

Longitudinal Data Analysis

Adam J Sullivan, PhD

04/25/2018

Estimation

REML vs. ML

Linear mixed model fit by REML ['lmerMod']

Formula: Reaction ~ 1 + Days + (1 + Days | Subject)

Data: sleepstudy

REML criterion at convergence: 1744

Scaled residuals:

Min	1Q	Median	3Q	Max
-3.954	-0.463	0.023	0.463	5.179

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
Subject	(Intercept)	612.1	24.74	
	Days	35.1	5.92	0.07
Residual		654.9	25.59	

REML vs. ML

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	251.41	6.82	36.8
Days	10.47	1.55	6.8

Correlation of Fixed Effects:

(Intr)
Days -0.138

REML vs ML

```
model.full.ml <- update(model.full, REML=FALSE)
```

REML vs ML

Formula: Reaction ~ 1 + Days + (1 + Days | Subject)

Data: sleepstudy

AIC	BIC	logLik	deviance	df.resid
1764	1783	-876	1752	174

Scaled residuals:

Min	1Q	Median	3Q	Max
-3.942	-0.466	0.029	0.464	5.179

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
Subject	(Intercept)	565.5	23.78	
	Days	32.7	5.72	0.08
Residual		654.9	25.59	

REML vs ML

Number of obs: 180, groups: Subject, 18

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	251.41	6.63	37.9
Days	10.47	1.50	7.0

Correlation of Fixed Effects:

(Intr)
Days -0.138

REML vs. ML

- **REML:** Restricted Maximum Likelihood
- Variance is the average squared distance to the
- Variance measured to the is less than true variance in finite samples
- This is the motivation behind Bessel's correction ($n - 1$ in the denominator instead of n when calculating variance using SS)

REML vs. ML

- Similar motivation for using REML in estimation of mixed-effects models: more accurate estimation of variance (and hence random effects!)
- "log likelihood" REML-fitted models dependent on parameterization and thus not comparable across models
- similarly: AIC, BIC, other measures of fit not comparable across models
- REML-models more accurate but not comparable with each other:
 - determine model structure with comparisons between ML-estimates
 - present final model with REML-estimates
 - `lme4` has some built-in protections to prevent REML-based comparisons, but don't depend on these!

Comparing our Models with ML and LRT

```
m1 <- update(model.intercepts, REML=F)  
m2 <- update(model.full, REML=F)  
anova(m1,m2)
```

Comparing our Models with ML and LRT

```
## Data: sleepstudy
## Models:
## m1: Reaction ~ 1 + Days + (1 | Subject)
## m2: Reaction ~ 1 + Days + (1 + Days | Subject)
##      Df  AIC   BIC logLik deviance Chisq Chi Df Pr(>Chisq)
## m1   4 1802 1815   -897     1794
## m2   6 1764 1783   -876     1752  42.1      2 7.1e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Comparing our Models with ML and LRT

```
m1 <- update(model.slopes, REML=F)  
m2 <- update(model.full, REML=F)  
anova(m1,m2)
```

Comparing our Models with ML and LRT

```
## Data: sleepstudy
## Models:
## m1: Reaction ~ 1 + Days + (0 + Days | Subject)
## m2: Reaction ~ 1 + Days + (1 + Days | Subject)
##      Df  AIC   BIC logLik deviance Chisq Chi Df Pr(>Chisq)
## m1   4 1782 1795   -887     1774
## m2   6 1764 1783   -876     1752  22.1      2  0.000016 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Covariance Structures

Variance Structure

- The simplest structure is below

$$\begin{bmatrix} \sigma_1^2 & \dots & \dots & \vdots \\ \vdots & \sigma_1^2 & \dots & \vdots \\ \vdots & \dots & \sigma_1^2 & \vdots \\ \dots & \dots & \dots & \sigma_1^2 \end{bmatrix}$$

- In this model between time points has 0 covariances. This is not usually the case with longitudinal data.

Compound Symmetry

$$\sigma^2 \begin{bmatrix} 1.0 & \rho & \rho & \rho \\ & 1.0 & \rho & \rho \\ & & 1.0 & \rho \\ & & & 1.0 \end{bmatrix} = \begin{bmatrix} \sigma_b^2 + \sigma_e^2 & \sigma_b^2 & \sigma_b^2 & \sigma_b^2 \\ & \sigma_b^2 + \sigma_e^2 & \sigma_b^2 & \sigma_b^2 \\ & & \sigma_b^2 + \sigma_e^2 & \sigma_b^2 \\ & & & \sigma_b^2 + \sigma_e^2 \end{bmatrix}$$

- The simplest covariance structure that includes within-subject correlated errors is compound symmetry (CS).
- Here we see correlated errors between time points within subjects, and note that these correlations are presumed to be the same for each set of times, regardless of how distant in time the repeated measures are made.

First Order Autoregressive AR(1)

$$\sigma^2 \begin{bmatrix} 1.0 & \rho & \rho^2 & \rho^3 \\ & 1.0 & \rho & \rho^2 \\ & & 1.0 & \rho \\ & & & 1.0 \end{bmatrix}$$

- The autoregressive (Lag 1) structure considers correlations to be highest for time adjacent times, and a systematically decreasing correlation with increasing distance between time points.
- For one subject, the error correlation between time 1 and time 2 would be $\rho^{t_1 - t_2}$.

First Order Autoregressive AR(1)

$$\sigma^2 \begin{bmatrix} 1.0 & \rho & \rho^2 & \rho^3 \\ & 1.0 & \rho & \rho^2 \\ & & 1.0 & \rho \\ & & & 1.0 \end{bmatrix}$$

- Between time 1 and time 3 the correlation would be less, and equal to $\rho^{t_1-t_3}$.
- Between time 1 and 4, the correlation is less yet, as $\rho^{t_1-t_4}$, and so on.
- Note, however, that this structure is only applicable for evenly spaced time intervals for the repeated measure.

Spatial Power

$$\sigma^2 \begin{bmatrix} 1.0 & \rho^{\frac{|t_1-t_2|}{|t_1-t_2|}} & \rho^{\frac{|t_1-t_3|}{|t_1-t_2|}} & \rho^{\frac{|t_1-t_4|}{|t_1-t_2|}} \\ & 1.0 & \rho^{\frac{|t_2-t_3|}{|t_1-t_2|}} & \rho^{\frac{|t_2-t_4|}{|t_1-t_2|}} \\ & & 1.0 & \rho^{\frac{|t_3-t_4|}{|t_1-t_2|}} \\ & & & 1.0 \end{bmatrix}$$

- When time intervals are not evenly spaced, a covariance structure equivalent to the AR(1) is the spatial power (SP(POW)).
- The concept is the same as the AR(1) but instead of raising the correlation to powers of 1, 2, 3, ... , the correlation coefficient is raised to a power that is actual difference in times (e.g. $|t_1 - t_2|$ for the correlation between time 1 and time 2).

Spatial Power

$$\sigma^2 \begin{bmatrix} 1.0 & \rho^{\frac{|t_1-t_2|}{|t_1-t_2|}} & \rho^{\frac{|t_1-t_3|}{|t_1-t_2|}} & \rho^{\frac{|t_1-t_4|}{|t_1-t_2|}} \\ & 1.0 & \rho^{\frac{|t_2-t_3|}{|t_1-t_2|}} & \rho^{\frac{|t_2-t_4|}{|t_1-t_2|}} \\ & & 1.0 & \rho^{\frac{|t_3-t_4|}{|t_1-t_2|}} \\ & & & 1.0 \end{bmatrix}$$

- This method requires having a quantitative expression of the times in the data so that it can be specified for calculation of the exponents in the SP(POW) structure. If an analysis is run wherein the repeated measures are equally spaced in time, the AR(1) and SP(POW) structures yield identical results.

Unstructured

$$\begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ & \sigma_2^2 & \sigma_{23} & \sigma_{24} \\ & & \sigma_3^2 & \sigma_{34} \\ & & & \sigma_4^2 \end{bmatrix}$$

- The Unstructured covariance structure (UN) is the most complex because it is estimating unique correlations for each pair of time points.
- It is not uncommon to find out that you are not able to use this structure.
- R will return an error message indicating that there are too many parameters to estimate with the data.

How can we fit these?

- If we use the `lme4` package and the `lmer()` function then we do not need to suggest a correlation structure for these models.
- If we use the `nlme` package and the `lme()` function, then we can specify the covariance structure and compare the models using LRT in order to see which covariance structure works the best.

Case Study

Case Study: Predicting College GPA

- We will assess factors predicting college grade point average (GPA).
- Each of the 200 students is assessed twice a year for the first three years of their education.
- We also have other variables such as job status, sex, high school GPA, whether they have been admitted to a program of choice.

The Data

Viewing the Data

What do we have?

- All student paths are shown in blue
- Sample of 10 shown in orange.
- Population Regression shown in red.

What can you see?

Basic Linear Model

```
gpa_lm = lm(gpa ~ occasion, data=gpa)
```

Basic Linear Model

	ESTIMATE	STD. ERROR	T VALUE	PR(> T)
(Intercept)	2.599	0.018	145.7	0
occasion	0.106	0.006	18.04	0

Fitting linear model: gpa ~ occasion

OBSERVATIONS	RESIDUAL STD. ERROR	R^2	ADJUSTED R^2
1200	0.3487	0.2136	0.2129

What do we see?

- We can see that on average students begin with a GPA of 2.6
- Then comparing students 1 semester apart, the student one semester further along would have on average a 0.11 increase in GPA.

Problems?

- We do not meet the assumptions for regular linear regression.
- We noticed a lot of variability at different time points that might not be captured.

Random Intercept Model

```
library(lme4)
gpa_mixed = lmer(gpa ~ occasion + (1|student), data=gpa)
summary(gpa_mixed)
```

Random Intercept Model

TERM	ESTIMATE	STD.ERROR	STATISTIC
(Intercept)	2.599	0.022	119.800
occasion	0.106	0.004	26.096
GRP	VARIANCE		SD
student	0.064		0.252
Residual	0.058		0.241

What do we see?

- We can see that on average students begin with a GPA of 2.6
- Then comparing students 1 semester apart, the student one semester further along would have on average a 0.11 increase in GPA.
- We can see that the standard deviation for student error is 0.252 and the residual standard deviation is 0.241.

Are there differences??

- There are different standard errors.
- Intercept standard error increased
 - this could suggest that we actually underestimated it before and now with the random effects model we can explore that variation more.

Why no p-values?

- There are many problems with p-values in a mixed effects setting.
- Read this [site](#)
- We can get confidence intervals though.

Confidence intervals

```
confint(gpa_mixed)
```

Confidence Intervals

	2.5%	97.5%
student	0.225	0.282
residual	0.231	0.252
Intercept	2.557	2.642
occasion	0.098	0.114

Random Effects

```
ranef(gpa_mixed)$student %>% head(5)  
coef(gpa_mixed)$student %>% head(5)
```


Random Effects

(INTERCEPT)

-0.071

-0.216

0.088

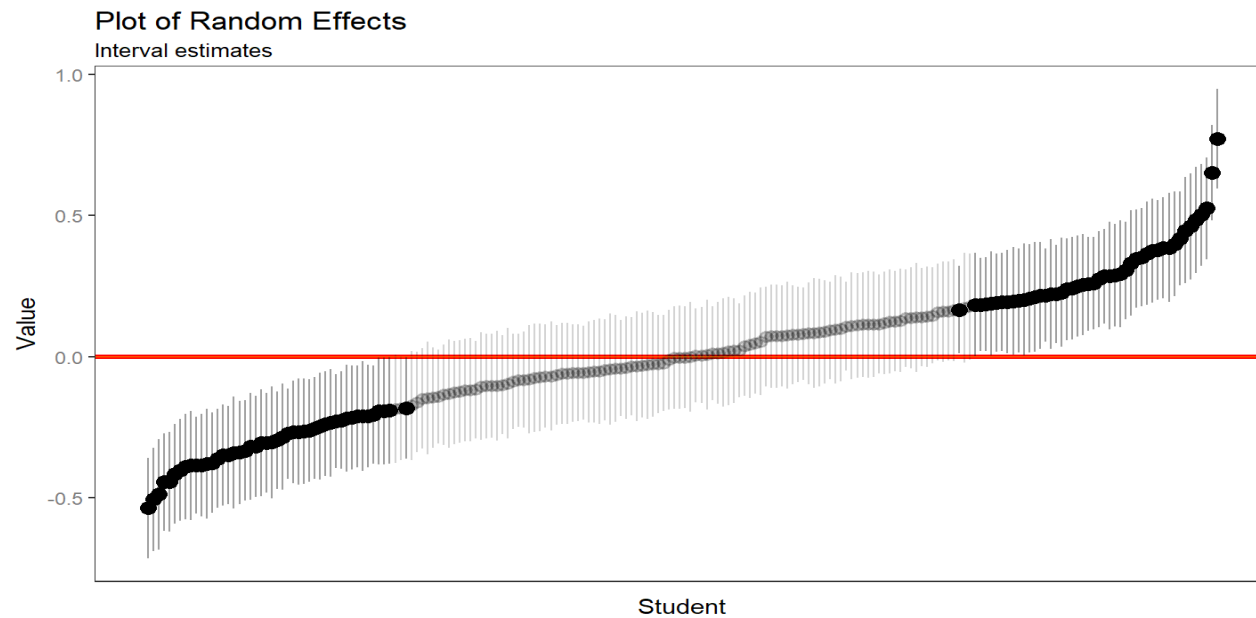
-0.187

0.030

Random Effects

(INTERCEPT)	OCCASION
2.528	0.106
2.384	0.106
2.687	0.106
2.413	0.106
2.630	0.106

Random Effects and Confidence Intervals for Each



Predicting GPA

- `re.form=NA` is ignoring the random effects

```
predict_no_re = predict(gpa_mixed, re.form=NA)  
predict_lm = predict(gpa_lm)
```

Predicting GPA

Predicting GPA with Mixed Effects added

Random Slopes and Intercepts

```
gpa_mixed = lmer(gpa ~ occasion + (1 + occasion|student), data=gpa)  
summary(gpa_mixed)
```

Random Slopes and Intercepts

TERM	ESTIMATE	STD.ERROR	STATISTIC	CONF.LOW	CONF.HIGH
(Intercept)	2.599	0.018	141.592	2.563	2.635
occasion	0.106	0.006	18.066	0.095	0.118
GRP	RE	VARIANCE		SD	
student	(Intercept)	0.045		0.213	
student	occasion	0.005		0.067	
Residual		0.042		0.206	

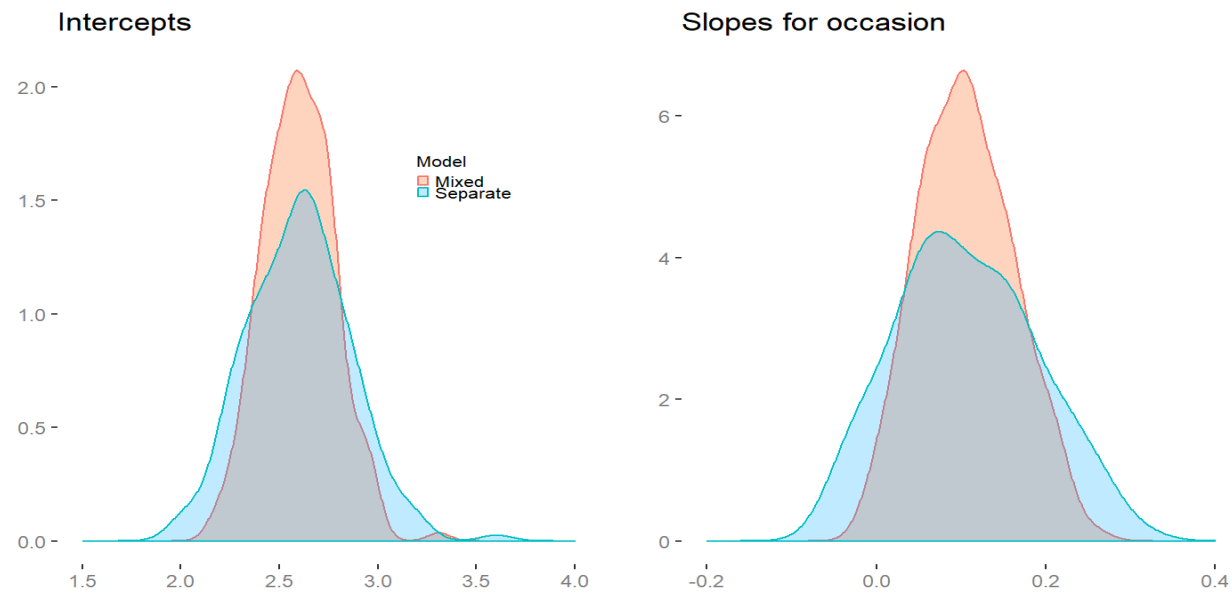
What do we see?

- We can see that on average students begin with a GPA of 2.6
- Then comparing students 1 semester apart, the student one semester further along would have on average a 0.11 increase in GPA.
- We can see that the standard deviation for student error is 0.213 and the residual standard deviation is 0.206 and the standard deviation of the occasion is 0.067.

Are there differences??

- There are different standard errors than other mixed model but same as linear.

Comparison Again to Separate Regression



Visualizing Model Differences

Further Differences

Which Model Should we use?

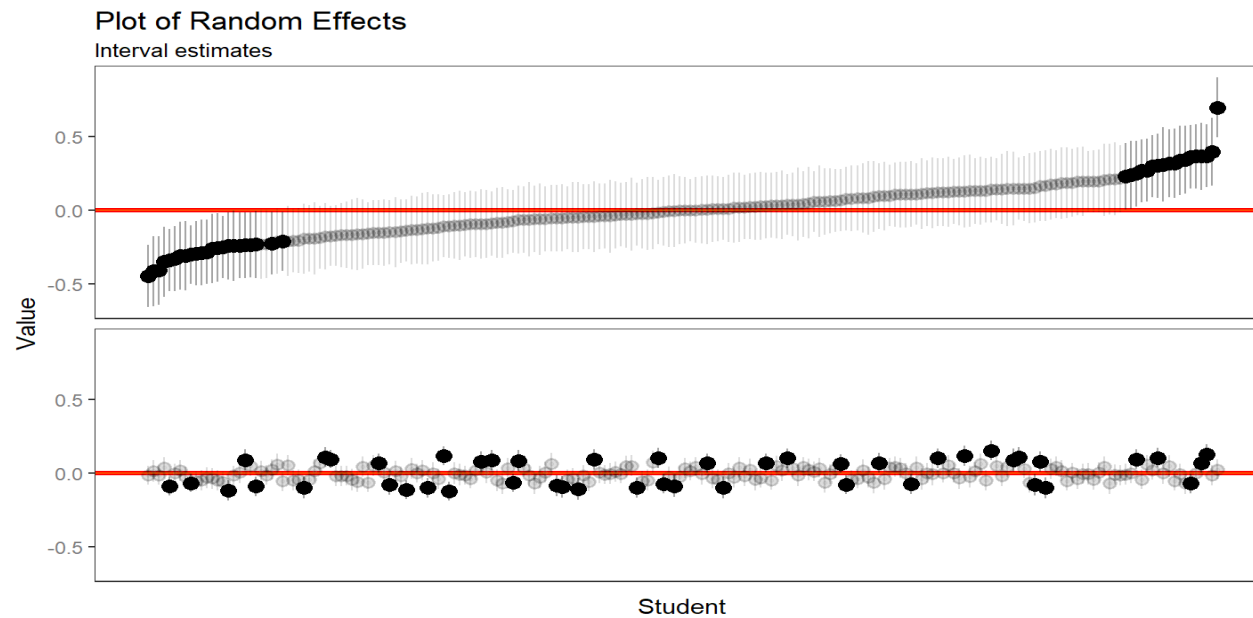
```
m1 <- lmer(gpa ~ occasion + (1|student), data=gpa, REML=F)
m2 <- lmer(gpa ~ occasion + (1 + occasion|student), data=gpa, REML=F)

anova(m1,m2)
```

Which Model Should we use?

```
## Data: gpa
## Models:
## m1: gpa ~ occasion + (1 | student)
## m2: gpa ~ occasion + (1 + occasion | student)
##      Df AIC BIC logLik deviance Chisq Chi Df Pr(>Chisq)
## m1   4 402 422   -197      394
## m2   6 258 289   -123      246   147    2   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Random Effects and Confidence Intervals for Each



Predictions