Linear Mixed-Effects Regression

Nathaniel E. Helwig

Assistant Professor of Psychology and Statistics
University of Minnesota (Twin Cities)

Updated 04-Jan-2017

Copyright

Outline of Notes

1) Correlated Data:

- Overview of problem
- Motivating Example
- Modeling correlated data

2) One-Way RM-ANOVA:

- Model Form & Assumptions
- Estimation & Inference
- Example: Grocery Prices

3) Linear Mixed-Effects Model:

- Random Intercept Model
- Random Intercepts & Slopes
- General Framework
- Covariance Structures
- Estimation & Inference
- Example: TIMSS Data

Correlated Data

What are Correlated Data?

So far we have assumed that observations are independent.

- Regression: $(y_i, \mathbf{x}_i)$ are independent for all n
- ANOVA: y_i are independent within and between groups

In a Repeated Measures (RM) design, observations are observed from the same subject at multiple occasions.

- Regression: multiple y_i from same subject
- ANOVA: same subject in multiple treatment cells

RM data are one type of correlated data, but other types exist.

Why are Correlated Data an Issue?

Thus far, all of our inferential procedures have required independence.

- Regression:
  $\hat{\mathbf{b}} \sim \mathrm{N}(\mathbf{b}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$ requires the assumption $(\mathbf{y}|\mathbf{X}) \sim \mathrm{N}(\mathbf{X}\mathbf{b}, \sigma^2\mathbf{I}_n)$
  where $\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$
- ANOVA:
  $\hat{L} \sim \mathrm{N}(L, \sigma^2 \sum_{j=1}^{a} c_j^2/n_j)$ requires the assumption $y_{ij} \overset{\mathrm{iid}}{\sim} \mathrm{N}(\mu_j, \sigma^2)$
  where $\hat{L} = \sum_{j=1}^{a} c_j \hat{\mu}_j$

Correlated data are (by definition) correlated.

- Violates the independence assumption
- Need to account for correlation for valid inference

TIMSS Data from 1997

Trends in International Mathematics and Science Study (TIMSS)[1]

- Ongoing study assessing STEM education around the world
- We will analyze data from 3rd and 4th grade students
- We have $n_T = 7,097$ students nested within $n = 146$ schools

```
> timss = read.table(paste(datapath,"timss1997.txt",sep=""),header=TRUE,
+                    colClasses=c(rep("factor",4),rep("numeric",3)))
> head(timss)
  idschool idstudent grade gender science  math hoursTV
1       10    100101     3   girl   146.7 137.0       3
2       10    100103     3   girl   148.8 145.3       2
3       10    100107     3   girl   150.0 152.3       4
4       10    100108     3   girl   146.9 144.3       3
5       10    100109     3    boy   144.3 140.3       3
6       10    100110     3    boy   156.5 159.2       2
```

[1] https://nces.ed.gov/TIMSS/

Issues with Modeling TIMSS Data

Data are collected from students nested within schools.

Nesting typically introduces correlation into data at level-1

- Students are level-1 and schools are level-2
- Dependence/correlation between students from same school

We need to account for this dependence when we model the data.

Fixed versus Random Effects

Thus far, we have assumed that parameters are unknown constants.

- Regression: $\mathbf{b}$ is some unknown (constant) coefficient vector
- ANOVA: μ_j are some unknown (constant) means
- These are referred to as fixed effects

Unlike fixed effects, random effects are NOT unknown constants

- Random effects are random variables in the population
- Typically assume that random effects are zero-mean Gaussian
- Typically want to estimate the variance parameter(s)

Models with fixed and random effects are called mixed-effects models.

Modeling Correlated Data with Random Effects

To model correlated data, we include random effects in the model.

- Random effects relate to assumed correlation structure for data
- Including different combinations of random effects can account for different correlation structures present in the data

Goal is to estimate fixed effects parameters (e.g., $\hat{\mathbf{b}}$) and random effects variance parameters.

- Variance parameters are of interest, because they relate to model covariance structure
- Could also estimate the random effect realizations (BLUPs)

One-Way Repeated Measures ANOVA

Model Form

The One-Way Repeated Measures ANOVA model has the form

$$y_{ij} = \rho_i + \mu_j + e_{ij}$$

for $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, a\}$ where

- $y_{ij} \in \mathbb{R}$ is the response for i-th subject in j-th factor level
- $\mu_j \in \mathbb{R}$ is the fixed effect for the j-th factor level
- $\rho_i \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma_\rho^2)$ is the random effect for the i-th subject
- $e_{ij} \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma_e^2)$ is a Gaussian error term
- n is number of subjects and a is number of factor levels

Note: each subject is observed a times (once in each factor level).

Model Assumptions

The fundamental assumptions of the one-way RM ANOVA model are:

1. x_{ij} and y_i are observed random variables (known constants)
2. $\rho_i \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma_\rho^2)$ is an unobserved random variable
3. $e_{ij} \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma_e^2)$ is an unobserved random variable
4. ρ_i and e_{ij} are independent of one another
5. $\mu_1, \ldots, \mu_a$ are unknown constants
6. $y_{ij} \sim \mathrm{N}(\mu_j, \sigma_Y^2)$ where $\sigma_Y^2 = \sigma_\rho^2 + \sigma_e^2$ is the total variance of Y

Using effect coding, $\mu_j = \mu + \alpha_j$ with $\sum_{j=1}^{a} \alpha_j = 0$

Assumed Covariance Structure (same subject)

For two observations from the same subject y_{ij} and y_{ik} we have

$$\begin{aligned} Cov(y_{ij}, y_{ik}) &= E[(y_{ij} - \mu_j)(y_{ik} - \mu_k)] \\ &= E[(\rho_i + e_{ij})(\rho_i + e_{ik})] \\ &= E[\rho_i^2 + \rho_i(e_{ij} + e_{ik}) + e_{ij}e_{ik}] \\ &= E[\rho_i^2] = \sigma_\rho^2 \end{aligned}$$

given that $E(\rho_i e_{ij}) = E(\rho_i e_{ik}) = E(e_{ij}e_{ik}) = 0$ by model assumptions.

Assumed Covariance Structure (different subjects)

For two observations from different subjects y_{hj} and y_{ik} we have

$$\begin{aligned} Cov(y_{hj}, y_{ik}) &= E[(y_{hj} - \mu_j)(y_{ik} - \mu_k)] \\ &= E[(\rho_h + e_{hj})(\rho_i + e_{ik})] \\ &= E[\rho_h\rho_i + \rho_h e_{ik} + \rho_i e_{hj} + e_{hj}e_{ik}] \\ &= 0 \end{aligned}$$

given that $E(\rho_h\rho_i) = E(\rho_h e_{ik}) = E(\rho_i e_{hj}) = E(e_{hj}e_{ik}) = 0$ due to the model assumptions.

Assumed Covariance Structure (general form)

The covariance between any two observations is

$$Cov(y_{hj}, y_{ik}) = \begin{cases} \sigma_\rho^2 = \omega\sigma_Y^2 & \text{if } h = i \text{ and } j \neq k \\ 0 & \text{if } h \neq i \end{cases}$$

where $\omega = \sigma_\rho^2/\sigma_Y^2$ is the correlation between any two repeated measurements from the same subject.

ω is referred to as the intra-class correlation coefficient (ICC).

Compound Symmetry

Assumptions imply covariance pattern known as compound symmetry

- All repeated measurements have same variance
- All pairs of repeated measurements have same covariance

With $a = 4$ repeated measurements the covariance matrix is

$$Cov(\mathbf{y}_i) = \begin{pmatrix} \sigma_Y^2 & \omega\sigma_Y^2 & \omega\sigma_Y^2 & \omega\sigma_Y^2 \\ \omega\sigma_Y^2 & \sigma_Y^2 & \omega\sigma_Y^2 & \omega\sigma_Y^2 \\ \omega\sigma_Y^2 & \omega\sigma_Y^2 & \sigma_Y^2 & \omega\sigma_Y^2 \\ \omega\sigma_Y^2 & \omega\sigma_Y^2 & \omega\sigma_Y^2 & \sigma_Y^2 \end{pmatrix} = \sigma_Y^2 \begin{pmatrix} 1 & \omega & \omega & \omega \\ \omega & 1 & \omega & \omega \\ \omega & \omega & 1 & \omega \\ \omega & \omega & \omega & 1 \end{pmatrix}$$

where $\mathbf{y}_i = (y_{i1}, y_{i2}, y_{i3}, y_{i4})$ is the i-th subject's vector of data.

Note on Compound Symmetry and Sphericity

Assumption of compound symmetry is more strict than we need.

For valid inference, we need the homogeneity of treatment-difference variances (HOTDV) assumption to hold, which states that

$$Var(y_{ij} - y_{ik}) = \theta$$

for any $j \neq k$, where θ is some constant.

- This is the sphericity assumption for covariance matrix

If compound symmetry is met, sphericity assumption will also be met.

$$\begin{aligned} Var(y_{ij} - y_{ik}) &= Var(y_{ij}) + Var(y_{ik}) - 2Cov(y_{ij}, y_{ik}) \\ &= 2\sigma_Y^2 - 2\sigma_\rho^2 = 2\sigma_e^2 \end{aligned}$$

Ordinary Least Squares Estimation

Parameter estimates are analogue of balanced two-way ANOVA:

$$\hat{\mu} = \frac{1}{na} \sum_{j=1}^{a} \sum_{i=1}^{n} y_{ij} = \bar{y}_{..}$$
$$\hat{\rho}_i = \left(\frac{1}{a} \sum_{j=1}^{a} y_{ij} \right) - \hat{\mu} = \bar{y}_{i.} - \bar{y}_{..}$$
$$\hat{\alpha}_j = \left(\frac{1}{n} \sum_{i=1}^{n} y_{ij} \right) - \hat{\mu} = \bar{y}_{.j} - \bar{y}_{..}$$

which implies that the fitted values have the form

$$\begin{aligned} \hat{y}_{ij} &= \hat{\mu} + \hat{\rho}_i + \hat{\alpha}_j \\ &= \bar{y}_{i.} + \bar{y}_{.j} - \bar{y}_{..} \end{aligned}$$

so that the residuals have the form $\hat{e}_{ij} = y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}$

Sums-of-Squares and Degrees-of-Freedom

The relevant sums-of-squares are given by

$$\begin{aligned} SST &= \textstyle\sum_{j=1}^{a} \sum_{i=1}^{n} (y_{ij} - \bar{y}_{\cdot\cdot})^2 \\ SSS &= a \textstyle\sum_{i=1}^{n} \hat{\rho}_i^2 \\ SSA &= n \textstyle\sum_{j=1}^{a} \hat{\alpha}_j^2 \\ SSE &= \textstyle\sum_{j=1}^{a} \sum_{i=1}^{n} \hat{e}_{ij}^2 \end{aligned}$$

where SSS = sum-of-squares for subjects; corresponding dfs are

$$\begin{aligned} df_{SST} &= na - 1 \\ df_{SSS} &= n - 1 \\ df_{SSA} &= a - 1 \\ df_{SSE} &= (n-1)(a-1) \end{aligned}$$

Extended ANOVA Table and F Tests

We typically organize the SS information into an ANOVA table:

Source	SS	df	MS	F	p-value
SSS	$a\sum_{i=1}^{n}\hat{\rho}_i^2$	$n-1$	MSS	F_s^*	p_s^*
SSA	$n\sum_{j=1}^{a}\hat{\alpha}_j^2$	$a-1$	MSA	F_a^*	p_a^*
SSE	$\sum_{j=1}^{a}\sum_{i=1}^{n}(y_{ij}-\hat{y}_{jk})^2$	$(n-1)(a-1)$	MSE		
SST	$\sum_{j=1}^{a}\sum_{i=1}^{n}(y_{ij}-\bar{y}_{\cdot\cdot})^2$	$na-1$			

$MSS = \frac{SSS}{n-1}$, $MSA = \frac{SSA}{a-1}$, $MSE = \frac{SSE}{(n-1)(a-1)}$

$F_s^* = \frac{MSS}{MSE} \sim F_{n-1,(n-1)(a-1)}$ and $p_s^* = P(F_{n-1,(n-1)(a-1)} > F_s^*)$,

$F_a^* = \frac{MSA}{MSE} \sim F_{a-1,(n-1)(a-1)}$ and $p_a^* = P(F_{a-1,(n-1)(a-1)} > F_a^*)$,

F_s^* statistic and p_s^*-value are testing $H_0 : \sigma_\rho^2 = 0$ versus $H_1 : \sigma_\rho^2 > 0$

- Testing random effect of subject, but not a valid test

F_a^* statistic and p_a^*-value are testing $H_0 : \alpha_j = 0 \; \forall j$ versus $H_1 : (\exists j \in \{1, \ldots, a\})(\alpha_j \neq 0)$

- Testing main effect of treatment factor

Expectations of Mean-Squares

The MSE is an unbiased estimator of σ_e^2, i.e., $E(MSE) = \sigma_e^2$.

The MSS has expectation $E(MSS) = \sigma_e^2 + a\sigma_\rho^2$

- If $MSS > MSE$, can use $\hat{\sigma}_\rho^2 = (MSS - MSE)/a$

The MSA has expectation $E(MSA) = \sigma_e^2 + \frac{n\sum_{j=1}^{a}\alpha_j^2}{a-1}$

Quantifying Violations of Sphericity

Valid inference requires sphericity assumption to be met.

- If sphericity assumption is violated, our F test is too liberal

George Box (1954) proposed a measure of sphericity

$$\epsilon = \frac{(\sum_{j=1}^{a} \lambda_j)^2}{(a-1)\sum_{j=1}^{a} \lambda_j^2}$$

where λ_j are the eigenvalues of $a \times a$ population covariance matrix.

- $\frac{1}{a-1} \leq \epsilon \leq 1$ such that $\epsilon = 1$ denotes perfect sphericity

If sphericity is violated, then $F_a^* \sim F_{\epsilon(a-1),\epsilon(a-1)(n-1)}$

Geisser-Greenhouse $\hat{\epsilon}$ Adjustment

Let $\mathbf{Y} = \{y_{ij}\}_{n \times a}$ denote the data matrix

- $\mathbf{Z} = \mathbf{C}_n\mathbf{Y}$ where $\mathbf{C}_n = \mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n'$ denotes $n \times n$ centering matrix
- $\hat{\boldsymbol{\Sigma}} = \frac{1}{n-1}\mathbf{Z}'\mathbf{Z}$ is sample covariance matrix
- $\hat{\boldsymbol{\Sigma}}_c = \mathbf{C}_a\hat{\boldsymbol{\Sigma}}\mathbf{C}_a$ is doubled-centered covariance matrix

The Geisser-Greenhouse $\hat{\epsilon}$ estimate is defined

$$\hat{\epsilon} = \frac{(\sum_{j=1}^{a} \hat{\lambda}_j)^2}{(a-1)\sum_{j=1}^{a} \hat{\lambda}_j^2}$$

where $\hat{\lambda}_j$ are eigenvalues of $\hat{\boldsymbol{\Sigma}}_c$.

Note that $\hat{\epsilon}$ is the empirical version of ϵ using $\hat{\boldsymbol{\Sigma}}_c$ to estimate $\boldsymbol{\Sigma}$.

Huynh-Feldt $\tilde{\epsilon}$ Adjustment

GG adjustment is too conservative when ϵ is close to 1.

Huynh and Feldt provide a corrected estimate of ϵ

$$\tilde{\epsilon} = \frac{n(a-1)\hat{\epsilon} - 2}{(a-1)[n-1-(a-1)\hat{\epsilon}]}$$

where $\hat{\epsilon}$ is the GG estimate of ϵ... note that $\tilde{\epsilon} \geq \hat{\epsilon}$.

HF adjustment is too liberal when ϵ is close to 1.

An R Function for One-Way RM ANOVA

```
aov1rm <- function(X){
  X = as.matrix(X)
  n = nrow(X)
  a = ncol(X)
  mu = mean(X)
  rhos = rowMeans(X) - mu
  alphas = colMeans(X) - mu
  ssa = n*sum(alphas^2)
  msa = ssa / (a - 1)
  mss = a*sum(rhos^2) / (n - 1)
  ehat = X - ( mu + matrix(rhos,n,a) + matrix(alphas,n,a,byrow=TRUE) )
  sse = sum(ehat^2)
  mse = sse / ( (a-1)*(n-1) )
  Fstat = msa / mse
  pval = 1 - pf(Fstat,a-1,(a-1)*(n-1))
  Cmat = cov(X)
  Jmat = diag(a) - matrix(1/a,a,a)
  Dmat = Jmat%*%Cmat%*%Jmat
  gg = ( sum(diag(Dmat))^2 ) / ( (a-1)*sum(Dmat^2) )
  hf = (n*(a-1)*gg - 2) / ( (a-1)*(n - 1 - (a-1)*gg) )
  pgg = 1 - pf(Fstat,gg*(a-1),gg*(a-1)*(n-1))
  phf = 1 - pf(Fstat,hf*(a-1),hf*(a-1)*(n-1))
  list(mu = mu, alphas = alphas, rhos = rhos,
       Fstat = c(F=Fstat,df1=(a-1),df2=(a-1)*(n-1)),
       pvals = c(pGG=pgg,pHF=phf,p=pval),
       epsilon = c(GG=gg,HF=hf),
       vcomps = c(sigsq.e=mse, sigsq.rho=((mss-mse)/a)) )
}
```

Multiple Comparisons

Can use same approaches as before (e.g., Tukey, Bonferroni, Scheffé).

MCs are extremely sensitive to violations of the HOTDV assumption.

$\hat{L} \sim \mathrm{N}(L, \frac{\sigma^2}{n} \sum_{j=1}^{a} c_j^2)$ where the MSE is used to estimate σ^2

- $\hat{L} = \sum_{j=1}^{a} c_j \hat{\mu}_j$ is a linear combination of factor means
- MSE is error estimate using all treatment groups
- If data violate HOTDV, then MSE will be a bad estimate of the variance for certain linear combinations

Grocery Example: Data Description

Grocery prices data from William B. King[2]

```
> groceries = read.table("~/Desktop/groceries.txt", header=TRUE)
> groceries
             subject storeA storeB storeC storeD
1            lettuce   1.17   1.78   1.29   1.29
2           potatoes   1.77   1.98   1.99   1.99
3               milk   1.49   1.69   1.79   1.59
4               eggs   0.65   0.99   0.69   1.09
5              bread   1.58   1.70   1.89   1.89
6             cereal   3.13   3.15   2.99   3.09
7        ground.beef   2.09   1.88   2.09   2.49
8        tomato.soup   0.62   0.65   0.65   0.69
9  laundry.detergent   5.89   5.99   5.99   6.99
10           aspirin   4.46   4.84   4.99   5.15
```

[2] http://ww2.coastal.edu/kingw/statistics/R-tutorials/

Grocery Example: Data Long Format

For many examples we will need data in “long format”

```
> grocery = data.frame(price = as.numeric(unlist(groceries[,2:5])),
+                      item = rep(groceries$subject,4),
+                      store = rep(LETTERS[1:4],each=10))
> grocery[1:12,]
   price              item store
1   1.17           lettuce     A
2   1.77          potatoes     A
3   1.49              milk     A
4   0.65              eggs     A
5   1.58             bread     A
6   3.13            cereal     A
7   2.09       ground.beef     A
8   0.62       tomato.soup     A
9   5.89 laundry.detergent     A
10  4.46           aspirin     A
11  1.78           lettuce     B
12  1.98          potatoes     B
```

Grocery Example: Check and Set Contrasts

```
> contrasts(grocery$store)
  B C D
A 0 0 0
B 1 0 0
C 0 1 0
D 0 0 1
> contrasts(grocery$store) <- contr.sum(4)
> contrasts(grocery$store)
  [,1] [,2] [,3]
A    1    0    0
B    0    1    0
C    0    0    1
D   -1   -1   -1
```

Grocery Example: `aov` with Fixed-Effects Syntax

```
> amod = aov(price ~ store + item, data=grocery)
> summary(amod)
            Df Sum Sq Mean Sq F value Pr(>F)
store        3   0.59   0.195   4.344 0.0127 *
item         9 115.19  12.799 284.722 <2e-16 ***
Residuals   27   1.21   0.045
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Grocery Example: `aov` with Mixed-Effects Syntax

```
> amod = aov(price ~ store + Error(item/store), data=grocery)
> summary(amod)

Error: item
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals  9  115.2    12.8

Error: item:store
          Df Sum Sq Mean Sq F value Pr(>F)
store      3 0.5859 0.19529   4.344 0.0127 *
Residuals 27 1.2137 0.04495
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Grocery Example: lmer Syntax (ML solution)

```
> library(lme4)
> nmod = lmer(price ~ 1 + (1 | item), data=grocery, REML=F)
> amod = lmer(price ~ store + (1 | item), data=grocery, REML=F)
> anova(amod,nmod)
Data: grocery
Models:
nmod: price ~ 1 + (1 | item)
amod: price ~ store + (1 | item)
     Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
nmod  3 59.546 64.613 -26.773   53.546
amod  6 53.731 63.864 -20.865   41.731 11.816      3   0.008042 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Grocery Example: lm Syntax (multivariate solution)

```
> library(car)
> lmod = lm(as.matrix(groceries[,2:5]) ~ 1)
> store = LETTERS[1:4]
> almod = Anova(lmod, type="III",
                idata=data.frame(store=store), idesign=~store)
> summary(almod,multivariate=FALSE)$univariate
                 SS num Df Error SS den Df       F   Pr(>F)
(Intercept) 240.688      1  115.193      9 18.8049 0.001887 **
store         0.586      3    1.214     27  4.3442 0.012730 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> summary(almod,multivariate=FALSE)$pval.adj
        GG eps Pr(>F[GG])    HF eps Pr(>F[HF])
store 0.639109  0.0309308 0.8082292 0.02033859
attr(,"na.action")
(Intercept)
          1
attr(,"class")
[1] "omit"
```

Grocery Example: `aov1rm` Syntax

```
> amod = aov1rm(groceries[,2:5])
> amod$Fstat
        F       df1       df2
 4.344209  3.000000 27.000000
> amod$pvals
       pGG        pHF          p
0.03093080 0.02033859 0.01273035
> amod$eps
       GG        HF
0.6391090 0.8082292
```

Linear Mixed-Effects Model

Random Intercept Model Form

A random intercept regression model has the form

$$y_{ij} = b_0 + b_1 x_{ij} + v_i + e_{ij}$$

for $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, m_i\}$ where

- $y_{ij} \in \mathbb{R}$ is the response for j-th measurement of i-th subject
- $b_0 \in \mathbb{R}$ is the fixed intercept for the regression model
- $b_1 \in \mathbb{R}$ is the fixed slope for the regression model
- $x_{ij} \in \mathbb{R}$ is the predictor for j-th measurement of i-th subject
- $v_i \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma_v^2)$ is the random intercept for the i-th subject
- $e_{ij} \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma_e^2)$ is a Gaussian error term

Random Intercept Model Assumptions

The fundamental assumptions of the RI model are:

1. Relationship between X and Y is linear
2. x_{ij} and y_{ij} are observed random variables (known constants)
3. $v_i \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma_v^2)$ is an unobserved random variable
4. $e_{ij} \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma_e^2)$ is an unobserved random variable
5. v_i and e_{ij} are independent of one another
6. b_0 and b_1 are unknown constants
7. $(y_{ij}|x_{ij}) \sim \mathrm{N}(b_0 + b_1 x_{ij}, \sigma_Y^2)$ where $\sigma_Y^2 = \sigma_v^2 + \sigma_e^2$

Note: v_i allows each subject to have unique regression intercept.

Assumed Covariance Structure

The (conditional) covariance between any two observations is

$$Cov(y_{hj}, y_{ik}) = \begin{cases} \sigma_v^2 = \omega\sigma_Y^2 & \text{if } h = i \text{ and } j \neq k \\ 0 & \text{if } h \neq i \end{cases}$$

where $\omega = \sigma_v^2/\sigma_Y^2$ is the correlation between any two repeated measurements from the same subject.

- If $h = i$, then $Cov(y_{ij}, y_{ik}) = E[(v_i + e_{ij})(v_i + e_{ik})] = E(v_i^2) = \sigma_v^2$
- If $h \neq i$, then $Cov(y_{hj}, y_{ik}) = E[(v_h + e_{hj})(v_i + e_{ik})] = 0$

Note: this covariance is conditioned on fixed effects x_{hj} and x_{ik}.

Random Intercept and Slope Model Form

A random intercept and slope regression model has the form

$$y_{ij} = b_0 + b_1 x_{ij} + v_{i0} + v_{i1} x_{ij} + e_{ij}$$

for $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, m_i\}$ where

- $y_{ij} \in \mathbb{R}$ is the response for j-th measurement of i-th subject
- $b_0 \in \mathbb{R}$ is the fixed intercept for the regression model
- $b_1 \in \mathbb{R}$ is the fixed slope for the regression model
- $x_{ij} \in \mathbb{R}$ is the predictor for j-th measurement of i-th subject
- $v_{i0} \stackrel{\text{iid}}{\sim} \mathrm{N}(0, \sigma_0^2)$ is the random intercept for the i-th subject
- $v_{i1} \stackrel{\text{iid}}{\sim} \mathrm{N}(0, \sigma_1^2)$ is the random slope for the i-th subject
- $e_{ij} \stackrel{\text{iid}}{\sim} \mathrm{N}(0, \sigma_e^2)$ is a Gaussian error term

Random Intercept and Slope Model Assumptions

The fundamental assumptions of the RIS model are:

1. Relationship between X and Y is linear
2. x_{ij} and y_{ij} are observed random variables (known constants)
3. $v_{i0} \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma_0^2)$ and $v_{i1} \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma_1^2)$ are unobserved random variable
4. $(v_{i0}, v_{i1}) \overset{\text{iid}}{\sim} \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{01} & \sigma_1^2 \end{pmatrix}$
5. $e_{ij} \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma_e^2)$ is an unobserved random variable
6. (v_{i0}, v_{i1}) and e_{ij} are independent of one another
7. b_0 and b_1 are unknown constants
8. $(y_{ij} | x_{ij}) \sim \mathrm{N}(b_0 + b_1 x_{ij}, \sigma_{Y_{ij}}^2)$ where $\sigma_{Y_{ij}}^2 = \sigma_0^2 + 2\sigma_{01} x_{ij} + \sigma_1^2 x_{ij}^2 + \sigma_e^2$

Note: v_{i0} allows each subject to have unique regression intercept, and v_{i1} allows each subject to have unique regression slope.

Assumed Covariance Structure

The (conditional) covariance between any two observations is

$$
\begin{aligned}
Cov(y_{hj}, y_{ik}) &= E[(v_{h0} + v_{h1}x_{hj} + e_{hj})(v_{i0} + v_{i1}x_{ik} + e_{ik})] \\
&= E[v_{h0}v_{i0}] + E[v_{h0}(v_{i1}x_{ik} + e_{ik})] \\
&\quad + E[v_{i0}(v_{h1}x_{hj} + e_{hj})] + E[(v_{h1}x_{hj} + e_{hj})(v_{i1}x_{ik} + e_{ik})] \\
&= E[v_{h0}v_{i0}] + E[v_{h0}v_{i1}x_{ik}] + E[v_{i0}v_{h1}x_{hj}] \\
&\quad + E[v_{h1}x_{hj}v_{i1}x_{ik}] + E[e_{hj}e_{ik}] \\
&= \begin{cases} \sigma_0^2 + \sigma_{01}(x_{ij} + x_{ik}) + \sigma_1^2 x_{ij}x_{ik} & \text{if } h = i \text{ and } j \neq k \\ 0 & \text{if } h \neq i \end{cases}
\end{aligned}
$$

Note: this covariance is conditioned on fixed effects x_{hj} and x_{ik}.

LME Regression Model Form

A linear mixed-effects regression model has the form

$$y_{ij} = b_0 + \sum_{k=1}^{p} b_k x_{ijk} + v_{i0} + \sum_{k=1}^{q} v_{ik} z_{ijk} + e_{ij}$$

for $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, m_i\}$ where

- $y_{ij} \in \mathbb{R}$ is response for j-th measurement of i-th subject
- $b_0 \in \mathbb{R}$ is fixed intercept for the regression model
- $b_k \in \mathbb{R}$ is fixed slope for the k-th predictor
- $x_{ijk} \in \mathbb{R}$ is j-th measurement of k-th fixed predictor for i-th subject
- $v_{i0} \stackrel{\text{iid}}{\sim} \mathrm{N}(0, \sigma_0^2)$ is random intercept for the i-th subject
- $v_{ik} \stackrel{\text{iid}}{\sim} \mathrm{N}(0, \sigma_k^2)$ is random slope for k-th predictor of i-th subject
- $z_{ijk} \in \mathbb{R}$ is j-th measurement of k-th random predictor for i-th subj.
- $e_{ij} \stackrel{\text{iid}}{\sim} \mathrm{N}(0, \sigma_e^2)$ is a Gaussian error term

LME Regression Model Assumptions

The fundamental assumptions of the LMER model are:

1. Relationship between X_k and Y is linear (given other predictors)
2. x_{ijk}, z_{ijk}, and y_{ij} are observed random variables (known constants)
3. $\mathbf{v}_i = (v_{i0}, v_{i1}, \ldots, v_{iq})'$ is an unobserved random vector such that
$$\mathbf{v}_i \overset{\text{iid}}{\sim} \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}) \text{ where } \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_0^2 & \sigma_{01} & \cdots & \sigma_{0q} \\ \sigma_{10} & \sigma_1^2 & \cdots & \sigma_{1q} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{q0} & \sigma_{q1} & \cdots & \sigma_q^2 \end{pmatrix}$$
4. $e_{ij} \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma_e^2)$ is an unobserved random variable
5. $\mathbf{v}_i$ and e_{ij} are independent of one another
6. $(b_0, b_1, \ldots, b_p)$ are unknown constants
7. $(y_{ij}|x_{ij}) \sim \mathrm{N}(b_0 + \sum_{k=1}^p b_k x_{ijk}, \sigma_{Y_{ij}}^2)$ where
$\sigma_{Y_{ij}}^2 = \sigma_0^2 + 2\sum_{k=1}^q \sigma_{0k} z_{ijk} + 2\sum_{1 \le k < l \le q} \sigma_{kl} z_{ijk} z_{ijl} + \sum_{k=1}^q \sigma_k^2 z_{ijk}^2 + \sigma_e^2$

LMER in Matrix Form

Using matrix notation, we can write the LMER model as

$$\mathbf{y}_i = \mathbf{X}_i\mathbf{b} + \mathbf{Z}_i\mathbf{v}_i + \mathbf{e}_i$$

for $i \in \{1, \ldots, n\}$ where

- $\mathbf{y}_i = (y_{i1}, \ldots, y_{im_i})'$ is i-th subject's response vector
- $\mathbf{X}_i = [\mathbf{1}, \mathbf{x}_{i1}, \ldots, \mathbf{x}_{ip}]$ is fixed effects design matrix with $\mathbf{x}_{ik} = (x_{i1k}, \ldots, x_{im_ik})'$
- $\mathbf{b} = (b_0, b_1, \ldots, b_p)'$ is fixed effects vector
- $\mathbf{Z}_i = [\mathbf{1}, \mathbf{z}_{i1}, \ldots, \mathbf{z}_{iq}]$ is random effects design matrix with $\mathbf{z}_{ik} = (z_{i1k}, \ldots, z_{im_ik})'$
- $\mathbf{v}_i = (v_{i0}, v_{i1}, \ldots, v_{iq})'$ is random effects vector
- $\mathbf{e}_i = (e_{i1}, e_{i2}, \ldots, e_{im_i})'$ is error vector

Assumed Covariance Structure

LMER model assumes that

$$\mathbf{y}_i \sim \mathrm{N}(\mathbf{X}_i\mathbf{b}, \boldsymbol{\Sigma}_i)$$

where

$$\boldsymbol{\Sigma}_i = \mathbf{Z}_i\boldsymbol{\Sigma}\mathbf{Z}_i' + \sigma^2\mathbf{I}_n$$

is the $m_i \times m_i$ covariance matrix for the i-th subject's data.

LMER model assumes that

$$Cov[\mathbf{y}_h, \mathbf{y}_i] = \mathbf{0}_{m_h \times m_i} \quad \text{if} \quad h \neq i$$

given that data from different subjects are assumed independent.

Covariance Structure Choices

Assumed covariance structure $\mathbf{\Sigma}_i = \mathbf{Z}_i \mathbf{\Sigma} \mathbf{Z}_i' + \sigma^2 \mathbf{I}_n$ depends on $\mathbf{\Sigma}$.

- Need to choose some structure for $\mathbf{\Sigma}$

Some possible choices of covariance structure:

- *Unstructured*: all $(q+1)(q+2)/2$ unique parameters of $\mathbf{\Sigma}$ are free
- *Variance components*: σ_k^2 free and $\sigma_{kl} = 0$ if $k \neq l$
- *Compound symmetry*: $\sigma_k^2 = \sigma_v^2 + \sigma^2$ and $\sigma_{kl} = \sigma_v^2$
- *Autoregressive(1)*: $\sigma_{kl} = \sigma^2 \rho^{|k-l|}$ where ρ is autocorrelation
- *Toeplitz*: $\sigma_{kl} = \sigma^2 \rho_{|k-l|+1}$ where $\rho_1 = 1$

Unstructured Covariance Matrix

All $(q+1)(q+2)/2$ unique parameters of $\boldsymbol{\Sigma}$ are free.

With $q = 3$ we have $\mathbf{v}_i = (v_{i0}, v_{i1}, v_{i2}, v_{i3})$ and

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_0^2 & \sigma_{01} & \sigma_{02} & \sigma_{03} \\ \sigma_{10} & \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{20} & \sigma_{21} & \sigma_2^2 & \sigma_{23} \\ \sigma_{30} & \sigma_{31} & \sigma_{32} & \sigma_3^2 \end{pmatrix}$$

where 10 free parameters are the 4 variance parameters $\{\sigma_k^2\}_{k=0}^3$ and the 6 covariance parameters $\{\sigma_{kl}\}_{1 \leq k < l \leq 3}$.

Variance Components Covariance Matrix

σ_k^2 free and $\sigma_{kl} = 0$ if $k \neq l \quad \Longleftrightarrow \quad q + 1$ free parameters

With $q = 3$ we have $\mathbf{v}_i = (v_{i0}, v_{i1}, v_{i2}, v_{i3})$ and

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_0^2 & 0 & 0 & 0 \\ 0 & \sigma_1^2 & 0 & 0 \\ 0 & 0 & \sigma_2^2 & 0 \\ 0 & 0 & 0 & \sigma_3^2 \end{pmatrix}$$

where 4 variance parameters $\{\sigma_k^2\}_{k=0}^3$ are the only free parameters.

Compound Symmetry Covariance Matrix

$\sigma_k^2 = \sigma_v^2 + \sigma^2$ and $\sigma_{kl} = \sigma_v^2 \quad \Longleftrightarrow \quad$ 2 free parameters

With $q = 3$ we have $\mathbf{v}_i = (v_{i0}, v_{i1}, v_{i2}, v_{i3})$ and

$$\boldsymbol{\Sigma} = (\sigma_v^2 + \sigma^2) \begin{pmatrix} 1 & \omega & \omega & \omega \\ \omega & 1 & \omega & \omega \\ \omega & \omega & 1 & \omega \\ \omega & \omega & \omega & 1 \end{pmatrix}$$

where $\omega = \frac{\sigma_v^2}{\sigma_v^2 + \sigma^2}$ is the correlation between v_{ij} and v_{ik} (when $j \neq k$), and σ_v^2 and σ^2 are the only two free parameters.

Autoregressive(1) Covariance Matrix

$\sigma_{kl} = \sigma^2 \rho^{|k-l|}$ where ρ is autocorrelation $\iff$ 2 free parameters

With $q = 3$ we have $\mathbf{v}_i = (v_{i0}, v_{i1}, v_{i2}, v_{i3})$ and

$$\boldsymbol{\Sigma} = \sigma^2 \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{pmatrix}$$

where the autocorrelation ρ and σ^2 are the only two free parameters.

Toeplitz Covariance Matrix

$\sigma_{kl} = \sigma^2 \rho_{|k-l|+1}$ where $\rho_1 = 1 \quad \Longleftrightarrow \quad q + 1$ free parameters

With $q = 3$ we have $\mathbf{v}_i = (v_{i0}, v_{i1}, v_{i2}, v_{i3})$ and

$$\boldsymbol{\Sigma} = \sigma^2 \begin{pmatrix} 1 & \rho_1 & \rho_2 & \rho_3 \\ \rho_1 & 1 & \rho_1 & \rho_2 \\ \rho_2 & \rho_1 & 1 & \rho_1 \\ \rho_3 & \rho_2 & \rho_1 & 1 \end{pmatrix}$$

where the correlations (ρ_1, ρ_2, ρ_3) and the variance σ^2 are the only 4 free parameters.

Generalized Least Squares

If σ^2 and $\mathbf{\Sigma}$ are known, we could use generalized least squares:

$$\begin{aligned} GSSE &= \min_{\mathbf{b} \in \mathbb{R}^{p+1}} \sum_{i=1}^{n} (\mathbf{y}_i - \mathbf{X}_i \mathbf{b})' \mathbf{\Sigma}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \mathbf{b}) \\ &= \min_{\mathbf{b} \in \mathbb{R}^{p+1}} \sum_{i=1}^{n} (\tilde{\mathbf{y}}_i - \tilde{\mathbf{X}}_i \mathbf{b})' (\tilde{\mathbf{y}}_i - \tilde{\mathbf{X}}_i \mathbf{b}) \end{aligned}$$

where

- $\tilde{\mathbf{y}}_i = \mathbf{\Sigma}_i^{-1/2} \mathbf{y}_i$ is transformed response vector for i-th subject
- $\tilde{\mathbf{X}}_i = \mathbf{\Sigma}_i^{-1/2} \mathbf{X}_i$ is transformed design matrix for i-th subject
- $\mathbf{\Sigma}_i^{-1/2}$ is symmetric square root such that $\mathbf{\Sigma}_i^{-1/2} \mathbf{\Sigma}_i^{-1/2} = \mathbf{\Sigma}_i^{-1}$

Solution: $\hat{\mathbf{b}} = \left(\sum_{i=1}^{n} \mathbf{X}_i' \mathbf{\Sigma}_i^{-1} \mathbf{X}_i \right)^{-1} \sum_{i=1}^{n} \mathbf{X}_i' \mathbf{\Sigma}_i^{-1} \mathbf{y}_i$

Maximum Likelihood Estimation

If σ^2 and $\boldsymbol{\Sigma}$ are unknown, we can use maximum likelihood estimation to estimate the fixed effects ($\mathbf{b}$) and the variance components (σ^2 and $\boldsymbol{\Sigma}$).

There are two types of maximum likelihood (ML) estimation:

- Standard ML underestimates variance components ▸ ML
- Restricted ML (REML) provides consistent estimates ▸ REML

REML is default in many softwares, but need to use ML if you want to conduct likelihood ratio tests.

Estimating Fixed and Random Effects

If we only care about $\mathbf{b}$ use $\hat{\mathbf{b}} = (\mathbf{X}'\hat{\boldsymbol{\Sigma}}_*^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\boldsymbol{\Sigma}}_*^{-1}\mathbf{y}$

- $\hat{\boldsymbol{\Sigma}}_* = \mathbf{Z}\hat{\boldsymbol{\Sigma}}_{\mathrm{b}}\mathbf{Z}' + \hat{\sigma}^2\mathbf{I}$ is the estimated covariance matrix

If we care about both $\mathbf{b}$ and $\mathbf{v}$, then we solve mixed model equations

$$\begin{pmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \sigma^2\boldsymbol{\Sigma}_{\mathrm{b}}^{-1} \end{pmatrix} \begin{pmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{v}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{pmatrix} \iff \begin{array}{l} \hat{\mathbf{b}} = (\mathbf{X}'\hat{\boldsymbol{\Sigma}}_*^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\boldsymbol{\Sigma}}_*^{-1}\mathbf{y} \\ \hat{\mathbf{v}} = \hat{\boldsymbol{\Sigma}}_{\mathrm{b}}\mathbf{Z}'\hat{\boldsymbol{\Sigma}}_*^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) \end{array}$$

where

- $\hat{\mathbf{b}}$ is the empirical best linear unbiased estimator (BLUE) of $\mathbf{b}$
- $\hat{\mathbf{v}}$ is the empirical best linear unbiased predictor (BLUP) of $\mathbf{v}$

Mixed Model Equations

Likelihood Ratio Tests

Given two nested models, the Likelihood Ratio Test (LRT) statistic is

$$D = -2\ln\left(\frac{L(\mathcal{M}_0)}{L(\mathcal{M}_1)}\right) = 2[LL(\mathcal{M}_1) - LL(\mathcal{M}_0)]$$

where

- $L(\cdot)$ and $LL(\cdot)$ are the likelihood and log-likelihood
- $\mathcal{M}_0$ is null model with p parameters
- $\mathcal{M}_1$ is alternative model with $q = p + k$ parameters

Wilks's Theorem reveals that as $n \to \infty$ we have the result

$$D \sim \chi^2_k$$

where χ^2_k denotes chi-squared distribution with k degrees of freedom.

Inference for Random Effects

Use LRT to test significance of variance and covariance parameters.

To test the significance of a variance or covariance parameter use

$$H_0 : \sigma_{jk} = 0 \quad \text{versus} \quad \begin{cases} H_1 : \sigma_{jk} > 0 & \text{if } j = k \\ H_1 : \sigma_{jk} \neq 0 & \text{if } j \neq k \end{cases}$$

where σ_{jk} denotes the entry in cell j, k of $\boldsymbol{\Sigma}$.

Can use LRT idea to test hypotheses and compare to

- χ^2_k distribution if $j \neq k$
- Mixture of χ^2_k and 0 if $j = k$ (for simple cases)

Inference for Fixed Effects

Can use LRT idea to test fixed effects also

$$H_0 : \beta_k = 0 \quad \text{versus} \quad H_1 : \beta_k \neq 0$$

and compare D to χ^2_k distribution.

Reminder: The χ^2_k approximation is large sample result.

Could consider bootstrapping data to obtain non-asymptotic significance results.

TIMSS Data from 1997

Trends in International Mathematics and Science Study (TIMSS)[3]

- Ongoing study assessing STEM education around the world
- We will analyze data from 3rd and 4th grade students
- We have $n_T = 7,097$ students nested within $n = 146$ schools

```
> timss = read.table(paste(myfilepath,"timss1997.txt",sep=""),header=TRUE,
+                    colClasses=c(rep("factor",4),rep("numeric",3)))
> head(timss)
  idschool idstudent grade gender science  math hoursTV
1       10    100101     3   girl   146.7 137.0       3
2       10    100103     3   girl   148.8 145.3       2
3       10    100107     3   girl   150.0 152.3       4
4       10    100108     3   girl   146.9 144.3       3
5       10    100109     3    boy   144.3 140.3       3
6       10    100110     3    boy   156.5 159.2       2
```

[3] https://nces.ed.gov/TIMSS/

Define Level-2 `math` and `hoursTV` Variables

```
# get mean math and hoursTV info by school
> grpMmath = with(timss,tapply(math,idschool,mean))
> grpMhoursTV = with(timss,tapply(hoursTV,idschool,mean))

> # merge school mean scores with timss data.frame
> timss = merge(timss,data.frame(idschool=names(grpMmath),
+                                grpMmath=as.numeric(grpMmath),
+                                grpMhoursTV=as.numeric(grpMhoursTV)))
> head(timss)
  idschool idstudent grade gender science  math hoursTV grpMmath grpMhoursTV
1       10    100101     3   girl   146.7 137.0       3 152.0452    2.904762
2       10    100103     3   girl   148.8 145.3       2 152.0452    2.904762
3       10    100107     3   girl   150.0 152.3       4 152.0452    2.904762
4       10    100108     3   girl   146.9 144.3       3 152.0452    2.904762
5       10    100109     3    boy   144.3 140.3       3 152.0452    2.904762
6       10    100110     3    boy   156.5 159.2       2 152.0452    2.904762
```

Define Level-1 `math` and `hoursTV` Variables

```
# define group-centered math and hoursTV
> timss = cbind(timss,grpCmath=(timss$math-timss$grpMmath),
+               grpChoursTV=(timss$hoursTV-timss$grpMhoursTV))

> head(timss)
  idschool idstudent grade gender science  math hoursTV grpMmath grpMhoursTV    grpCmath grpChoursTV
1       10    100101     3   girl   146.7 137.0       3 152.0452    2.904762 -15.0452381   0.0952381
2       10    100103     3   girl   148.8 145.3       2 152.0452    2.904762  -6.7452381  -0.9047619
3       10    100107     3   girl   150.0 152.3       4 152.0452    2.904762   0.2547619   1.0952381
4       10    100108     3   girl   146.9 144.3       3 152.0452    2.904762  -7.7452381   0.0952381
5       10    100109     3    boy   144.3 140.3       3 152.0452    2.904762 -11.7452381   0.0952381
6       10    100110     3    boy   156.5 159.2       2 152.0452    2.904762   7.1547619  -0.9047619
```

Some Simple Random Intercept Models

```
> # random one-way ANOVA (ANOVA II Model)
> ramod = lmer(science ~ 1 + (1|idschool), data=timss, REML=FALSE)

> # add math as fixed effect
> rimod = lmer(science ~ 1 + math + (1|idschool), data=timss, REML=FALSE)

> # likelihood-ratio test for math
> anova(rimod,ramod)
Data: timss
Models:
ramod: science ~ 1 + (1 | idschool)
rimod: science ~ 1 + math + (1 | idschool)
      Df   AIC   BIC logLik deviance  Chisq Chi Df Pr(>Chisq)
ramod  3 51495 51516 -25744    51489
rimod  4 48490 48518 -24241    48482 3006.6      1  < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

More Complex Random Intercept Model

```
> ri5mod = lmer(science ~ 1 + grpCmath + grpMmath + grade + gender
+               + grpChoursTV + grpMhoursTV + (1|idschool), data=timss, REML=FALSE)
> ri5mod
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: science ~ 1 + grpCmath + grpMmath + grade + gender
                   + grpChoursTV + grpMhoursTV + (1 | idschool)
   Data: timss
      AIC       BIC    logLik  deviance  df.resid
 48370.84  48432.65 -24176.42  48352.84      7088
Random effects:
 Groups   Name        Std.Dev.
 idschool (Intercept) 1.859
 Residual             7.193
Number of obs: 7097, groups:  idschool, 146
Fixed Effects:
(Intercept)     grpCmath     grpMmath       grade4   gendergirl
    26.2078       0.5528       0.8616       0.9395      -1.1407
grpChoursTV  grpMhoursTV
    -0.1246      -1.9785
```

Random Intercept and Slopes (Unstructured)

```
> risucmod = lmer(science ~ 1 + grpCmath + grpMmath + grade + gender
+                 + grpChoursTV + grpMhoursTV + (grpCmath+grpChoursTV|idschool),
+                 data=timss, REML=FALSE) REML=FALSE)
> risucmod
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: science ~ 1 + grpCmath + grpMmath + grade + gender
                 + grpChoursTV + grpMhoursTV + (grpCmath + grpChoursTV | idschool)
   Data: timss
      AIC       BIC    logLik  deviance  df.resid
 48341.60  48437.74 -24156.80  48313.60      7083
Random effects:
 Groups   Name        Std.Dev. Corr
 idschool (Intercept) 1.89212
          grpCmath    0.09541   0.46
          grpChoursTV 0.36491   0.36 -0.27
 Residual             7.12812
Number of obs: 7097, groups:  idschool, 146
Fixed Effects:
(Intercept)     grpCmath     grpMmath       grade4   gendergirl
    15.6041       0.5593       0.9309       0.8990      -1.1839
grpChoursTV  grpMhoursTV
    -0.1152      -1.9144
```

Random Intercept and Slopes (Variance Components)

```
> risvcmod = lmer(science ~ 1 + grpCmath + grpMmath + grade + gender
+                 + grpChoursTV + grpMhoursTV + (grpCmath+grpChoursTV||idschool),
+                 data=timss, REML=FALSE)
> risvcmod
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: science ~ 1 + grpCmath + grpMmath + grade + gender + grpChoursTV
                  + grpMhoursTV + ((1 | idschool) + (0 + grpCmath | idschool)
                  + (0 + grpChoursTV | idschool))
   Data: timss
      AIC       BIC    logLik  deviance  df.resid
 48344.04  48419.58 -24161.02  48322.04      7086
Random effects:
 Groups     Name        Std.Dev.
 idschool   (Intercept) 1.86618
 idschool.1 grpCmath    0.09643
 idschool.2 grpChoursTV 0.36752
 Residual               7.12626
Number of obs: 7097, groups:  idschool, 146
Fixed Effects:
(Intercept)     grpCmath     grpMmath       grade4   gendergirl
    26.2279       0.5600       0.8616       0.9343      -1.1856
grpChoursTV  grpMhoursTV
    -0.1203      -1.9774
```

Likelihood Ratio Test on Covariance Components

```
> anova(risucmod,risvcmod)
Data: timss
Models:
risvcmod: science ~ 1 + grpCmath + grpMmath + grade + gender + grpChoursTV +
risvcmod:     grpMhoursTV + ((1 | idschool) + (0 + grpCmath | idschool) +
risvcmod:     (0 + grpChoursTV | idschool))
risucmod: science ~ 1 + grpCmath + grpMmath + grade + gender + grpChoursTV +
risucmod:     grpMhoursTV + (grpCmath + grpChoursTV | idschool)
         Df   AIC   BIC logLik deviance  Chisq Chi Df Pr(>Chisq)
risvcmod 11 48344 48420 -24161    48322
risucmod 14 48342 48438 -24157    48314 8.4417      3    0.03771 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We reject $H_0 : \sigma_{jk} = 0 \; \forall j \neq k$ at a significance level of $\alpha = 0.05$.
We retain $H_0 : \sigma_{jk} = 0 \; \forall j \neq k$ at a significance level of $\alpha = 0.01$.

More Complex Random Effects Structure

```
> risicmod = lmer(science ~ 1 + grpCmath + grpMmath + grade + gender
+                 + grpChoursTV + grpMhoursTV + (1|idschool)
+                 + (0+grpCmath+grpChoursTV|idschool), data=timss, REML=FALSE)
> risicmod
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: science ~ 1 + grpCmath + grpMmath + grade + gender + grpChoursTV +
    grpMhoursTV + (1 | idschool) + (0 + grpCmath + grpChoursTV |      idschool)
   Data: timss
      AIC       BIC    logLik  deviance  df.resid
 48345.49  48427.90 -24160.74  48321.49      7085
Random effects:
 Groups     Name        Std.Dev. Corr
 idschool   (Intercept) 1.86615
 idschool.1 grpCmath    0.09659
            grpChoursTV 0.36331  -0.26
 Residual               7.12655
Number of obs: 7097, groups:  idschool, 146
Fixed Effects:
(Intercept)     grpCmath     grpMmath       grade4   gendergirl
    26.2272       0.5598       0.8616       0.9358      -1.1854
grpChoursTV  grpMhoursTV
    -0.1165      -1.9775
```

Appendix

Likelihood Function

A vector $\mathbf{y} = (y_1, \ldots, y_n)'$ with multivariate normal distribution has pdf:

$$f(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-n/2}|\boldsymbol{\Sigma}|^{-1/2}e^{-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})}$$

where $\boldsymbol{\mu}$ is the mean vector and $\boldsymbol{\Sigma}$ is the covariance matrix.

Thus, the likelihood function for the model is given by

$$L(\mathbf{b}, \boldsymbol{\Sigma}, \sigma^2|\mathbf{y}_1, \ldots, \mathbf{y}_n) = \prod_{i=1}^{n}(2\pi)^{-m_i/2}|\boldsymbol{\Sigma}_i|^{-1/2}e^{-\frac{1}{2}(\mathbf{y}_i-\mathbf{X}_i\mathbf{b})'\boldsymbol{\Sigma}_i^{-1}(\mathbf{y}_i-\mathbf{X}_i\mathbf{b})}$$

where $\boldsymbol{\Sigma}_i = \mathbf{Z}_i\boldsymbol{\Sigma}\mathbf{Z}_i' + \sigma^2\mathbf{I}$ with $\mathbf{X}_i$ and $\mathbf{Z}_i$ known design matrices.

Maximum Likelihood Estimates

Plugging $\hat{\mathbf{b}} = \left(\sum_{i=1}^{n} \mathbf{X}_i'\mathbf{\Sigma}_i^{-1}\mathbf{X}_i\right)^{-1} \sum_{i=1}^{n} \mathbf{X}_i'\mathbf{\Sigma}_i^{-1}\mathbf{y}_i$ into the likelihood, we can write the log-likelihood

$$\ln\{L(\mathbf{\Sigma}, \sigma^2|\mathbf{y}_1, \ldots, \mathbf{y}_n)\} = -\frac{n_T}{2}\ln(2\pi) - \frac{1}{2}\sum_{i=1}^{n}\ln(|\mathbf{\Sigma}_i|) - \frac{1}{2}\sum_{i=1}^{n}\mathbf{r}_i'\mathbf{\Sigma}_i^{-1}\mathbf{r}_i$$

where $n_T = \sum_{i=1}^{n} m_i$ and $\mathbf{r}_i = \mathbf{y}_i - \mathbf{X}_i\hat{\mathbf{b}}$.

We can now maximize $\ln\{L(\mathbf{\Sigma}, \sigma^2|\mathbf{y}_1, \ldots, \mathbf{y}_n)\}$ to get MLEs $\hat{\mathbf{\Sigma}}$ and $\hat{\sigma}^2$.

Problem: our MLE estimates $\hat{\mathbf{\Sigma}}$ and $\hat{\sigma}^2$ depend on having the correct mean structure in the model, so we tend to underestimate.

Return

REML Error Contrasts

We need to work with the "stacked" model form: $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{v} + \mathbf{e}$

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_n \end{pmatrix}, \mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_n \end{pmatrix}, \mathbf{Z} = \begin{pmatrix} \mathbf{Z}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{Z}_n \end{pmatrix}, \mathbf{v} = \begin{pmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \vdots \\ \mathbf{v}_n \end{pmatrix}, \mathbf{e} = \begin{pmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \vdots \\ \mathbf{e}_n \end{pmatrix}$$

Note that $\mathbf{y} \sim \mathbf{N}(\mathbf{X}\mathbf{b}, \mathbf{\Sigma}_*)$ where $\mathbf{\Sigma}_* = \mathbf{Z}\mathbf{\Sigma}_{\mathrm{b}}\mathbf{Z}' + \sigma^2\mathbf{I}$ is block diagonal and the matrix $\mathbf{\Sigma}_{\mathrm{b}} = \mathrm{bdiag}(\mathbf{\Sigma})$ is $n(q+1) \times n(q+1)$ block diagonal matrix.

Form $\mathbf{w} = \mathbf{K}'\mathbf{y}$ where $\mathbf{K}$ is an $n_T \times (n_T - p - 1)$ matrix where $\mathbf{K}'\mathbf{X} = \mathbf{0}$

- Doesn't matter what $\mathbf{K}$ we choose so pick one such that $\mathbf{K}'\mathbf{K} = \mathbf{I}$
- $\mathbf{w} \sim \mathrm{N}(\mathbf{0}, \mathbf{K}'\mathbf{\Sigma}_*\mathbf{K})$ does not depend on the model mean structure

REML Log-likelihood Function

The log-likelihood of the model written in terms of $\mathbf{w}$ is

$$\ln\{L(\boldsymbol{\Sigma}, \sigma^2 | \mathbf{w})\} = -\frac{n_T - p - 1}{2}\ln(2\pi) - \frac{1}{2}\ln(|\mathbf{K}'\boldsymbol{\Sigma}_*\mathbf{K}|) - \frac{1}{2}\mathbf{w}'[\mathbf{K}'\boldsymbol{\Sigma}_*\mathbf{K}]^{-1}\mathbf{w}$$

As long as $\mathbf{K}'\mathbf{X} = \mathbf{0}$ and $\mathrm{rank}(\mathbf{X}) = p + 1$, it can be shown that:

- $\ln(|\mathbf{K}'\boldsymbol{\Sigma}_*\mathbf{K}|) = \ln(|\boldsymbol{\Sigma}_*|) + \ln(|\mathbf{X}'\boldsymbol{\Sigma}_*^{-1}\mathbf{X}|)$
- $\mathbf{y}'\mathbf{K}[\mathbf{K}'\boldsymbol{\Sigma}_*\mathbf{K}]^{-1}\mathbf{K}'\mathbf{y} = \mathbf{r}'\boldsymbol{\Sigma}_*^{-1}\mathbf{r}$ where $\mathbf{r} = \mathbf{y} - \mathbf{X}\hat{\mathbf{b}}$
- $\hat{\mathbf{b}} = (\mathbf{X}'\boldsymbol{\Sigma}_*^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}_*^{-1}\mathbf{y} = \left(\sum_{i=1}^n \mathbf{X}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i\right)^{-1}\sum_{i=1}^n \mathbf{X}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{y}_i$

Restricted Maximum Likelihood Estimates

We can rewrite the restricted model log-likelihood as

$$\ln\{\tilde{L}(\boldsymbol{\Sigma}, \sigma^2|\mathbf{y})\} = -\frac{\tilde{n}_T}{2}\ln(2\pi) - \frac{1}{2}\ln(|\boldsymbol{\Sigma}_*|) - \frac{1}{2}\ln(|\mathbf{X}'\boldsymbol{\Sigma}_*^{-1}\mathbf{X}|) - \frac{1}{2}\mathbf{r}'\boldsymbol{\Sigma}_*^{-1}\mathbf{r}$$

where $\tilde{n}_T = n_T - p - 1$.

For comparison the log-likelihood using stacked model notation is

$$\ln\{L(\boldsymbol{\Sigma}, \sigma^2|\mathbf{y})\} = -\frac{n_T}{2}\ln(2\pi) - \frac{1}{2}\ln(|\boldsymbol{\Sigma}_*|) - \frac{1}{2}\mathbf{r}'\boldsymbol{\Sigma}_*^{-1}\mathbf{r}$$

Maximize $\ln\{\tilde{L}(\boldsymbol{\Sigma}, \sigma^2|\mathbf{y})\}$ to get REML $\hat{\boldsymbol{\Sigma}}$ and $\hat{\sigma}^2$. Return

Joint Likelihood and Log-Likelihood Function

Note that the pdf of $\mathbf{y}$ given $(\mathbf{b}, \mathbf{v}, \sigma^2)$ is:

$$f(\mathbf{y}|\mathbf{b}, \mathbf{v}, \sigma^2) = (2\pi)^{-n_T/2}|\sigma^2\mathbf{I}|^{-1/2}e^{-\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{X}\mathbf{b}-\mathbf{Z}\mathbf{v})'(\mathbf{y}-\mathbf{X}\mathbf{b}-\mathbf{Z}\mathbf{v})}$$

Using $f(\mathbf{v}|\mathbf{\Sigma}_{\mathrm{b}}) = (2\pi)^{-\frac{n(q+1)}{2}}|\mathbf{\Sigma}_{\mathrm{b}}|^{-1/2}e^{-\frac{1}{2}\mathbf{v}'\mathbf{\Sigma}_{\mathrm{b}}^{-1}\mathbf{v}}$, we have that:

$$\begin{aligned} f(\mathbf{y}, \mathbf{v}|\mathbf{b}, \sigma^2, \mathbf{\Sigma}_{\mathrm{b}}) &= f(\mathbf{y}|\mathbf{b}, \mathbf{v}, \sigma^2)f(\mathbf{v}|\mathbf{\Sigma}_{\mathrm{b}}) \\ &= (2\pi)^{-\frac{n_T+n(q+1)}{2}}|\sigma^2\mathbf{I}|^{-1/2}|\mathbf{\Sigma}_{\mathrm{b}}|^{-1/2} \\ &\quad \times e^{-\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{X}\mathbf{b}-\mathbf{Z}\mathbf{v})'(\mathbf{y}-\mathbf{X}\mathbf{b}-\mathbf{Z}\mathbf{v})-\frac{1}{2}\mathbf{v}'\mathbf{\Sigma}_{\mathrm{b}}^{-1}\mathbf{v}} \end{aligned}$$

The log-likelihood of $(\mathbf{b}, \mathbf{v})$ given $(\mathbf{y}, \sigma^2, \mathbf{\Sigma}_{\mathrm{b}})$ is of the form

$$\ln\{L(\mathbf{b}, \mathbf{v}|\mathbf{y}, \sigma^2, \mathbf{\Sigma}_{\mathrm{b}})\} \propto -(\mathbf{y} - \mathbf{X}\mathbf{b} - \mathbf{Z}\mathbf{v})'(\mathbf{y} - \mathbf{X}\mathbf{b} - \mathbf{Z}\mathbf{v}) - \sigma^2\mathbf{v}'\mathbf{\Sigma}_{\mathrm{b}}^{-1}\mathbf{v} + c$$

where c is some constant that does not depend on $\mathbf{b}$ or $\mathbf{v}$.

Solving Mixed Model Equations

$\max_{\mathbf{b},\mathbf{v}} \ln\{L(\mathbf{b},\mathbf{v}|\mathbf{y},\sigma^2,\boldsymbol{\Sigma}_{\mathrm{b}})\} \Longleftrightarrow \min_{\mathbf{b},\mathbf{v}} -\ln\{L(\mathbf{b},\mathbf{v}|\mathbf{y},\sigma^2,\boldsymbol{\Sigma}_{\mathrm{b}})\}$ and

$$
\begin{aligned}
-\ln\{L(\mathbf{b},\mathbf{v}|\mathbf{y},\sigma^2,\boldsymbol{\Sigma}_{\mathrm{b}}) &= \mathbf{y}'\mathbf{y} - 2\mathbf{y}'(\mathbf{X}\mathbf{b}+\mathbf{Z}\mathbf{v}) + \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b} + 2\mathbf{b}'\mathbf{X}'\mathbf{Z}\mathbf{v} \\
&\quad + \mathbf{v}'\mathbf{Z}'\mathbf{Z}\mathbf{v} + \sigma^2\mathbf{v}'\boldsymbol{\Sigma}_{\mathrm{b}}^{-1}\mathbf{v} + c \\
&= \mathbf{y}'\mathbf{y} - 2\mathbf{y}'\mathbf{W}\mathbf{u} + \mathbf{u}'(\mathbf{W}'\mathbf{W} + \sigma^2\tilde{\boldsymbol{\Sigma}}_{\mathrm{b}}^{-1})\mathbf{u} + c
\end{aligned}
$$

where

- $\mathbf{u} = (\mathbf{b}',\mathbf{v}')'$ contains the fixed and random effects coefficients
- $\mathbf{W} = (\mathbf{X},\mathbf{Z})$ contains the fixed and random effects design matrices
- $\tilde{\boldsymbol{\Sigma}}_{\mathrm{b}}^{-1} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{\mathrm{b}}^{-1} \end{pmatrix}$, which is $\boldsymbol{\Sigma}_{\mathrm{b}}^{-1}$ augmented with zeros corresponding to $\mathbf{X}$ in $\mathbf{W}$

Solving Mixed Model Equations (continued)

Taking the derivative of the negative log-likelihood w.r.t. **u** gives

$$\frac{\partial - \ln\{L(\mathbf{b}, \mathbf{v}|\mathbf{y}, \sigma^2, \boldsymbol{\Sigma}_{\mathrm{b}})}{\partial \mathbf{u}} = -2\mathbf{W}'\mathbf{y} + 2(\mathbf{W}'\mathbf{W} + \sigma^2 \tilde{\boldsymbol{\Sigma}}_{\mathrm{b}}^{-1})\mathbf{u}$$

and setting to zero and solving for **u** gives

$$\hat{\mathbf{u}} = (\mathbf{W}'\mathbf{W} + \sigma^2 \tilde{\boldsymbol{\Sigma}}_{\mathrm{b}}^{-1})^{-1}\mathbf{W}'\mathbf{y}$$

which gives us the mixed model equations and result

$$\begin{pmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \sigma^2 \boldsymbol{\Sigma}_{\mathrm{b}}^{-1} \end{pmatrix} \begin{pmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{v}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{pmatrix} \iff \begin{array}{l} \hat{\mathbf{b}} = (\mathbf{X}'\hat{\boldsymbol{\Sigma}}_*^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\boldsymbol{\Sigma}}_*^{-1}\mathbf{y} \\ \hat{\mathbf{v}} = \hat{\boldsymbol{\Sigma}}_{\mathrm{b}}\mathbf{Z}'\hat{\boldsymbol{\Sigma}}_*^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) \end{array}$$

▸ Return