

Mixed Effects Models

D2K Course Staff

July 25 2019

Introduction

Mixed effects models are often used in cases where we want to take repeated measurements or nuisance variables in to account without treating them as a fixed variable in the final model. The model will generally look something like:

$$Y_i = \mu + \alpha Z_{j(i)} + \beta X_i + \epsilon_i, i \in 1, \dots, n, j \in 1, \dots, m.$$
$$\epsilon_i \sim N(0, \sigma^2)$$

In this case, we have n total individual observations which are each in one of m groups of a variable in the data. The x_i are called fixed effects and are the variables that we want to measure the effect of in our final model. The $Z_{j(i)}$ are the random effects that we want to account for but do not wish to do inference on; each individual i is a member of one of the groups j . μ is the intercept term that we normally include in regression models.

This will probably make more sense with the example below. We will analyze the `nlschools` data set from the `MASS` package, which contains data from students in schools in the Netherlands. The variables in the data:

- `lang`: language test score.
- `IQ`: verbal IQ score.
- `class`: class ID.
- `GS`: class size.
- `SES`: socioeconomic status.
- `COMB`: multi-grade class?

We will use the `lme4` package to fit our mixed effects models.

```
library(MASS)
#?nlsschools
data(nlsschools)
head(nlsschools)
```

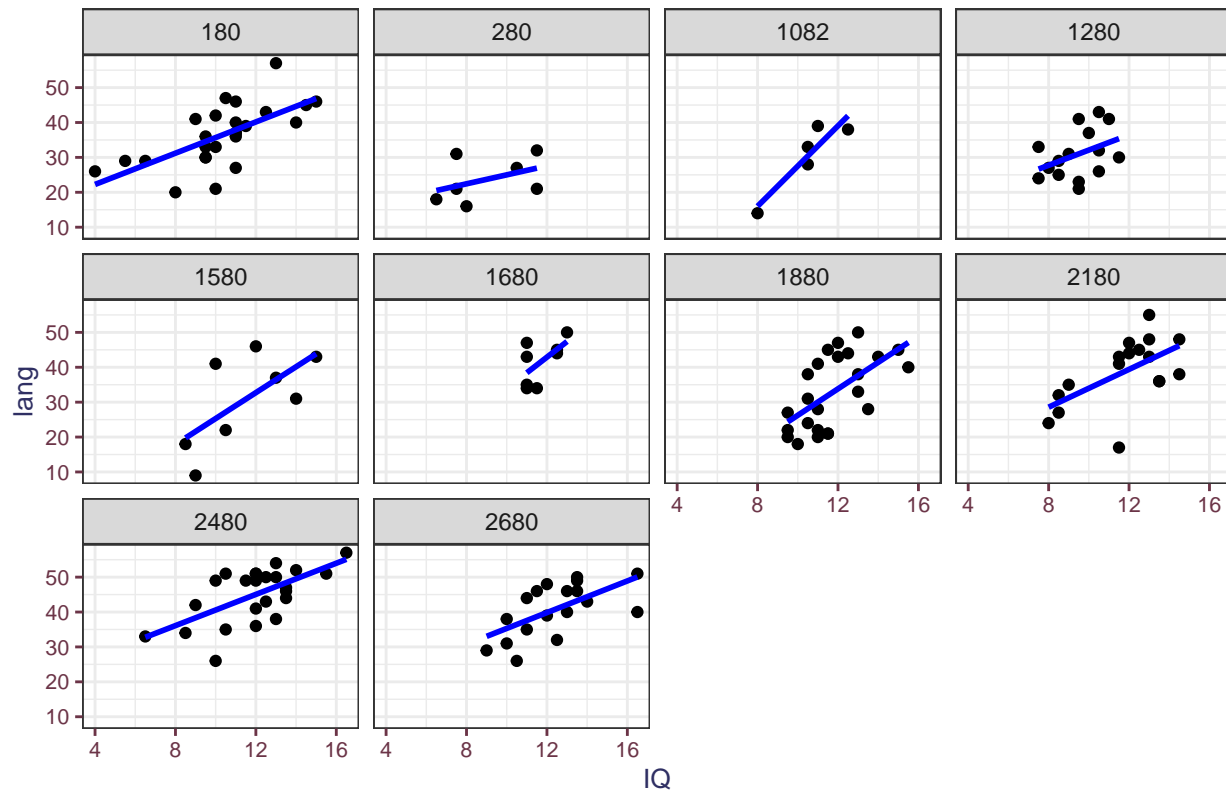
```
   lang   IQ class GS SES COMB
1    46 15.0   180 29  23     0
2    45 14.5   180 29  10     0
3    33  9.5   180 29  15     0
4    46 11.0   180 29  23     0
5    20  8.0   180 29  10     0
6    30  9.5   180 29  10     0
```

Intuition

Say for the `nlsschools` data we want to predict a student's language test score based on their verbal IQ score, controlling for the other variables in the data set. However, it seems that students were sampled in blocks from individual classes. Thus, we have repeated measures on individuals from the same class. This would be a case where we would include class ID as a random effect. We certainly want to take the class ID in to account in our model, as there may be class-level effects on the language test score that are not directly measured in the other variables in the data; perhaps the teachers for certain classes are better or worse than others, or perhaps some classes are honors-level courses while others are remedial level. Indeed, the plots below appear to show that the relationship between a student's language test score and their verbal IQ score appears to be different for the different classes. However, because of block sampling scheme, we can't just include this as a fixed effect since these class IDs do not represent all possible classrooms in the Netherlands. In particular, if we wanted to predict for a student in a new class not in the data, a model with class ID as a fixed effect could not be used. Thus, we include class ID as a random effect.

```
ggplot(data = nlsschools[1:151, ]) +
  geom_point(aes(x = IQ, y = lang)) +
  geom_smooth(aes(x = IQ, y = lang), se = FALSE,
              method = "lm", color = "blue") +
  facet_wrap(~ class) +
  labs(title = "Effect by School") +
  theme1
```

Effect by School



How are these random effects accounted for?

Random effects are most often included in a mixed effects model as a parametric distribution around a “global” underlying average effect. One might think of this as a sampling distribution for the random effect. For example, we might assume that the classroom level effect on the relationship between a student’s language test score and their verbal IQ score is normally distributed with mean 0 and variance τ^2 around the true underlying parameter values.

Random Intercept

The first model example we will look at is a random intercept model. This model assumes that the class-level effect changes the average language score of the individual students. Mathematically, this is:

$$Y_i = \mu + \alpha_{j(i)} + \Lambda W_i + \beta x_i + \epsilon_i, i \in 1, \dots, n, j \in 1, \dots, m.$$
$$\epsilon_i \sim N(0, \sigma^2), \alpha_j \sim N(0, \tau^2).$$

- μ is the global intercept - α is the class-level random intercept effect - x_i is the fixed effect of interest - W_i are the individual level covariates - ϵ_i is the individual level error

The results are shown below. Note the syntax of the formula - we specify a random intercept by adding (1|class). We can see the estimates for the coefficients for IQ as well as the covariates under “Fixed effects:”, and we can see the estimates for τ^2 and σ^2 under “Random effects:”.

```
library(lme4)

model11 <- lmer(lang ~ IQ + GS + SES + COMB + (1|class), data = nlschools)
summary(model11)
```

```
Linear mixed model fit by REML ['lmerMod']
Formula: lang ~ IQ + GS + SES + COMB + (1 | class)
Data: nlschools
```

REML criterion at convergence: 15132.9

Scaled residuals:

	Min	1Q	Median	3Q	Max
	-3.9588	-0.6488	0.0486	0.7204	3.0765

Random effects:

Groups	Name	Variance	Std.Dev.
class	(Intercept)	8.507	2.917
Residual		39.998	6.324

Number of obs: 2287, groups: class, 133

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	10.63524	1.50012	7.090
IQ	2.24696	0.07133	31.502
GS	-0.01562	0.04770	-0.327
SES	0.16538	0.01479	11.185
COMB1	-2.02161	0.60055	-3.366

Correlation of Fixed Effects:

	(Intr)	IQ	GS	SES
IQ		-0.484		
GS	-0.798		-0.002	
SES	-0.062	-0.292		-0.057
COMB1	-0.168	0.033	-0.006	0.017

Random Slope

The first model example we will look at is a random intercept model. This model assumes that the class-level effect changes the linear relationship between the language score and verbal IQ of the individual students. Mathematically, this is:

$$Y_i = \mu + \alpha_{j(i)}x_i + \Lambda W_i + \beta x_i + \epsilon_i, i \in 1, \dots, n, j \in 1, \dots, m.$$
$$\epsilon_i \sim N(0, \sigma^2), \alpha_j \sim N(0, \tau^2).$$

- μ is the global intercept - α is the class-level random slope effect - x_i is the fixed effect of interest - W_i are the individual level covariates - ϵ_i is the individual level error

The results are shown below. We specify a random slope WITHOUT a random intercept by adding (IQ - 1|class). Again, we can see the estimates for the coefficients for IQ as well as the covariates under “Fixed effects:”, and we can see the estimates for τ^2 and σ^2 under “Random effects:”.

```
library(lme4)
```

```
model12 <- lmer(lang ~ IQ + GS + SES + COMB + (IQ - 1|class), data = nlschools)
summary(model12)
```

Linear mixed model fit by REML ['lmerMod']

Formula: lang ~ IQ + GS + SES + COMB + (IQ - 1 | class)

Data: nlschools

REML criterion at convergence: 15157.4

Scaled residuals:

	Min	1Q	Median	3Q	Max
	-3.9450	-0.6453	0.0590	0.7063	3.4214

Random effects:

Groups	Name	Variance	Std.Dev.
class	IQ	0.0561	0.2369
Residual		40.5701	6.3695

Number of obs: 2287, groups: class, 133

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	10.70203	1.43466	7.460
IQ	2.26638	0.07481	30.295
GS	-0.02471	0.04466	-0.553
SES	0.16820	0.01484	11.334
COMB1	-2.45973	0.56308	-4.368

Correlation of Fixed Effects:

	(Intr)	IQ	GS	SES
IQ		-0.508		
GS	-0.805		0.037	
SES	-0.063	-0.284		-0.060
COMB1	-0.140	-0.031	0.009	0.020

Model Comparison & Inference

We should not use the t-values returned by `lmer` to do statistical inference. Period. Instead, we should use some sort of likelihood ratio testing between different potential models. There are a couple of methods to do this:

ANOVA

ANOVA can be used in situations where we want to test for the statistical significance of individual or sets of variables. This is done by fitting 2 models: one with the variable(s) of interest and one without. These are then compared formally using an F-test, one of the classic likelihood ratio tests. The key in this case is that one model is nested in the other, meaning that the variables included in the smaller model must be a subset of the variables included in the larger model. Below, we fit a model without IQ before we run the ANOVA.

```
model3 <- lmer(lang ~ GS + SES + COMB + (1|class), data = nlschools)
anova(model3, model1)
```

```
Data: nlschools
```

```
Models:
```

```
model3: lang ~ GS + SES + COMB + (1 | class)
```

```
model1: lang ~ IQ + GS + SES + COMB + (1 | class)
```

	Df	AIC	BIC	logLik	deviance	Chisq	Chi	Df	Pr(>Chisq)
model3	6	15956	15990	-7971.9	15944				
model1	7	15133	15173	-7559.3	15119	825.09		1	< 2.2e-16 ***

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The results are shown above. In this example, the model with IQ is significantly better than the one without, based on the p-value from the ANOVA output.

AIC/BIC

Information criterion tests are used in the case where we do not have nested models. The upside is, of course, that we can more flexibly compare two models. The downside is that there is no formal statistical test or standard for these types of tests to show that one model is definitively better than the other. We can say in general that a smaller AIC or BIC is better; by most standards, a difference of 6 is generally considered enough evidence that one model is better than another. Here, we first create a non-nested model by creating a random independent variable, then fit a model with said random variable and without IQ.

```
nlschools$random <- rnorm(nrow(nlschools))
model4 <- lmer(lang ~ GS + SES + COMB + (1|class) + random, data = nlschools)
AIC(model4, model1)
```

	df	AIC
model4	7	15967.21
model1	7	15146.86

```
BIC(model4, model1)
```

	df	BIC
model4	7	16007.35
model1	7	15187.00

For both of these criteria, the AIC and BIC of the model without IQ is much larger than the respective measures on the model with IQ. There is almost certainly enough evidence to say that the model with IQ is better.

Extensions

Obviously, we can add more random effects to the model. In the particular example above, we could add both a random slope and intercept at the class level using `(IQ|class)`. Our data could have multiple levels of hierarchies (e.g., students within classes within schools within districts...) to add as random effects. We could also use different distributions or parameterizations for the random effects. The possibilities are endless!

Mixed effects models can also be applied to regression models outside of ordinary linear regressions. In particular, `glmer` in the `lme4` package allows for generalized linear mixed effects models (analogous to the `glm` extension of `lm`).