Australian National University

Comparison between linear mixed model (LMM) and analysis of variance (ANOVA) for separation of within- and between-subject effects

Jin Yoon
Statistical Consulting Unit
The Australian National University
15 July 2011

Data Structure

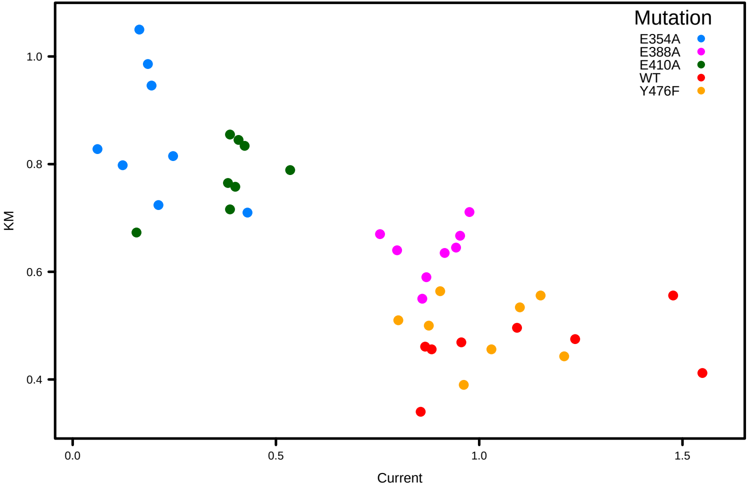
SEQ	Mutation	Current(x)	KM(y)
1	E354A	0.061	0.828
⋮	⋮	⋮	⋮
8	E354A	0.164	1.05
9	E410A	0.157	0.673
⋮	⋮	⋮	⋮
32	Y476F	1.209	0.443
33	WT	0.867	0.461
⋮	⋮	⋮	⋮
40	WT	0.883	0.456

Model

- Mixed models are now becoming widely used in biology.
- Evolutionary and ecological hypotheses often specifically address either within- or between-subject differences
- Statistical models which produce unambiguous and unbiased estimates of such within- and between-subject effects
- Some confusion concerning the extent to which fixed independent variables quantify within- and between-subject effects on response variables.

Problems

- Fixed independent variables, x, as well as dependent variables, y, can vary at multiple levels of aggregation due to the nonexperimental studies.
- Although random effects(intercept) in mixed models can account for between-subject variation in y, they do not automatically account for between-subject variation in x.
- An association between x and y could be caused by a within-subject effect of x on y, but also by a between-subject effect of x on y.


Mutation
E354A
E388A
E410A
WT
Y476F
KM
1.0
0.8
0.6
0.4
0.0
0.5
1.0
1.5
Current

Within-Subject Centering

- The standard random effects model is given by

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \gamma_j + \varepsilon_{ij} \tag{1}$$

  where x_{ij} is the value of measurement i from subject j.
- γ_j is random effect of subject j.
- $\gamma_j \sim \mathbf{N}(0, \sigma^2_{\gamma_j})$ and $\varepsilon_{ij} \sim \mathbf{N}(0, \sigma^2_{\varepsilon_{ij}})$
- Within-subject centering: subtracting the subject's mean value from each observation value, $x_{ij} - \bar{x}_j$.

Within-Subject Centering

- Centering around subjects' means effectively get rid of any between-subject effect variation
- Derived new independent variable $(x_{ij} - \bar{x}_j)$ to use as a fixed effect that expresses only the within-subject variation component.
- Derived a second new fixed independent variable, $\bar{x}_j$, for only between-subject variation component.

Within-Subject Centering

- The new model with these new fixed effects is given by

$$y_{ij} = \beta_0 + \beta_W(x_{ij} - \bar{x}_j) + \beta_B\bar{x}_j + \gamma_j + \varepsilon_{ij}. \tag{2}$$

- Allow to test whether either the within-subject effect, β_W, or the between-subject effect, β_B is itself significant.
- If these two effects seem to differ, then want to compare their estimated effects, β_W and β_B, to see whether they are statistically different from each other.

Within-Subject Centering

- We can rewrite the model (2)

$$y_{ij} = \beta_0 + \beta_W x_{ij} + (\beta_B - \beta_W)\bar{x}_j + \gamma_j + \varepsilon_{ij}. \qquad (3)$$

- Since all these statistical models test for each effect while controlling for all other effects in the model, the within-subject effects, β_W in (2) and (3) be identical.
- Between-subject effect in (3) represents the difference between the between- and within-subject effect, $\beta_B - \beta_W$.

Proposed method

- Use ANOVA adjusted with covariate
- No treatment structure
- Subject (or group) as block structure

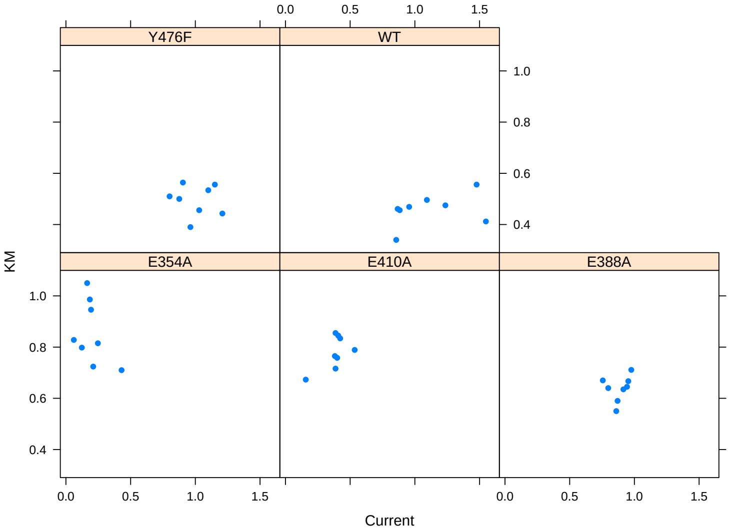
Example

- Biochemistry data from an invitro expression system
- In this experiment, interested in studying the effect of mutations of a transporter protein on the KM values as a parameter of transporter kinetics and transport rate (electrical current).
- Create 4 mutations (1 wild type) with sample size of 8 in order to study which highly conserved areas of the protein are essential for the activity of protein.

Results

- Parameter estimates of models (1)-(3) and ANOVA with covariate
- p-value < 0.05 (Red colour)

Parameter	Model(1)	Model(2)	Model(3)	ANOVA
β_0	0.646(0.08)	0.646(0.02)	0.646(0.02)	0.645
β_1	-0.019(0.08)			-0.019(0.08)
β_W		0.044(0.08)	0.044(0.08)	0.044(0.08)
β_B		-0.423(0.05)		-0.423(0.05)
$\beta_B - \beta_W$			-0.467(0.10)	
σ_γ^2	0.028	0.001	0.001	0.028
σ_ε^2	0.006	0.006	0.006	0.006


Y476F
WT
E354A
E410A
E388A
KM
Current
0.0
0.5
1.0
1.5
0.4
0.6
0.8
1.0

Results

- Although model (1) even suggested there was no association between x and y, whereas there was an effect of between-subject.
- For the case of balanced data, ANOVA(adjusted with covariate) gives within- and between-subject effects and combined effect as well.
- No need two new variables to estimate within- and between-subject effects

Results

- For unbalanced case, model (2) and (3) are the same but model (1) and ANOVA give slightly different parameter estimates

Parameter	Model(1)	Model(2)	Model(3)	ANOVA
β_0	0.610(0.07)	0.647(0.03)	0.647(0.03)	0.645
β_1	-0.064(0.08)			-0.067(0.08)
β_W		-0.018(0.08)	-0.018(0.08)	-0.018(0.08)
β_B		-0.429(0.08)		-0.423(0.07)
$\beta_B - \beta_W$			-0.411(0.11)	
σ_γ^2	0.020	0.002	0.002	0.020
σ_ε^2	0.007	0.007	0.007	0.007

Conclusion

- For balanced cases, don't need centering for separate the within- and between-subject effects. Simply use ANOVA (adjusted for covariate) for parameter estimates
- Don't need new variables
- Combine ANOVA with covariate and REML for variance component
- Need further research for the unbalanced cases