1}$$

- $i = 1, ..., m$ number of individuals and $j = 1, ..., n_i$ number of repeated measurements
- $\epsilon_{ij} \sim \mathcal{N}(0, \epsilon^2)$ are i.i.d.
- β_0 is the "fixed" population intercept.
- γ_{0i} is the individual- or cluster-specific (random) deviation from the population intercept β_0
- $\beta_0 + \gamma_{0i}$ is the (random) intercept for individual i
- β_1 is a "fixed" population slope parameter that is common across individuals

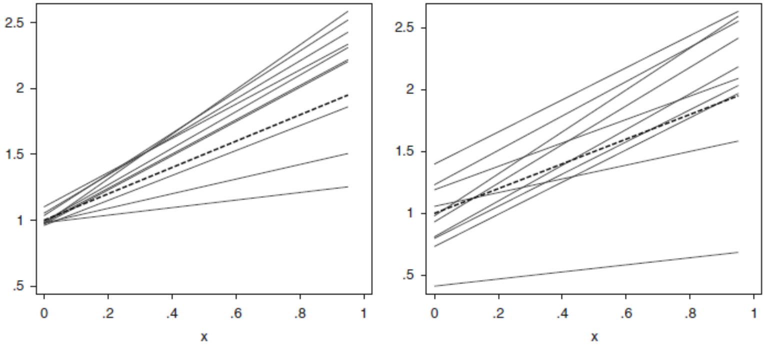
Random Intercept Model

Individuals or clusters are a random sample from a larger population, γ_{0i} are assumed to be random with

$$\gamma_{0i} \sim_{i.i.d.} \mathcal{N}(0, \tau^2) \tag{2}$$

- Mean is zero because the populationmean is already represented by the fixed effect β_0
- We assume mutual independence between the γ_{0i} and the ϵ_{ij}

Expanding the Model



Random Slope Model

In case of individual-specific slope parameters (random slopes) we can model this by

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \gamma_{1i} * x_{ij} + \epsilon_{ij} \tag{3}$$

But in most cases

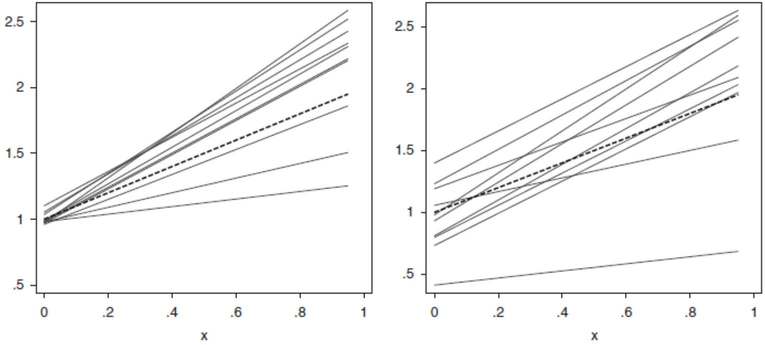
$$y_{ij} = \beta_0 + (\beta_1 + \gamma_{1i}) * x_{ij} + \gamma_{0i} + \epsilon_{ij} \tag{4}$$

Where

- β_0 is the “fixed” population intercept.
- γ_{1i} is the individual- or cluster-specific (random) deviation from the population slope β_1
- β_1 is a “fixed” population slope parameter that is common across individuals
- $\beta_1 + \gamma_{1i}$ is the (random) slope for individual i

Random Slope Model

Let's have another look



$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \gamma_{1i} * x_{ij} + \epsilon_{ij} \quad (5)$$

$$y_{ij} = (\beta_0 + \gamma_{0i}) + (\beta_1 + \gamma_{1i}) * x_{ij} + \epsilon_{ij} \quad (6)$$

Matrix Notation

In matrix notation this gives a general model:

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{Z}\boldsymbol{b} + \epsilon \tag{7}$$

With:

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix}, \boldsymbol{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_M \end{bmatrix}, \boldsymbol{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_M \end{bmatrix}, \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_M \end{bmatrix},$$

$$\boldsymbol{Z} = \begin{bmatrix} Z_1 & 0 & \cdots & 0 \\ 0 & Z_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & Z_M \end{bmatrix}.$$

Estimation

In a general linear mixed effect model,we have

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{Z}\gamma + \epsilon \tag{8}$$

Assuming$\gamma \sim \mathcal{N}(0, \epsilon^2 D)$ and$\epsilon \sim \mathcal{N}(0, \epsilon^2)$, we get

$$var(y) = var(Z\gamma) + var(\epsilon) = \sigma^2 ZDZ^T + \sigma^2 I \tag{9}$$

the unconditional distribution of y as

$$\boldsymbol{y} \sim N(\boldsymbol{X}\beta, \sigma^2(I + ZDZ^T)) \tag{10}$$

For simplicity,$\sigma^2(I + ZDZ^T)$ will be replaced by $H(\alpha)$,

$$\boldsymbol{y} \sim \mathcal{N}\left(\boldsymbol{X}\beta, \boldsymbol{H}(\alpha) = \begin{bmatrix} H_1(\alpha) & 0 & \cdots & 0 \\ 0 & H_2(\alpha) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & H_M(\alpha) \end{bmatrix}\right), \tag{11}$$

Estimation

- Q1: What do we need to estimate?
- A1:we should estimate β and $H(\alpha)$

Estimation

Q2:how do we estimate β and $H(\alpha)$
A2:??......how about assuming $H(\alpha)$ is known

Estimation- Known Variance-covariance Parameters

Assuming that $H(\alpha)$ are known, we can just maximize the joint log-likelihood of y, γ with respect to β and γ and get

$$\widehat{\beta} = \boldsymbol{X}^T \boldsymbol{H}^{-1} \boldsymbol{X}^{-1} \boldsymbol{X}^T \boldsymbol{H}^{-1} \boldsymbol{y} \tag{12}$$

$$\widehat{\gamma} = DZ^T H^{-1} (\boldsymbol{y} - X\widehat{\beta}) \tag{13}$$

This implies that $\widehat{\beta}$ is a weighted least squares estimator with the inverse of the $H(\alpha) = \sigma^2(I + ZDZ^T)$ of β as the weight matrix.

Estimation- Unknown Variance-covariance Parameters

1. Maximum likelihood (ML) approach
2. Restricted maximum likelihood(REML) approach
3. Which one is better? why?

Estimation-Unknown Variance-covariance Parameters

Based on the likelihood of $y \sim \mathcal{N}(X\beta, H(\alpha))$, the corresponding log-likelihood is

$$logL(\beta, \alpha | y, X) = -\frac{1}{2}[\log \det H(\alpha) + (y - X\beta)^T H(\alpha)^{-1}(y - X\beta)] \quad (14)$$

ML: with respect to β(while holding α fixed),we have

$$\widehat{\beta}(\alpha) = (X^T H(\alpha)^{-1} X)^{-1} X^T H(\alpha)^{-1} y \quad (15)$$

Inserting $\widehat{\beta}(\alpha)$ in $logL(\beta, \alpha | y, X)$ reults in the profile log-likelihood

$$\mathcal{L}_{ML}(\alpha) = -\frac{1}{2}[\log \det H(\alpha) + (y - X\widehat{\beta}(\alpha)^T H(\alpha)^{-1}(y - X\widehat{\beta}(\alpha)))] \quad (16)$$

Finally we can get ML estimator $\widehat{\alpha}_{ML}$.

Estimation- Unknown Variance-covariance Parameters

Based on the likelihood of $y \sim \mathcal{N}(X\beta, H(\alpha))$, the corresponding log-likelihood is

$$logL(\beta, \alpha|y, X) = -\frac{1}{2}[\log\ \det H(\alpha) + (y - X\beta)^T H(\alpha)^{-1}(y - X\beta)] \quad (17)$$

REML:Integrating out β from the likelihood,that restricted log-likelihood is

$$\mathcal{L}_R(\alpha) = log(\int L(\beta, \alpha)d\beta) \quad (18)$$

$$\begin{aligned}\mathcal{L}_R(\alpha|A^T y) = -\frac{1}{2}(&\log\ \det H(\alpha) + \log\ \det X^T H(\alpha)^{-1}X \\ &+ (y - X\widehat{\beta})^T H(\alpha)^{-1}(y - X\widehat{\beta}))\end{aligned} \quad (19)$$

where $\widehat{\beta}(\alpha) = (X^T H(\alpha)^{-1}X)^{-1}X^T H(\alpha)^{-1}y$.
Finally we can get REML estimator $\widehat{\alpha}_{REML}$.

If vector a is orthogonal to all columns of X, i.e., $a^T X = 0$, then $a^T y$ is known as an error contrast. The distribution of y is

$$\boldsymbol{y} \sim \mathcal{N}\left(\boldsymbol{X}\beta, \boldsymbol{H}(\alpha) = \begin{bmatrix} H_1(\alpha) & 0 & \cdots & 0 \\ 0 & H_2(\alpha) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & H_M(\alpha) \end{bmatrix}\right), \tag{20}$$

Define $A = \left[a_1 a_2 ... a_{(N-k)}\right]$. $A^T X = 0$ and $E(A^T y) = 0$. The error contrast vector

$$A^T y = A^T X\beta + \epsilon = A^T \epsilon \sim \mathcal{N}(0, A^T HA) \tag{21}$$

is free of β.

Which one is better ?

- we prefer $\widehat{\alpha}_{REML}$ in LMMs as an estimator of α.The ML estimator for σ^2 is biased.

In the classical linear model,the estimator

$$\hat{\sigma}^2_{unbiased} = \frac{1}{N-K}(y - X\hat{\beta})^T(y - X\hat{\beta}) = E(\hat{\sigma}^2_{REML}) \tag{22}$$

$$E(\hat{\sigma}^2_{ML}) = \frac{1}{N}(y - X\hat{\beta})^T(y - X\hat{\beta}) = \frac{N-K}{N}\hat{\sigma}^2_{unbiased} < \hat{\sigma}^2_{unbiased} \tag{23}$$

In LMMs,the reason of biased maximum likelihood estimator is that we neglect the loss of **degree of freedom**(DoF) for estimating β.
The intuition behind ReML is to maximize a modified likelihood that is free of mean components instead of the original likelihood as in ML.

Inference for linear mixed model

Recalling from the estimation part,

$$y_i \sim \mathcal{N}(X_i\beta, \sigma^2(I + Z_i D Z_i^T)) \tag{24}$$

we want to test mean structure β and variance components α(variance, covariance elements in D and σ^2) with respect to y_i

- Testing for fixed effects
  $\rightarrow$ Approximate F test
  $\rightarrow$ Likelihood Ratio Test(LRT)
- Testing for variance components
  $\rightarrow$ Likelihood Ratio Test(LRT)
- Two popular alternatives
  $\rightarrow$ Parametric Bootstrap
  $\rightarrow$ Kenward and Roger

Testing for fixed effects

Recalling from estimation chapter

$$\widehat{\beta} = (X^T H^{-1} X)^{-1} X^T H^{-1} y \tag{25}$$

Approximate F test

- for any known matrix L,

$$H_0 : L\beta = 0, versus \quad H_A : L\beta \neq 0 \tag{26}$$

- it is now based on an F-approximation to the distribution of

$$F = \frac{(\widehat{\beta} - \beta)^T L^T [L(\sum_{i=1}^{N} X_i^T H_i^{-1}(\widehat{\alpha}) X_i)^{-1} L^T]^{-1} L(\widehat{\beta} - \beta)}{rank(L)} \tag{27}$$

- F has asymptotically a $\frac{1}{d}{\chi_d}^2$ distribution
- Problems with F test: degree of freedom; not necessarily F-distributed

Testing for fixed effects

Likelihood Ratio Test

- conditions: comparison of **nested** models with **different mean structure**
- $H_0 : \boldsymbol{\beta} \in \Theta_{\beta,0}$, for some subspace $\Theta_{\beta,0}$ of the parameter space Θ_β of the fixed effects $\boldsymbol{\beta}$.
- LRT is defined as

$$-2\ln\lambda_N = -2\ln\left[\frac{L_{ML}(\widehat{\boldsymbol{\theta}}_{ML,0})}{L_{ML}(\widehat{\boldsymbol{\theta}}_{ML})}\right], \tag{28}$$

  where $\widehat{\boldsymbol{\theta}}_{ML,0}$ and $\widehat{\boldsymbol{\theta}}_{ML}$ are the ML estimates obtained from maximizing L_{ML} over $\Theta_{\beta,0}$ and Θ_β, respectively.
- $-2\ln\lambda_N$: asymptotically a chi-squared distribution with certain degrees of freedom
- LRT is not valid with REML estimation

Testing for variance components

Likelihood Ratio Test

- conditions: comparison of **nested** models with **different covariance structure**
- H_0: $\boldsymbol{\alpha}$ in $\Theta_{\alpha,0}$, for some subspace $\Theta_{\alpha,0}$ of the parameter space Θ_α of the variance components $\boldsymbol{\alpha}$.
- LRT is again defined as

$$-2\ln\lambda_N = -2\ln\left[\frac{L_{REML}(\widehat{\boldsymbol{\theta}}_{REML,0})}{L_{REML}(\widehat{\boldsymbol{\theta}}_{REML})}\right], \tag{29}$$

  where $\widehat{\boldsymbol{\theta}}_{REML,0}$ and $\widehat{\boldsymbol{\theta}}_{REML}$ are now the maximum likelihood estimates obtained from maximizing L_{REML} over $\Theta_{\alpha,0}$ and Θ_α, respectively.
- $-2\ln\lambda_N$: asymptotically a chi-squared distribution with certain degrees of freedom(only if not on the boundary!)
- using REML estimation to get more accurate estimator.

Two popular alternatives

- parametric bootstrap
  $\rightarrow$ P-values generated by LRT for fixed effects tend to be too small
  $\rightarrow$ P-values generated by LRT for variance components tend to be larger(eg. $H_0 : \hat{\sigma}^2 = 0$)
  $\rightarrow$ using parametric bootstrap to find more accurate p-values for LRT
- Kenward and Roger
  $\rightarrow$ Improving the small sample properties by approximating the distribution of F by an $F_{d,m}$ distribution
  $\rightarrow$ efficiently calculating the denominator degrees of freedom

Monte Carlo Simulation

- Imagine we want to simulate the effect of a Physics test score on a Math test score
- We expect correlation between the subjects scores within the same school $\rightarrow$ each school has an effect on its students
- We want to account for this by using random effects

Naive Random Intercept Model

$$mathscr_{ij} = \beta_0 + \beta_1 physcr_{ij} + \gamma_{0i} + \epsilon_{ij} \tag{30}$$

Here:

- "individuals"/clusters i are the schools
- repeated measurements j are students physics test scores (of school i)
- γ_{0i} school random effect

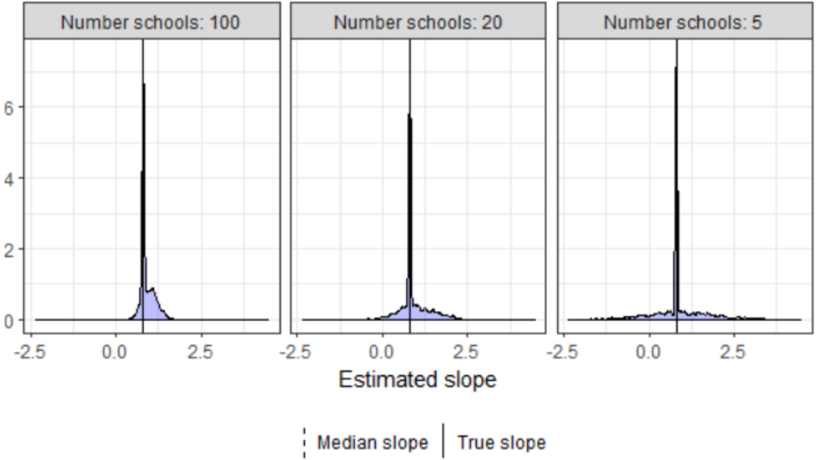
Simulation Parameters

Remember: $mathscr_{ij} = \beta_0 + \beta_1 physcr_{ij} + \gamma_{0i} + \epsilon_{ij}$
We set:

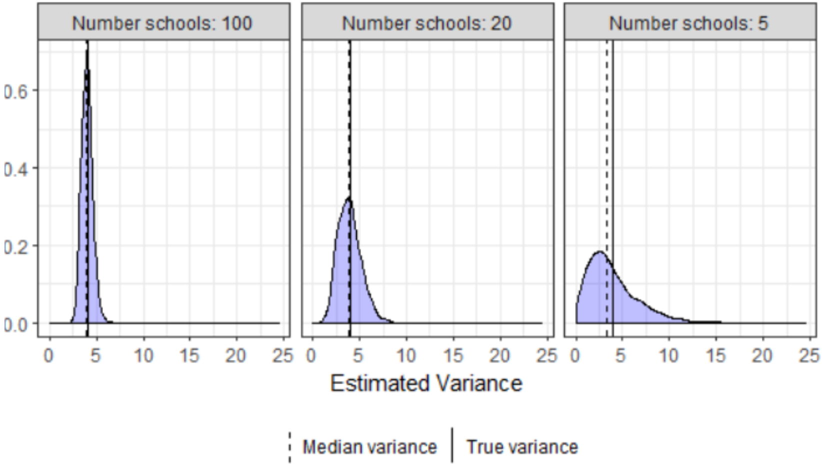
- $\beta_0 = 1$
- Slope of fixed effect $\beta_1 = 0.8$
- number of students per school $j = 1, ..., 30$
- random effect $\sim \mathcal{N}(0, 4)$
- $\epsilon_{ij} \sim \mathcal{N}(0, 1)$

Plug everything in the model an simulate data 1000 times for 5, 30, 100 schools → estimate fixed effect parameter and random effect variance each time

Estimated Fixed Effects Slope



Estimated Random Effects Variance



Thank you for your attention!