L4: LMM - Random intercept model and Repeated Measures ANOVA

BIOS6643

EJC

Department of Biostatistics & Informatics, CU Anschutz

Background

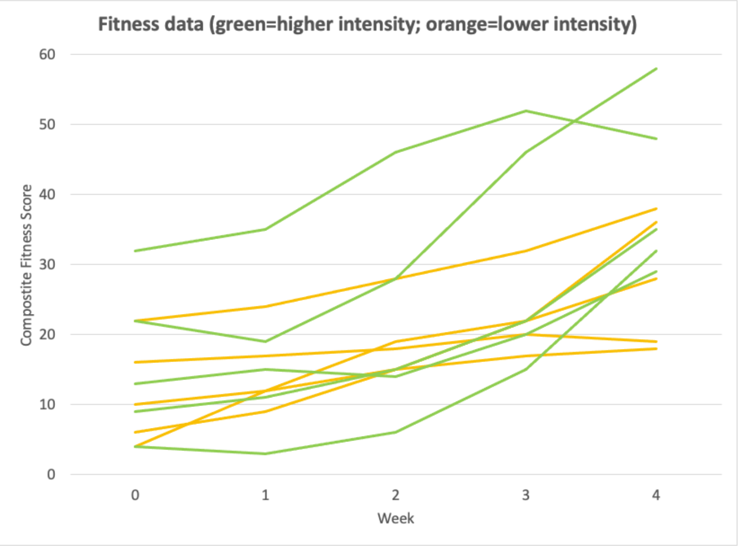
- One of the simplest ways to account for correlated data in a linear mixed model is to add a random intercept term.
  - For example, adding a random intercept term for subjects will induce correlation between measures within subjects, even when repeated measures are not accounted for in the error covariance matrix.
  - The covariance structure (referred to as compound symmetric) in this case is simplistic and often not realistic for longitudinal data (covariance between any pair of responses over time is the same regardless of the pair of time points being considered), but is far better than not accounting for correlation at all.
  - When the random intercept term is for, say, schools, then the covariance structure might be more realistic.

- For the random-intercept-for subjects model we assume the random intercepts (b_i) are drawn from a normal distribution with mean 0 and variance σ_b^2 (i.e., between-subject variance).
- When the error covariance matrix has the form $\sigma_\epsilon^2 I$, the model variance for a response at any time point is the sum of residual variance (or within-subject variance after accounting for fixed effects) and the between-subject variance.
- The correlation between any 2 time points is the intraclass correlation coefficient (ICC) $\sigma_b^2/(\sigma_b^2 + \sigma_\epsilon^2)$.
  - The analysis, or at least much of it, can be carried out using what is referred to Repeated Measures ANVOA (RM ANOVA), which has been around much longer than mixed models have, at least in practice. The model for the RM ANOVA can be considered as a special case of the LMM for balanced data.

Fitness data

- 10 subjects were randomized to one of two fitness programs, one lower intensity and the other higher.
  - Subjects were evaluated using an overall composite fitness score, which ranges from 0 to 75. Although it is an integer score, given the many possible levels, using a linear model has been shown to be adequate for the data.
  - Subjects were evaluated at baseline (Week 0), and then at 4 successive weeks after starting the program (e.g., Week 1 as at the end of the first week), making for 5 times points per subject.
- The fitness longitudinal data are shown below.
  - Data suggests that the lower intensity program has bigger gains in early weeks, while the higher intensity program has stronger gains in later weeks.
  - Data will be fit with a model to determine whether apparent differences are statistically significant.

Fitness data figure



Understanding variation in the data

- Analysis of variance (ANOVA) tables are intuitive, as they partition total (corrected) sums of squares into sources, providing a sense of relative amounts of variation in the data.
- Repeated measures ANOVA (RM ANOVA) uses the standard ANOVA approach, but makes adjustments to tests to account for the repeated measures taken within subjects.
- Linear mixed models can be used to achieve the same analysis, so there is no need to perform an RM ANOVA (in SAS via *PROC GLM*). The RM ANOVA just helps give us an intuitive understanding for the sources of variation.
- In general, inference in LMMs are not based on ANOVA tables, but in some cases like this one (for balanced data), inference is the same.

- In order to consider variation and the RM ANOVA approach, consider the following model.

$$Y_{hij} = \mu + \gamma_{h(\text{Group})} + \tau_{j(\text{Time})} + (\gamma\tau)_{hj(\text{Group}\times\text{Time})} + b_{i:h(\text{Subject:Group})} + \epsilon_{hij(\text{error})};$$

where $b_{i(h)} \overset{iid}{\sim} \mathcal{N}(0,\, \sigma_b^2)$ independent of $\epsilon_{hij} \overset{iid}{\sim} \mathcal{N}(0,\, \sigma_\epsilon^2)$.

- Sum to 0 restrictions can be placed on group (G), time (T) and $G \times T$ effects.
    - Although the subject term is random, for RM ANOVA, subjects within groups is treated as a fixed-effect term, at least initially (i.e., within *PROC GLM*, and *ID(PROGRAM)* is added as a term in the *MODEL* statement). This allows us to incorporate all sources of variation in the table.

Repeated measures ANOVA

The ANOVA table, including expected mean squares (E(MS)) may be written as follows.

- Note: Q_T is a function of time effects; the greater the value, the more the difference between τ_j parameters.
- Similar for $G, G \times T$ are functions of group and interaction effects.
- In this ANOVA table, n_h = number of subjects in group h, n_{tot} =total sample size.

Source	DF	SS	MS	E(MS)
G	$s-1$	$r \sum n_h(\bar{Y}_{h\bullet\bullet} - \bar{Y}_{\bullet\bullet\bullet})^2$	$SS_G/(s-1)$	$\sigma_\varepsilon^2 + r\sigma_b^2 + Q_G$
T	$r-1$	$n_{tot} \sum(\bar{Y}_{\bullet\bullet j} - \bar{Y}_{\bullet\bullet\bullet})^2$	$SS_T/(r-1)$	$\sigma_\varepsilon^2 + Q_T$
G×T	$(s-1)(r-1)$	$\sum\sum n_h(\bar{Y}_{h\bullet j} - \bar{Y}_{h\bullet\bullet} - \bar{Y}_{\bullet\bullet j} + \bar{Y}_{\bullet\bullet\bullet})^2$	$SS_{G*T}/[(s-1)(r-1)]$	$\sigma_\varepsilon^2 + Q_{GT}$
Subject(Group)	$n_{tot} - s$	$r \sum\sum(\bar{Y}_{hi\bullet} - \bar{Y}_{h\bullet\bullet})^2$	$SS_{S(G)}/(n_{tot}-s)$	$\sigma_\varepsilon^2 + r\sigma_b^2$
Residual	$(n_{tot}-s)(r-1)$	$\sum\sum\sum(Y_{hij} - \bar{Y}_{h\bullet j} - \bar{Y}_{hi\bullet} + \bar{Y}_{h\bullet\bullet})^2$	$SS_R/[(n_{tot}-s)(r-1)]$	σ_ε^2
Total corrected	$n_{tot}r - 1$	$\Sigma\Sigma\Sigma(Y_{hij} - \bar{Y}_{...})^2$		

Tests (based on sum-to-0 restrictions)

- Group $\times$ Time

  $H_0 : \forall\ (\gamma\tau)_{hj} = 0$ (or $Q_{GT} = 0$)
  Use $F = MS_{GT}/MS_R$

- Group

  $H_0 : \forall\ \gamma_h = 0$ (or $Q_G = 0$)
  Use $F = MS_G/MS_{S(G)}$

- Subject (i.e., Subject (Group))
  $H_0 : \sigma_b^2 = 0$
  Use $F = MS_{S(G)}/MS_R$

- Time $H_0 : \forall\ \tau_j = 0$ (or $Q_T = 0$)
  Use $F = MS_T/MS_R$

Estimating the ICC may be more informative than running a test for subject variance.

ANOVA table

The observed ANOVA table for our model is as follows.

				Uncorrected		Corrected	
Source	DF	SS	MS	F	P-value	F	P-value
Program	1	450.00	450.00	~~20.59~~	~~<0.0001~~	450.00/500.49=0.90	0.37
Time	4	2788.12	697.03	31.90	<0.0001		
Program×Time	4	199.40	49.85	2.28	0.0822		
Subject(Prg.)	8	4003.92	500.49	22.90	<0.0001		
Residual	32	699.28	21.85	20.03	<0.0001		
Total	49	8140.72					

- If you do not identify the repeated measures within subjects, the Program effect is much more significant than it should be (crossed out).
- Essentially, the Program effect is a between-subject effect, so it makes sense that the denominator MS for the corresponding F statistic is based on the Subject(Program) source of variability; this F statistic is much smaller. The other model sources use the standard residual source ($MS_{Residual}$) in the denominator of F.
- Subject(Program) allows us to estimate subject variability not due to Program effects; using Subject instead of Subject(Program) would not allow us to tease out Program variability from Subject variability.

- In general, the correct form of F can be guided by examining the expected mean squares.
- Under the null hypothesis of no effect for the source in question, the expected MS should be the same in the numerator and denominator.
  - For example, for the Group $\times$ Time test, the $E[MS]$ is $\sigma_\epsilon^2 + Q_{GT}$. Under the null, $Q_{GT} = 0$, reducing the quantity to σ_ϵ^2. Thus, the standard $MS_{Residual}$ is the correct SS in the denominator.
  - On the other hand, for group, $E[MS] = \sigma_\epsilon^2 + r\sigma_b^2 + Q_G$. Under the null, $Q_G = 0$, reducing the quantity to $\sigma_\epsilon^2 + r\sigma_b^2$, which is $E[MS_{Subject(group)}]$, showing that $MS_{Subject(group)}$ is the correct denominator term.

To carry out the RM ANOVA using PROC GLM, the basic code is shown below.

```
proc glm data=fitness;
 class program time id;
 model y=program time program*time id(program)
        / solution;
 random id(program) / test;
 output out=out1 p=pred_glm; run;
```

- The total variability in the data, 8140.72, is the sum of squared distances from the overall mean to the data points. This is also often called 'Corrected Total Sum of Squares', where the correction is for the mean.
- The first partition of the data sums of squares is into portions attributed to Model and Error, 7441.44 and 699.28, respectively, demonstrating that the model can account for a large portion of variation in the data.
    - Dividing these quantities by their respective degrees of freedom (17 and 32) yield Mean Square quantities of 437.7 and 21.85, respectively.

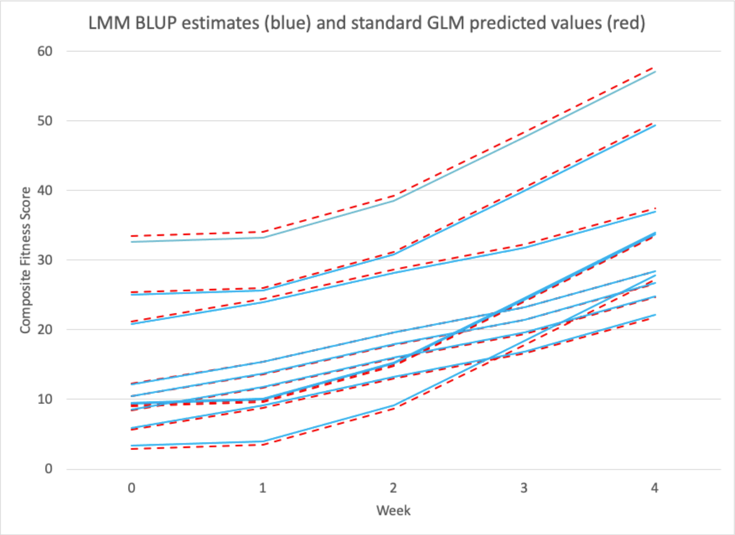
- Method of moments may be used to obtain estimates of variance components in terms of Mean Square quantities.
  - In particular, $E[MS_{Subject(Group)}] = \sigma_\epsilon^2 + r\sigma_b^2$ and $E[MS_{Residual}] = \sigma_\epsilon^2$
  - So we set the left side to MS quantities, put hats on variance terms on the right, and then solve for these estimated variance terms, to yield: $\hat{\sigma}_\epsilon^2 = MS_{Residual}$ and $\hat{\sigma}_b^2 = (MS_{Subject(Group)} - MS_{Residual})/r$
- For our data, these estimates are $\hat{\sigma}_\epsilon^2 = 21.8525$ and $\hat{\sigma}_b^2 = (500.49–21.8525)/5 = 95.7525.$
  - These variances show that between-subject variability is about 5 times larger than the within-subject variability that does not include variation due to the fixed effects.
- Two noticeable features in the data are the intercept variations in the 'noodles' and the increase in noodles over time;
  - these are the two greatest sources of variation in the mean-corrected SS: $2788/8140 = 34\%$ for Time and $4004/8140 = 49\%$ for subjects within programs, a total of 83% of variation in the data.

The LMM approach

- To fit the same model using an LMM, we treat subject as a true random effect (subject intercept here).
- Similarities between RM ANOVA and LMM approaches:
  - Using REML, the estimated variances are exactly the same as using method of moments with RM ANOVA.
  - The estimates of fixed effects for *Program*, *Time* and *Program* $\times$ *Time* effects are exactly the same.

- Differences between approaches
  - Subjects(Program) is treated as a fixed effect with the RM ANOVA approach, and random for the LMM approach. This leads to differences in subject-specific estimates.
  - Specifically, since empirical Bayes methods are used to estimate random effects for subjects, the predicted values for subjects that incorporate random effect estimates will be shrunk back to the overall mean to some degree, relative to the RM ANOVA (*PROC GLM*) estimates that treat subject effect as fixed effects.
  - This is demonstrated in the following graph; the LMM estimates are in solid blue and the GLM estimates are in dashed red.
    - The differences in this case are not great, but clearly the LMM estimates are compressed to the middle relative to the *PROC GLM* estimates.
    - The more reliable subject data are (more values, less variability), the less the shrinkage. This probably explains the small amount of shrinkage here.

Best Linear Unbiased Prediction (BLUP)



Degrees of freedom

- For any other differences between RM ANOVA and LMM, it is recommended to use the latter, since subjects are modeled using random effects, i.e., they are considered as having been sampled from a normal population, and inference properly accounts for this.
- If we are interested in inference just for the sample of subjects used, it makes sense to treat them as fixed effects, but usually we are more interested in the general population they were sampled from.
- Differences in approaches is reflected in the lower standard errors of estimates in the RM ANOVA approach relative to the LMM approach.
  - With RM ANOVA, Subjects-within-Programs is modeled as a fixed effect; hence inference is conditioned on the particular subjects at hand.
  - For the LMM approach, inference for fixed effects is based on the marginal model (averaging over subjects in the population), naturally (and appropriately) leading to larger SE's.

PROC MIXED and options

- Basic SAS code to fit the model above is shown below. The *OUTP* will provide *BLUP* estimates for each value in the data set ($\hat{\boldsymbol{Y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} + \boldsymbol{Z}\boldsymbol{b}$), while *OUTPM* provides $\hat{\boldsymbol{Y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}$.
- The *LSMEANS* statement will provide estimates for each *Program* $\times$ *time* combination.
- Adding the *'diff'* option to the right of the slash will provide comparisons between all pairs of differences in these combinations.
- There are also options to control for multiple testing using the Adjust option. For more detail, see the SAS Help Documentation.
- The *'solution'* option in the *RANDOM* statement provides estimates and t-tests for random effect estimates (the same solution option could be added in the *MODEL* statement, but the *LSMEANS* options gives us what we need in this case).

```
proc mixed data=fitness;
class program time id;
model y=program time program*time / outp=out2 outpm=out3;
random intercept / subject=id(program) solution;
lsmeans program*time / cl; run;
```

Abbreviated SAS output (REML; containment method for DF):

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
Intercept	id	95.7275
Residual		21.8525

Solution for Random Effects

Effect	id	Estimate	SE Pred	DF	t Value	Pr > |t|
Intercept	1	-3.0220	4.7423	32	-0.64	0.5285
Intercept	2	0.6121	4.7423	32	0.13	0.8981
Intercept	3	9.2191	4.7423	32	1.94	0.0607
Intercept	4	-5.6998	4.7423	32	-1.20	0.2382
Intercept	5	-1.1094	4.7423	32	-0.23	0.8165
Intercept	6	9.0278	4.7423	32	1.90	0.0660
Intercept	7	16.6785	4.7423	32	3.52	0.0013
Intercept	8	-6.4648	4.7423	32	-1.36	0.1823
Intercept	9	-12.5854	4.7423	32	-2.65	0.0123
Intercept	10	-6.6561	4.7423	32	-1.40	0.1701

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
program	1	8	0.90	0.3708
time	4	32	31.90	<.0001
program*time	4	32	2.28	0.0822

LSMEANS for program*time (SE=4.8493, DF=32)

program	time	Estimate	t Value	Pr > |t|	Lower	Upper
a	0	11.6000	2.39	0.0228	1.7222	21.4778
a	1	14.8000	3.05	0.0045	4.9222	24.6778
a	2	19.0000	3.92	0.0004	9.1222	28.8778
a	3	22.6000	4.66	<.0001	12.7222	32.4778
a	4	27.8000	5.73	<.0001	17.9222	37.6778
b	0	16.0000	3.30	0.0024	6.1222	25.8778
b	1	16.6000	3.42	0.0017	6.7222	26.4778
b	2	21.8000	4.50	<.0001	11.9222	31.6778
b	3	31.0000	6.39	<.0001	21.1222	40.8778
b	4	40.4000	8.33	<.0001	30.5222	50.2778

- One methodological difference in fitting LMM's is that in order to conduct inference, we develop statistical quantities that have approximate t or F distributions, and then estimate the denominator degrees of freedom (often referred to as DDF or DDFM) to conduct 'correct' inference.
- There are 6 or 7 different methods that can be used, and in SAS, the default methods used will depend on how the model is specified.
  - This will be discussed more later, but for our purposes now, an important thing to realize is that SAS and R have different default methods, which is why results may appear slightly different.
- For the SAS code above, the default DDFM method used is the 'Containment', since there is a RANDOM statement.
  - In order to change the method, you can add the DDFM option to the right of the slash in the MODEL statement.

LMM in R

- Basic R code using the **lme4::lmer()** function from the **lme4** package.
- Note that there are 3 basic DDFM methods available, two approximate (Satterthwaite, Kenward-Roger), and asymptotic.
- Asymptotic is not recommended for smaller data sets, as it is likely to lead to inflated Type I error rates and CI's that are too narrow.

```
library(lme4)
library(emmeans)
library(lmerTest)  #Allows for satterthwaite df
library(pbkrtest)  #Allows for Kenward-Roger df

runny <- lmer(y ~ (time*program) + (1|id), data=fitness_dat)
summary(runny)
emmeans(runny, ~ time*program, lmer.df="satterthwaite")
```

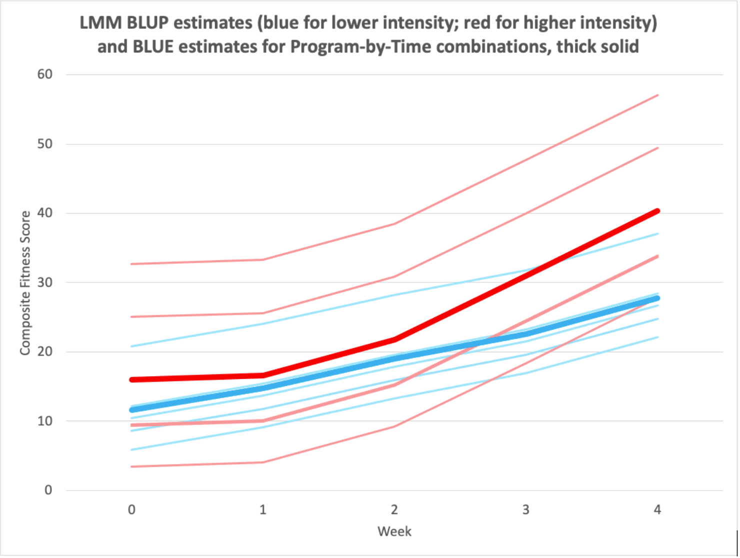
```
> emmeans(runny, ~time*program, lmer.df="satterthwaite")
 time program emmean   SE   df lower.CL upper.CL
 _0   a         11.6 4.85 10.9    0.921     22.3
 _1   a         14.8 4.85 10.9    4.121     25.5
 _2   a         19.0 4.85 10.9    8.321     29.7
 _3   a         22.6 4.85 10.9   11.921     33.3
 _4   a         27.8 4.85 10.9   17.121     38.5
 _0   b         16.0 4.85 10.9    5.321     26.7
 _1   b         16.6 4.85 10.9    5.921     27.3
 _2   b         21.8 4.85 10.9   11.121     32.5
 _3   b         31.0 4.85 10.9   20.321     41.7
 _4   b         40.4 4.85 10.9   29.721     51.1
Degrees-of-freedom method: satterthwaite
Confidence level used: 0.95

> aov <- anova(runny)
> aov
Type III Analysis of Variance Table with Satterthwaite\'s method
             Sum Sq Mean Sq NumDF DenDF F value    Pr(>F)
time        2788.12  697.03     4    32 31.8970 9.419e-11 ***
program       19.65   19.65     1     8  0.8991   0.37078
time:program 199.40   49.85     4    32  2.2812   0.08218 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

- Note that the results above match the RM ANOVA approach for the balanced fitness data.
- The estimated marginal means (**emmeans()**) are the same as the LSMEANS from SAS's approach.
- The only difference between SAS and R analyses is in the CI's and p-values, which is due to different DDFM methods used (Satterthwaite here, Containment using SAS).

Summary of DDF's for different DDFM methods for Fitness data, with the Random intercept model

Term	Contain/BW	Sat/KR	Residual
Program	8*	8	40
Time	32	32, 10.9	40
Program*Time	32	32, 10.9	40

*For SAS, need to specify ID(program) as subject in the RANDOM statement.


LMM BLUP estimates (blue for lower intensity; red for higher intensity) and BLUE estimates for Program-by-Time combinations, thick solid
Composite Fitness Score
60
50
40
30
20
10
0
0
1
2
3
4
Week

Summary