Linear Mixed Effects Model

Monte Carlo Group

Liang Zou, Michael Mcbride, Xinlu Zhang
Caleb Donoho, Ashley Lu

22 January 2018

Introduce the definition of LMM

Linear mixed effects models simply model the fixed and random effects as having a linear form. Similar to the Generalized Linear Model, an outcome variable is contributed to by additive fixed and random effects (as well as an error term). Using the familiar notation, the linear mixed effect model takes the form:

$$y_{ij} = b_0 + b_1 x_{ij} + v_i + \epsilon_{ij}$$

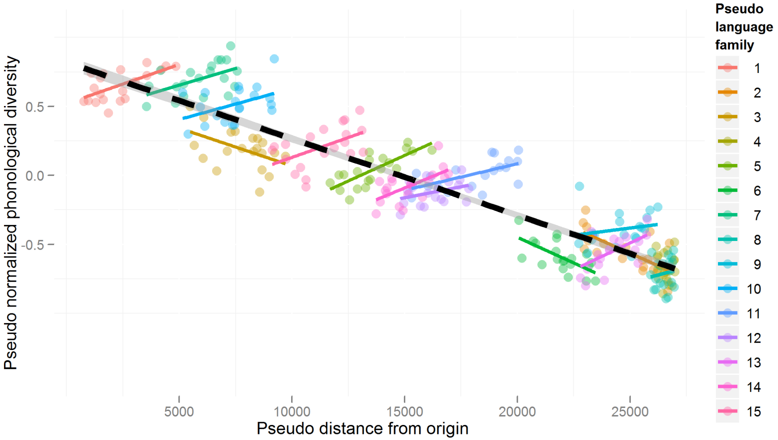
- y_{ij} is the response value for a particular ij case.
- b_0 is fixed intercept for regression model.
- b_1 is fixed slope for regression model.
- x_{ij} is fixed-effect variable for observation j-th measurement of i-th subject.
- v_i is random intercept of i-th subject.
- ϵ_{ij} is a Gussian error term which assumed to be multivariate normally distributed.

Assumption of the models

- The relationship between x and y is linear
- Y is observed random variable.
- X is a design matrix for fixed effect.
- $v_i \sim N(0, \sigma_v^2)$ is an unobserved random noise.
- $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$ is an unobserved random noise.
- v_i and ϵ_{ij} are independent of each other.
- b_0 and b_1 are unknown constants.

When and Why LMM is necessary and carried out.

We use Linear Mixed Effect Models when we believe that the data comes from partitions on the sample space, that each partition is internally correlated, and that the partition means are determined by some global parameters plus some noise. This type of situation is illustrated on the graph shown below (image obtained from here):



When and Why LMM is necessary and carried out.

As we can see, the mean within each group follows the larger line, while the data within each group follows it's own sub-population line. In situations like these we might want to be able to make both population and sub-population inferences, or discuss how much the grouping affects the variation. Using an LMM allows us do all of these things. Essentially, we use LMMs when we believe the response variable is sampled from different distributions, the parameters for which are sampled from a parent distribution, and we want our inference to be reflective of this model.

Statistical Analysis packages

There are several packages in R, which contains for fitting LMMs, for instance nlme, lme4.0, or MCMCglmm. For use in R, it should be noted that lme4 and nlme do not interact well with each other.

```
require(lme4)
citation("lme4")
require(nlme)
citation("nlme")
require(MCMCglmm)
citation("MCMCglmm")
```

Statistical Analysis packages

Two main differences between “lme4” and “nlme”:

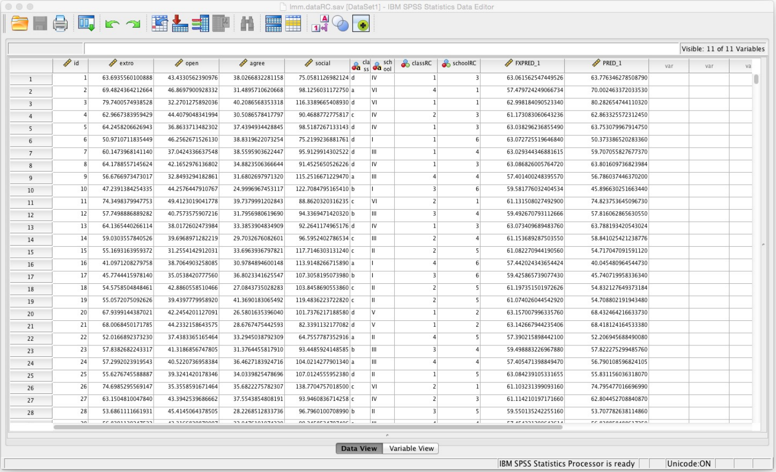
- “nlme” can only fit outcomes with normal distribution
- “lme4” can use other link functions (e.g. logistic regression)
- “nlme” is flexible with more variance-covariance structures
- “lme4” can only specify diagonal and unstructured covariance matrices

Statistical Analysis packages: MCMCglmm

- Estimate the parameters of random effects via MCMC
- Two types of MCMC: Gibbs sampler and Metropolis-Hasting
- General procedure of MCMC:
- Set up initial values
- Run the chain multiple times until they become stable
- Approximation

Perform dataset

We are using the dataset called Imm.dataRC and using SPSS to perform the analysis.



	id	extro	open	agree	social	class	school	classRC	schoolRC	FXPRED_1	PRED_1	var	var	va
1	1	63.6935560100888	43.4330562390976	38.0266832281158	75.0581126982124	d	IV	1	3	63.061562547449526	63.776346278508790			
2	2	69.4824364212664	46.8697900928332	31.4895710620668	98.1256031172750	a	VI	4	1	57.479724249066734	70.002463372033530			
3	3	79.7400574938528	32.2701275892036	40.2086568353318	116.3389665408930	d	VI	1	1	62.998184090523340	80.282654744110320			
4	4	62.9667383959429	44.4079048341994	30.5086578417797	90.4688772775817	c	IV	2	3	61.173083060643236	62.863325572312450			
5	5	64.2458206626943	36.8633713482302	37.4394934428845	98.5187267133143	d	IV	1	3	63.038296236855490	63.753079967914750			
6	6	50.9710711835449	46.2562671526130	38.8319622073254	75.2199236881761	d	I	1	6	63.072725519646840	50.373386520283360			
7	7	60.1473968141140	37.0424336637548	38.5595903622447	95.9129914302522	d	III	1	4	63.029344346881615	59.707055827677370			
8	8	64.1788557145624	42.1652976136802	34.8823506366644	91.4525650526226	d	IV	1	3	63.086826005764720	63.801609736823984			
9	9	56.6766973473017	32.8493294182861	31.6802697971320	115.2516671229470	a	III	4	4	57.401400248395570	56.786037446370200			
10	10	47.2391384254335	44.2576447910767	24.9996967453117	122.7084795165410	b	I	3	6	59.581776032404534	45.896630251663440			
11	11	74.3498379947753	49.4123019041778	39.7379991202843	88.8620320316235	c	VI	2	1	61.131508027492900	74.823753645096730			
12	12	57.7498886889282	40.7573575907216	31.7956980619690	94.3369471420320	b	III	3	4	59.492670793112666	57.816062865630550			
13	13	64.1365440266114	38.0172602473984	33.3853904834909	92.2641174965176	d	IV	1	3	63.073409689483760	63.788193420543024			
14	14	59.0303557840526	39.6968971282219	29.7032676082601	96.5952402786534	c	III	2	4	61.153689287503550	58.841025421238776			
15	15	55.1693163959372	31.2554142912031	33.6963936797821	117.7146303131240	c	II	2	5	61.082270944190560	54.717047091591120			
16	16	41.0971208279758	38.7064903258085	30.9784894600148	113.9148266715890	a	I	4	6	57.442024343654424	40.045480964544730			
17	17	45.7744415978140	35.0538420777560	36.8023341625547	107.3058195073980	b	I	3	6	59.425865739077430	45.740719958336340			
18	18	54.5758504848461	42.8860558510466	27.0843735028283	103.8458690553860	c	II	2	5	61.197351501972626	54.832127649373184			
19	19	55.0572075092626	39.4397779958920	41.3690183065492	119.4836223722820	c	II	2	5	61.074026044542920	54.708802191943480			
20	20	67.9399144387021	42.2454201127091	26.5801635396040	101.7376217188580	d	V	1	2	63.157007996335760	68.432464216633730			
21	21	68.0068450171785	44.2332158643575	28.6767475442593	82.3391132177082	d	V	1	2	63.142667944235406	68.418124164533380			
22	22	52.0166892373230	37.4383365165464	33.2945038792309	64.7557787352916	a	II	4	5	57.390215898442100	52.206945688490080			
23	23	57.8382682243317	41.3186856747805	31.3764455817910	93.4485924148585	b	III	3	4	59.498883226967880	57.822275299485760			
24	24	57.2992023919543	40.5220736958384	36.4627183924716	104.0214277901340	a	III	4	4	57.405471398849470	56.790108596824105			
25	25	55.6276745588887	39.3241420178346	34.0339825478696	107.0124555952380	d	II	1	5	63.084239105331655	55.831156036318070			
26	26	74.6985295569147	35.3558591671464	35.6822275782307	138.7704757018500	c	VI	2	1	61.103231399093160	74.795477016696990			
27	27	63.1504810047840	43.3942539686662	37.5543854808191	93.9460836714258	c	IV	2	3	61.114210197171660	62.804452708840870			
28	28	53.6861111661931	45.4145064378505	28.2268512833736	96.7960100708990	b	II	3	5	59.550135242255160	53.707782638114860			

Perform dataset

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	id	Numeric	3	0		None	None	8	Right	Scale	Input
2	extro	Numeric	16	13		None	None	16	Right	Scale	Input
3	open	Numeric	16	13		None	None	16	Right	Scale	Input
4	agree	Numeric	16	13		None	None	16	Right	Scale	Input
5	social	Numeric	16	13		None	None	16	Right	Scale	Input
6	class	String	1	0		None	None	4	Left	Nominal	Input
7	school	String	3	0		None	None	5	Left	Nominal	Input
8	classRC	Numeric	8	0		None	None	10	Right	Nominal	Input
9	schoolRC	Numeric	8	0		None	None	10	Right	Nominal	Input
10	FXPRED_1	Numeric	19	15	Fixed Predicte...	None	None	21	Right	Scale	Input
11	PRED_1	Numeric	19	15	Predicted Values	None	None	21	Right	Scale	Input
12											

Perform dataset

- The data contains 1200 cases evenly distributed among 24 nested groups (4 classes within 6 schools).
- extro: the interval scaled outcome variable Extroversion.
- open: predicted by fixed effects for the interval scaled predictor Openness to new experiences.
- agree: the interval scaled predictor Agreeableness.
- social: the interval scaled predictor Social engagement.
- classRC: the nominal scaled predictor Class.
- schoolRC: the random (nested) effect of Class (classRC) within School (schoolRC) as well as the random effect of School.

Syntax and interpretation of Linear Mixed Effect Model.

The Case Processing Summary simply shows that the cases are balanced among the categories of the categorical variables and no cases were excluded.

Case Processing Summary

		Count	Marginal Percentage
classRC	1	300	25.0%
	2	300	25.0%
	3	300	25.0%
	4	300	25.0%
schoolRC	1	200	16.7%
	2	200	16.7%
	3	200	16.7%
	4	200	16.7%
	5	200	16.7%
	6	200	16.7%
Valid		1200	100.0%
Excluded		0	
Total		1200	

Syntax and interpretation of Linear Mixed Effect Model.

Rather large table contains all the descriptive statistics (only the very top of the table is shown here).

Descriptive Statistics

classRC	schoolRC		Count	Mean	Standard Deviation	Coefficient of Variation
1	1	extro	50	80.3780775	2.72223072	3.4%
		open	50	40.3949348	5.99143304	14.8%
		agree	50	36.2490009	6.16716305	17.0%
		social	50	101.347351	17.6249114	17.4%
	2	extro	50	68.3509948	.438343995	0.6%
		open	50	40.2310796	6.13149141	15.2%
		agree	50	34.4292989	5.65646971	16.4%
		social	50	98.4141077	15.4740068	15.7%
	3	extro	50	63.7827007	.281376636	0.4%
		open	50	40.4489301	5.18116491	12.8%
		agree	50	35.5121035	4.70341338	13.2%
		social	50	100.324500	14.2921193	14.2%
	4	extro	50	59.7253379	.259824951	0.4%
		open	50	38.3536854	5.69767114	14.9%
		agree	50	36.4827588	4.75289937	13.0%
		social	50	102.088786	15.2093760	14.9%
	5	extro	50	55.8117921	.242368327	0.4%
		open	50	40.2527442	5.70058013	14.2%
		agree	50	35.8178298	4.19656365	11.7%
		social	50	100.216916	14.9726997	14.9%
	6	extro	50	50.3883065	.639401003	1.3%
		open	50	40.1017695	4.74579412	11.8%
		agree	50	34.5376332	4.45036364	12.9%
		social	50	95.4782588	18.3242857	19.2%
	Total	extro	300	63.0728683	9.68321522	15.4%
		open	300	39.9638573	5.59578837	14.0%
		agree	300	35.5047708	5.05514415	14.2%
		social	300	99.6449864	16.0645459	16.1%
2	1	extro	50	74.8053293	1.20904041	1.6%
		open	50	40.8589429	5.34758369	13.1%
		agree	50	35.9456349	4.30277384	12.0%
		social	50	99.8706241	16.3033499	16.3%

Syntax and interpretation of Linear Mixed Effect Model.

The Model Dimension table simply shows the model in terms of which variables (and their number of levels) are fixed and / or random effects and the number of parameters being estimated

Model Dimension[a]

		Number of Levels	Covariance Structure	Number of Parameters
Fixed Effects	Intercept	1		1
	open	1		1
	agree	1		1
	social	1		1
	classRC	4		3
Random Effects	classRC(schoolRC) + schoolRC[b]	30	Variance Components	2
Residual				1
Total		38		10

a. Dependent Variable: extro.

b. As of version 11.5, the syntax rules for the RANDOM subcommand have changed. Your command syntax may yield results that differ from those produced by prior versions. If you are using version 11 syntax, please consult the current syntax reference guide for more information.

Syntax and interpretation of Linear Mixed Effect Model.

The table displays fit indices. For each index; the lower the number, the better the model fits the data.

Information Criteria[a]

-2 Restricted Log Likelihood	3528.106
Akaike's Information Criterion (AIC)	3534.106
Hurvich and Tsai's Criterion (AICC)	3534.126
Bozdogan's Criterion (CAIC)	3552.359
Schwarz's Bayesian Criterion (BIC)	3549.359

The information criteria are displayed in smaller-is-better form.

a. Dependent Variable: extro.

Syntax and interpretation of Linear Mixed Effect Model.

This table is our Estimates of Fixed Effect. This is what we'd expect to see in a normal linear regression model with our β values, their standard error, degrees of freedom, t-value, significance, and 95% interval.

Estimates of Fixed Effects[a]

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	57.383879	4.056589	5.299	14.146	.000	47.130810	67.636947
open	.006130	.004963	1173.270	1.235	.217	-.003608	.015868
agree	-.007736	.005698	1173.203	-1.358	.175	-.018916	.003444
social	.000531	.001852	1173.359	.287	.774	-.003103	.004165
[classRC=1]	5.665733	.983721	15.001	5.759	.000	3.568999	7.762468
[classRC=2]	3.704930	.983709	15.001	3.766	.002	1.608212	5.801648
[classRC=3]	2.054798	.983726	15.002	2.089	.054	-.041945	4.151541
[classRC=4]	0[b]	0	.	.	.	.	.

a. Dependent Variable: extro.

b. This parameter is set to zero because it is redundant.

Syntax and interpret of Linear Mixed Effect Model.

- The RC variables contain the same information as the original variables, they simply have been recoded.
- SPSSautomatically choose the category with the highest numerical value (or the lowest alphabetical letter) as the reference category for categorical variables.
- In the lme4 package in R, the software automatically picks the lowest numerical value (or the earliest alphabetically letter) as the reference category for categorical variables.

Syntax and interpretation of Linear Mixed Effect Model.

Covariance Parameters

Estimates of Covariance Parameters[a]

Parameter		Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Residual		.968368	.039986	24.218	.000	.893085	1.049998
classRC(schoolRC)	Variance	2.883600	1.060023	2.720	.007	1.402910	5.927070
schoolRC	Variance	95.171929	60.651592	1.569	.117	27.293005	331.868776

a. Dependent Variable: extro.

Correlation Matrix for Estimates of Covariance Parameters[a]

Parameter		Residual	classRC (schoolRC) Variance	schoolRC Variance
Residual		1	-.001	.000
classRC(schoolRC)	Variance	-.001	1	-.004
schoolRC	Variance	.000	-.004	1

a. Dependent Variable: extro.

Syntax and interpretation of Linear Mixed Effect Model.

- We have the Covariance Matrix for the Estimates of Fixed effects table to determine the covariance between our fixed effects. The only significant covariances we see is between our class variables with themselves which is to be expected.
- The Correlation Matrix for Estimates of Fixed Effects table shows us a similar story as our covariance matrix where we find relatively small and insignificant correlations between our variables except between our classes.

Conclusion

- The LMM model is a useful tool when observing both fixed and random effects as in world data.
- In particular, our use of the LMM is able to be useful in addressing nested effects such as with classes within schools.
- To create LMM, SPSS is able to easily build and return descriptive statistics to address the fixed effects.
- In our interpretation of LMM, we have multiple tables to focus on the fixed effects in quantity and confidence intervals as our random effects are known via assumption.
- We should intend to use a LMM when we sample from subgroups which have different distributions determined by the parent group.

Q&A

Reference

- http://bayes.acs.unt.edu:8083/BayesContent/class/Jon/SPSS_SC/Module9/M9_LMM/SPSS_M9_LMM.htm
- http://users.stat.umn.edu/~helwig/notes/lmer-Notes.pdf
- http://www.bodowinter.com/tutorial/bw_LME_tutorial2.pdf