

UC Irvine ISI-BUDS Day 14

Zhaoxia Yu

7/29/2022

Study Goals

UC Irvine
ISI-BUDS Day 14

Zhaoxia Yu

LM, LME, GLM,
and GLMM

LME Examples:
Example 2

LME Examples:
Example 3

GLMM

- ▶ LM, LME, GLM, and GLMM
- ▶ LME Examples: Example 2
- ▶ LME Examples: Example 3
- ▶ Generalized Linear Mixed-Effects Model (GLMM)
- ▶ The slides are based on my published work:
<https://doi.org/10.1016/j.neuron.2021.10.030>
https://yu-zhaoxia.github.io/MM_in_Neuroscience/

LM, LME, GLM, and GLMM

LM and its Matrix Form

UC Irvine
ISI-BUDS Day 14

Zhaoxia Yu

LM, LME, GLM,
and GLMM

LME Examples:
Example 2

LME Examples:
Example 3

GLMM

► The model

$$Y_i = \beta_0 + x_{i1} \times \beta_1 + \dots + x_{ip} \times \beta_p + \epsilon_i, i = 1, \dots, n$$

► Assumptions

- $E(\epsilon_i) = 0$, which is equivalent to $E(Y_i|X_i) = \beta_0 + x_{i1} \times \beta_1 + \dots + x_{ip} \times \beta_p$
- $Var(\epsilon_i) = \sigma^2$, which is equivalent to $Var(Y_i|X_i) = \sigma^2$.
- $(\epsilon_1, \dots, \epsilon_n)$ are mutually independent

- The equivalent matrix form is $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where the random errors have a zero mean and a common variance, and are independent with each other

LM and LME

UC Irvine
ISI-BUDS Day 14

Zhaoxia Yu

LM, LME, GLM,
and GLMM

LME Examples:
Example 2

LME Examples:
Example 3

GLMM

- ▶ The components of LM: $\mathbf{Y} = \mathbf{X}\beta + \epsilon$
 - ▶ a linear predictor $\mathbf{X}\beta$
 - ▶ random errors ϵ are independent, have a zero mean and a constant variance.
 - ▶ $\epsilon \sim N(0, \sigma^2 \mathbf{I})$ is used for deriving t- and F-tests.
Typically this assumption is very critical as long as the sample size is not too small
- ▶ The components of LME: $\mathbf{Y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \epsilon$
 - ▶ fixed-effects: a linear predictor $\mathbf{X}\beta$
 - ▶ random-effects: $\mathbf{Z}\mathbf{u}$, where $\mathbf{u} \sim N(0, G)$. E.g., $G = \sigma_b^2 \mathbf{I}$.
 - ▶ random errors: $\epsilon \sim N(0, \sigma^2 \mathbf{I})$, independent with \mathbf{u} .

- ▶ The components of GLM:
 - ▶ a linear predictor $X\beta$
 - ▶ a link function to connect $E(Y|X)$ and $X\beta$:
 $g(E(Y|X)) = X\beta$
 - ▶ a distribution for Y
- ▶ The components of GLMM:
 - ▶ fixed-effects: a linear predictor $X\beta$
 - ▶ random-effects: $Z\mathbf{u}$, where $\mathbf{u} \sim N(0, G)$. E.g., $G = \sigma_b^2 \mathbf{I}$.
 - ▶ a link function to connect $E(Y|X, \mathbf{u})$ and $X\beta$:
 $g(E(Y|X, \mathbf{u})) = X\beta + Z\mathbf{u}$
 - ▶ a distribution for Y

LME Examples: Exmaple 2

LME Examples: Exmample 2

UC Irvine
ISI-BUDS Day 14

Zhaoxia Yu

LM, LME, GLM,
and GLMM

LME Examples:
Exmample 2

LME Examples:
Example 3

GLMM

- ▶ Research question: determine how in vivo calcium (Ca^{++}) activity of PV cells (measured longitudinally by the genetically encoded Ca^{++} indicator GCaMP6s) changes over time after ketamine treatment
- ▶ Study: Ca^{++} event frequencies were measured at 24h, 48h, 72h, and 1 week after ketamine treatment in four mice
- ▶ Want to compare Ca^{++} event frequency at 24h to the other three time points.
- ▶ In total, Ca^{++} event frequencies of 1,724 neurons were measured.

Example 2: Data

UC Irvine
ISI-BUDS Day 14

Zhaoxia Yu

LM, LME, GLM,
and GLMM

LME Examples:
Example 2

LME Examples:
Example 3

GLMM

```
library(nlme)
library(lme4)
library(lmerTest)
Ex2=read.csv("https://www.ics.uci.edu/~zhaoxia/Data/BeyondTan
Ex2$treatment_idx=Ex2$treatment_idx-4
Ex2$treatment_idx=as.factor(Ex2$treatment_idx)
### covert the variable of mouse IDs to a factor
Ex2$midx=as.factor(Ex2$midx)
```

Example 2: Wrong analysis

UC Irvine
ISI-BUDS Day 14

Zhaoxia Yu

LM, LME, GLM,
and GLMM

LME Examples:
Example 2

LME Examples:
Example 3

GLMM

```
lm.obj=lm(res~treatment_idx, data=Ex2)
summary(lm.obj)$coefficients
```

##		Estimate	Std. Error	t value	Pr(>
##	(Intercept)	0.71490545	0.01233741	57.9461618	0.000000e
##	treatment_idx2	-0.07802047	0.01701121	-4.5864155	4.835037e
##	treatment_idx3	0.00914741	0.01718859	0.5321791	5.946707e
##	treatment_idx4	0.04971562	0.01633230	3.0440051	2.369903e

Example 2: Wrong analysis

UC Irvine
ISI-BUDS Day 14

Zhaoxia Yu

LM, LME, GLM,
and GLMM

LME Examples:
Example 2

LME Examples:
Example 3

GLMM

- ▶ The LM (including ANOVA, t-test) analysis results indicate
 - ▶ significantly reduced Ca^{++} activity at 48h relative to 24h with $p = 4.8 \times 10^{-6}$
 - ▶ significantly increased Ca^{++} activity at 1week compared to 24h with $p = 2.4 \times 10^{-3}$
 - ▶ However, if we account for repeated measures due to cells clustered in mice using LME, the changes are no longer significant

Example 2: LME

UC Irvine
ISI-BUDS Day 14

Zhaoxia Yu

LM, LME, GLM,
and GLMM

LME Examples:
Example 2

data= Ex2
LME Examples:
Example 3

GLMM

```
lmer.obj=lmerTest::lmer(res~treatment_idx+(1|midx), data=Ex2)  
summary(lmer.obj)$coefficients
```

##		Estimate	Std. Error	df	t va
##	(Intercept)	0.699786009	0.03484986	4.901964	20.0800
##	treatment_idx2	-0.017490109	0.01726513	1723.485832	-1.0130
##	treatment_idx3	0.009353984	0.01657856	1720.292658	0.5642
##	treatment_idx4	0.029448530	0.01656107	1719.621372	1.7781

Example 2: LM vs LME

UC Irvine
ISI-BUDS Day 14

Zhaoxia Yu

LM, LME, GLM,
and GLMM

LME Examples:
Example 2

LME Examples:
Example 3

GLMM

Estimated changes of Ca⁺ event frequency (the baseline is 24h after treatment)

	48h	72h	1wk
LM est	-0.078 \pm 0.017	0.009 \pm 0.017	0.050 \pm 0.016
LM p	4.8×10^{-6}	0.595	2.4×10^{-3}
LME est	-0.017 \pm 0.017	0.009 \pm 0.017	0.029 \pm 0.017
LME p	0.311	0.573	0.076

Pooling data naively is not a good idea

UC Irvine
ISI-BUDS Day 14

Zhaoxia Yu

LM, LME, GLM,
and GLMM

LME Examples:
Example 2

LME Examples:
Example 3

GLMM

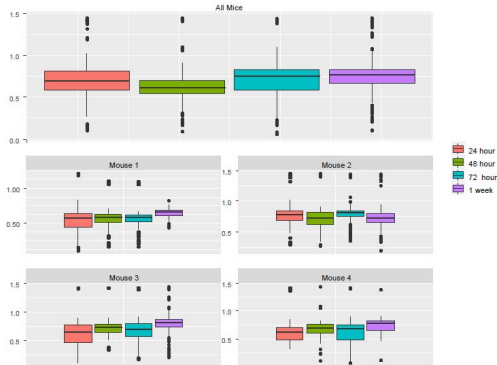


Figure 1: The boxplots of Ca++ event frequencies measured at four time points. (A) Boxplot of Ca++ event frequencies using the pooled neurons from four mice. (B) boxplots of Ca++ event frequencies stratified by individual mice.

Pooling data naively is not a good idea

UC Irvine
ISI-BUDS Day 14

Zhaoxia Yu

LM, LME, GLM,
and GLMM

LME Examples:
Example 2

LME Examples:
Example 3

GLMM

- ▶ Consider the change in Ca^{++} activities from 24h to 48h
- ▶ Pooled data from all mice:
 - ▶ The box plots suggest reduction in Ca^{++} activities
- ▶ Individual mice data:
 - ▶ The box plots of Mouse 2 suggest a noticeable reduction
 - ▶ However, there was almost no change in Mouse 1
 - ▶ Mouse 3 and Mouse 4 might suggest small increases, rather than decreases

Pooling data naively is not a good idea

UC Irvine
ISI-BUDS Day 14

Zhaoxia Yu

LM, LME, GLM,
and GLMM

LME Examples:
Example 2

LME Examples:
Example 3

GLMM

- Why do the pooled data follow the pattern of Mouse 2?

	24h	48h	72h	1wk	Total
Mouse 1	81	254	88	43	466(27%)
Mouse 2	206	101	210	222	739 (43%)
Mouse 3	33	18	51	207	309 (18%)
Mouse 4	63	52	58	37	210 (12%)
Total	383	425	407	509	1,724 (100%)

- Mouse 2 contributed 43% cells!

Remark: on the minimum number of levels for using random-effects

UC Irvine
ISI-BUDS Day 14

Zhaoxia Yu

LM, LME, GLM,
and GLMM

LME Examples:
Example 2

LME Examples:
Example 3

GLMM

- ▶ In Example 2, the number of levels in the random-effects variable is four, as there are four mice.
- ▶ According to Gelman and Hill 2006, it does not hurt to use random-effects in this situation.
- ▶ There is no unique answer on the minimum number of levels for using random-effects.

Remark: on the minimum number of levels for using random-effects

UC Irvine
ISI-BUDS Day 14

Zhaoxia Yu

LM, LME, GLM,
and GLMM

LME Examples:
Example 2

LME Examples:
Example 3

GLMM

- ▶ An alternative is to include the animal ID variable as factor with fixed animal effects.
- ▶ Neither of two approaches is the same as fitting an LM to the pooled cells naively.
- ▶ In a more extreme case, for an experiment using only two monkeys for example,
 - ▶ naively pooling data (such as neurons) is NOT recommended.
 - ▶ a more appropriate approach is to analyze the animals separately and then check whether the results from the two animals are consistent

LME Examples: Example 3

Example 3: Data Structure

UC Irvine
ISI-BUDS Day 14

Zhaoxia Yu

LM, LME, GLM,
and GLMM

LME Examples:
Example 2

LME Examples:
Example 3

GLMM

- ▶ Ca^{++} event integrated amplitudes are compared between baseline and 24h after ketamine treatment.
- ▶ 1244 cells were sampled from 11 mice
- ▶ each cell was measured twice (baseline and after ketamine treatment)
- ▶ correlation arises from both cells and animals, which creates a three-level structure:
 - ▶ measurements within cells and cells within animals.

```
library(nlme)
library(lme4)
library(lmerTest)
Ex3=read.csv("https://www.ics.uci.edu/~zhaoxia/Data/BeyondTan
```

Example 3: LM vs LME

UC Irvine
ISI-BUDS Day 14

Zhaoxia Yu

LM, LME, GLM,
and GLMM

LME Examples:
Example 2

wrong analysis: using the linear model

```
summary(lm(res~treatment, data=Ex3[!is.na(Ex3$res),]))
```

wrong analysis using t tests (paired or unpaired)

```
t.test(Ex3[Ex3$treatment==1,"res"], Ex3[Ex3$treatment==2,"res"])
```

```
t.test(Ex3[Ex3$treatment==1,"res"], Ex3[Ex3$treatment==2,"res"])
```

```
t.test(Ex3[Ex3$treatment==1,"res"], Ex3[Ex3$treatment==2,"res"])
```

#LME

```
lme.obj1=lme(res~ treatment, random =~1| midx/cidx, data= Ex3)
```

```
summary(lme.obj1)
```

Example 3
GLMM

Example 3: LM vs LME

UC Irvine
ISI-BUDS Day 14

Zhaoxia Yu

LM, LME, GLM,
and GLMM

LME Examples:
Example 2

LME Examples:
Example 3

GLMM

- ▶ LME and LM produce similar estimates for the fix-effects coefficients
- ▶ the standard error of the LM is larger; the p-value based on LME is smaller (0.0036 for LM vs 0.0001 for LME).
- ▶ In this example, since the two measures from each cell are positively correlated, the variance of the differences is smaller when treating the data as paired rather than independent.
- ▶ As a result, LME produces a smaller p-value
- ▶ Rigorous statistical analysis is not a hunt for the smallest p value (commonly known as p-hacking or significance chasing)

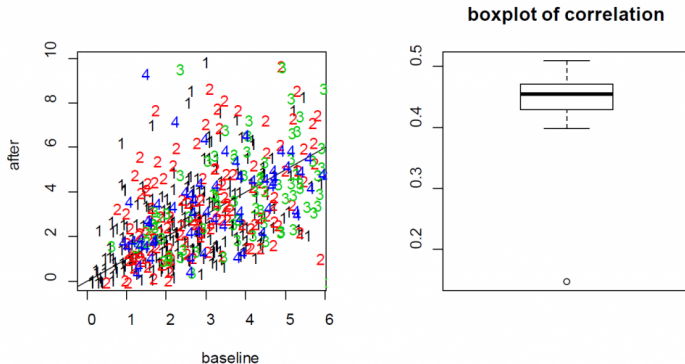


Figure 2: (Left) the scatter plot of Ca^{++} event integrated amplitude at baseline vs 24h after treatment for the neurons from four mice (labeled as 1, 2, 3 and 4) indicates that the baseline and after-treatment measures are positively correlated. (Right) boxplot of the baseline and after-treatment correlations of the 11 mice.

A note on “nested” random effects

UC Irvine
ISI-BUDS Day 14

Zhaoxia Yu

LM, LME, GLM,
and GLMM

LME Examples:
Example 2

LME Examples:
Example 3

GLMM

- ▶ When specifying the nested random effects, we used “random = ~1 | midx/cidx”.
- ▶ This leads to random effects at two levels: the mouse level and the cells-within-mouse level.
- ▶ This specification is important if same cell IDs might appear in different mice.
- ▶ When each cell has its unique ID, just like “cidx” variable in Example 3, it does not matter and “random = list(midx=~1, cidx=~1)” leads to exactly the same model.

A note on “nested” random effects

UC Irvine
ISI-BUDS Day 14

Zhaoxia Yu

LM, LME, GLM,
and GLMM

LME Examples:
Example 2

LME Examples:
Example 3

GLMM

to verify that the cell IDs are indeed unique

```
length(unique(Ex3$cidx))
```

#lme.obj2 is the same as lme.obj

```
lme.obj2=lme(res~ treatment, random =list(midx=~1, cidx=~1),  
summary(lme.obj2)
```

On models with more random effects

UC Irvine
ISI-BUDS Day 14

Zhaoxia Yu

LM, LME, GLM,
and GLMM

LME Examples:
Example 2

LME Examples:
Example 3

GLMM

- ▶ The above LME model only involves random intercepts.
- ▶ There might be random effects due to multiple sources.
- ▶ A model with more random-effects might be a better choice.
- ▶ Visualization is a useful exploratory tool to help identify an appropriate model.

On models with more random effects

UC Irvine
ISI-BUDS Day 14

Zhaoxia Yu

LM, LME, GLM,
and GLMM

LME Examples:
Exmaple 2

LME Examples:
Example 3

GLMM

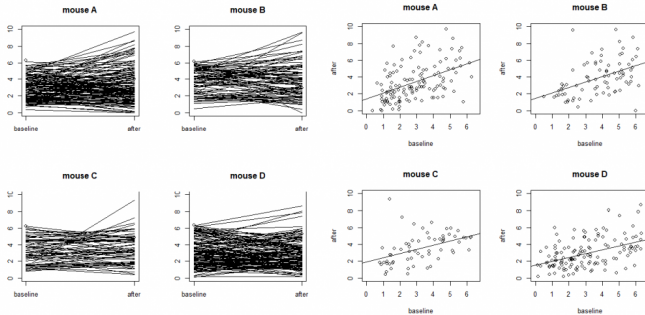


Figure 3: Ca⁺⁺ event integrated amplitudes at baseline vs 24h after treatment for the neurons from four mice (labeled as A, B, C and D) with each dot representing a neuron. The four plots on the left are “spaghetti” plots of the four animals with each line representing the values at baseline and 24h after treatment for a neuron; the four plots on the right report the before-after scatter plots (with fitted least-squares regression lines, respectively).

Compare Models with Different Random Effects

UC Irvine
ISI-BUDS Day 14

Zhaoxia Yu

LM, LME, GLM,
and GLMM

LME Examples:
Exmaple 2

LME Examples:
Example 3

GLMM

On models with more random effects

UC Irvine
ISI-BUDS Day 14

Zhaoxia Yu

LM, LME, GLM,
and GLMM

LME Examples:
Example 2

LME Examples:
Example 3

GLMM

- ▶ Tests can be used to compare models with different random effects
 - ▶ Need to be careful. See 6.4 of https://yu-zhaoxia.github.io/MM_in_Neuroscience/
- ▶ For example 3, the model I chose have the following random-effects:

```
“random=list(midx=~1, cidx=~treatment)”
```

- ▶ It improves lme.obj1 substantially.
- ▶ Adding more random-effects does not lead to much improvement

GLMM

Generalized Linear Mixed-Effects Model (GLMM)

UC Irvine
ISI-BUDS Day 14

Zhaoxia Yu

LM, LME, GLM,
and GLMM

LME Examples:
Example 2

LME Examples:
Example 3

GLMM

- ▶ The components of aGLMM:
 - ▶ fixed-effects: a linear predictor $X\beta$
 - ▶ random-effects: $Z\mathbf{u}$, where $\mathbf{u} \sim N(0, G)$. E.g., $G = \sigma_b^2 \mathbf{I}$.
 - ▶ a link function to connect $E(Y|X, \mathbf{u})$ and $X\beta$:

$$g(E(Y|X, \mathbf{u})) = X\beta + Z\mathbf{u}$$

- ▶ a distribution for Y

GLMM Examples: A Simulated Data Set

UC Irvine
ISI-BUDS Day 14

Zhaoxia Yu

LM, LME, GLM,
and GLMM

LME Examples:
Example 2

LME Examples:
Example 3

GLMM

- ▶ The simulation used parameters estimated from real data
- ▶ Eight mice were trained to do task
- ▶ The behavior outcome is whether the animals make the correct predictions
 - ▶ 512 trials in total: 216 correct trials, 296 wrong trials
- ▶ Mean neuronal activity levels (dF/F) were recorded for each trial
- ▶ We would like to model behaviors using neuronal data (decoding)

Use lme4::glmer to fit a GLMM

UC Irvine
ISI-BUDS Day 14

Zhaoxia Yu

LM, LME, GLM,
and GLMM

LM5: Fitting
Example 2

LME Examples:
Example 3

GLMM

```
library(lme4)
library(pbkrtest)
waterlick=read.table("https://www.ics.uci.edu/~zhaoxia/Data/B
summary(waterlick)
```

##	mouseID	lick	dff
##	Min. :1.000	Min. :0.0000	Min. : -8.838
##	1st Qu.:2.000	1st Qu.:0.0000	1st Qu.: 1.240
##	Median :4.500	Median :0.0000	Median : 4.702
##	Mean :4.527	Mean :0.4219	Mean : 4.810
##	3rd Qu.:6.000	3rd Qu.:1.0000	3rd Qu.: 8.426
##	Max. :8.000	Max. :1.0000	Max. :20.456

```
#change the mouseID to a factor
waterlick[,1]=as.factor(waterlick[,1])
```

Use lme4::glmer to fit a GLMM

UC Irvine
ISI-BUDS Day 14

Zhaoxia Yu

LM, LME, GLM,
and GLMM

LME Examples:
Example 2

LME Examples:
Example 3

GLMM

```
obj.glmm=glmer(lick~dff+(1|mouseID),  
data=waterlick,family="binomial")  
#summary(obj.glmm)  
#compute increase in odds and a 95% CI  
exp(c(0.06235, 0.06235-1.96*0.01986, 0.06235+1.96*0.01986))-1
```

```
## [1] 0.06433480 0.02370091 0.10658157
```

Interpret GLMM results

UC Irvine
ISI-BUDS Day 14

Zhaoxia Yu

LM, LME, GLM,
and GLMM

LME Examples:
Example 2

LME Examples:
Example 3

GLMM

- ▶ The estimate of odd is 6.4% increase and a 95% confidence interval is 2.3% to 10.7%
- ▶ The interpretation of the fixed effects for GLMM is complicated by both
 - ▶ the random effects and
 - ▶ non-linear link functions
- ▶ Among typical mice, the odds of making correct licks increased by 6.4% (95% C.I.: 2.4%-10.7%) with one unit increase in dF/F .

LRT test

UC Irvine
ISI-BUDS Day 14

Zhaoxia Yu

LM, LME, GLM,
and GLMM

LME Examples:
Example 2

LME Examples:
Example 3

GLMM

- Likelihood ratio test can be done by comparing the model with and the model without the “dff” variance (neuronal activity). Large-sample approximation is used.

```
#fit a smaller model, the model with the dff variable removed  
obj.glmm.smaller=glmer(lick~(1|mouseID),  
data=waterlick,family="binomial")  
#use the anova function to compare the likelihoods of the two  
anova(obj.glmm, obj.glmm.smaller)  
#alternatively, one can use the "drop1" function to test the  
drop1(obj.glmm, test="Chisq")
```

Improve accuracy of p-values

UC Irvine
ISI-BUDS Day 14

Zhaoxia Yu

LM, LME, GLM,
and GLMM

LME Examples:
Example 2

LME Examples:
Example 3

GLMM

- ▶ The large-sample approximations in GLMM might not be accurate
- ▶ We show how to conduct a parametric bootstrap test

#The code might take a few minutes

```
PBmodcomp(obj.glmm, obj.glmm.smaller)
```

- ▶ By default, 1000 samples were generated to obtain an empirical null distribution of the likelihood ratio statistic

Convergence Issues

UC Irvine
ISI-BUDS Day 14

Zhaoxia Yu

LM, LME, GLM,
and GLMM

LME Examples:
Example 2

LME Examples:
Example 3

GLMM

- ▶ GLMM is harder to converge than LME.
 - ▶ Increase the number of iterations
 - ▶ Switch to a different numerical maximization methods
 - ▶ Modify models such as eliminate some random effects

https://rstudio-pubs-static.s3.amazonaws.com/33653_57fc7b8e5d484c909b615d8633c01d51.html

<https://bbolker.github.io/mixedmodels-misc/glmmFAQ.html>

<https://m-clark.github.io/posts/2020-03-16-convergence/>

Convergence Issues

UC Irvine
ISI-BUDS Day 14

Zhaoxia Yu

LM, LME, GLM,
and GLMM

LME Examples:
Example 2

LME Examples:
Example 3

GLMM

- ▶ Consider more robust methods such generalized estimating equation (GEE)
- ▶ Oftentimes, Bayesian approaches are easier to converge