北京航空航天大学
经济管理学院
BEIHANG UNIVERSITY
SCHOOL OF ECONOMICS AND MANAGEMENT

Generalized Linear Models

Lecture 8: Hierarchical models and longitudinal data

Outline

Linear Mixed Models (LMM)

- Handle data where observations are not independent.
- Uncorrelated error is important but often violated.
- Violations occur when error terms are not independent but instead cluster by one or more grouping variables.
    - For instance, predicted student test scores and errors in predicting them may cluster by classroom, school, and municipality.
- LMM can lead to substantially different conclusions compared to conventional regression analysis.

Linear Mixed Models (LMM)

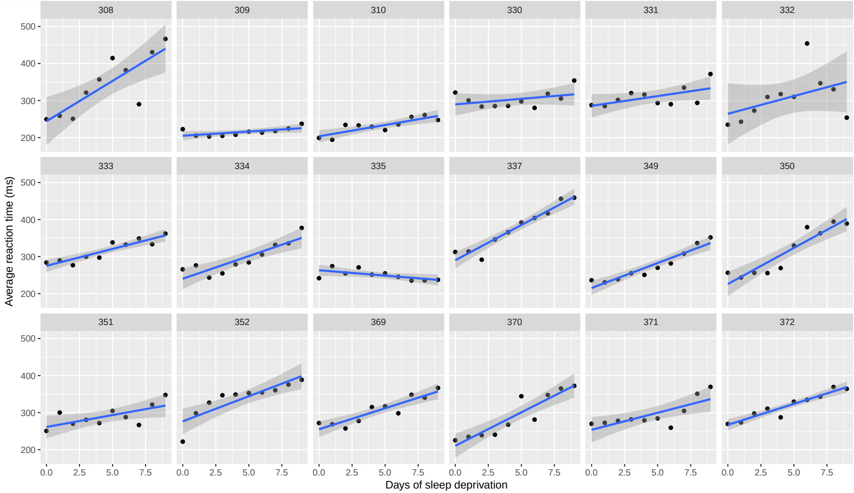
Linear mixed models are a generalization of general linear models to better support analysis of a continuous dependent variable for *random effects*.

Bear in mind:

Fixed effects: variables that we expect will have an effect on the dependent/response variable.

Random effects: grouping factors and they are categorical. Often we are not specifically interested in their impact on the response variable. Additionally, the data for our random effect is just a sample of all the possibilities.

A revisit to `sleepstudy` data



Question

Why not include dummy variables?

Question

Why not include dummy variables?

Benefits of using random effects

1. We can generalize to a wider population.
2. Fewer parameters are needed.
3. Information can be shared between groups.

Linear Mixed Models (LMM)

1. *Hierarchical effects:* When variables are measured at more than one level. Assess the effects of higher levels on the intercepts and coefficients at the lowest level.
    - scores at student level and teacher–student ratios at school level
    - sentencing lengths at the offender level, gender of judges at the court level, and budgets of judicial districts at the district level
2. *Repeated measures:* For when observations are correlated rather than independent. e.g., before–after studies, time series data.

Outline

Nested effects

Levels of one factor vary only within levels of another factor.

- Workers within job locations
- Units within campus

Crossed effects

- Any non-nested effects are “crossed”.
- That is, every level of one factor can potentially interact with every level of another factor.
- Incomplete crossing occurs when not all combinations of factors exist in the data.

Hierarchical linear models (HLM)

- Models with nested (hierarchical) structure.
- Commonly used in psychology, education, and other social sciences where survey data is naturally clustered hierarchically.

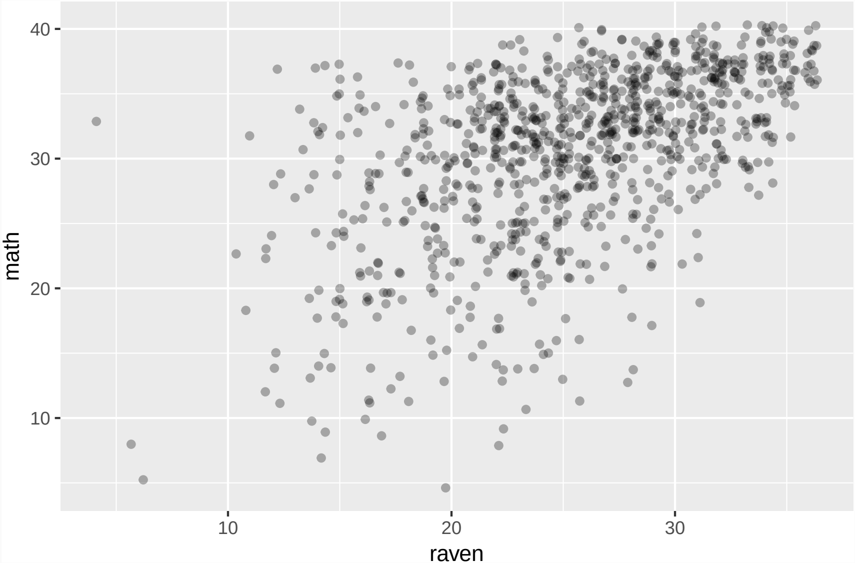
Junior School Project (1988)

Variables: `student`, `class`, `school`, `gender`, `social`, `raven`, `math`, `english`, `year`

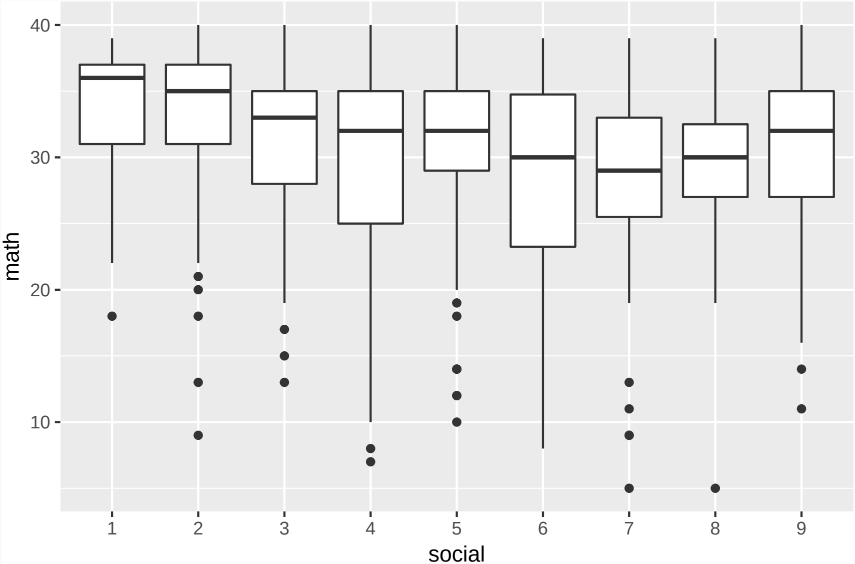
Nesting: `school:class:student`

Other variables crossed.

Plot our data - `faraway:: jsp`



Plot our data



Let's try simple lm

```
glin <- lm(math ~ raven*gender*social,jspr)
anova(glin)

glin <- lm(math ~ raven*social,jspr)
anova(glin)

glin <- lm(math ~ raven+social,jspr)
summary(glin)
```

Problems

1. All 953 students are independent?
2. Remember that they come from 50 different schools.

```
table(jspr$school)
```

```
##
##  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19
## 26 11 14 24 26 18 11 27 21  0 11 23 22 13  7 16  6 18 14
## 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 44 45
## 20 22 15 13 27 35 23 44 27 16 28 17 12 14 10 10 41  5 11
```

Include random effects

We need an analysis that uses the individual-level information, but also reflects the grouping in the data.

```
mmod <-
  lmer(math ~ raven * social * gender +
  (1 | school) +
  (1 | school:class), data = jspr)
```

Model selection

```
all3 <- lmer(math ~ raven * social * gender +
         (1 | school) + (1 | school:class),
         data = jspr, REML = FALSE)
all2 <- update(all3, . ~ . - raven:social:gender)
notrs <- update(all2, . ~ . -raven:social)
notrg <- update(all2, . ~ . -raven:gender)
notsg <- update(all2, . ~ . -social:gender)
onlyrs <- update(all2, . ~ . -
                   social:gender-raven:gender)
all1 <-  update(all2, . ~ . -social:gender -
                  raven:gender - social:raven)
nogen <- update(all1, . ~ . -gender)
```

Model selection

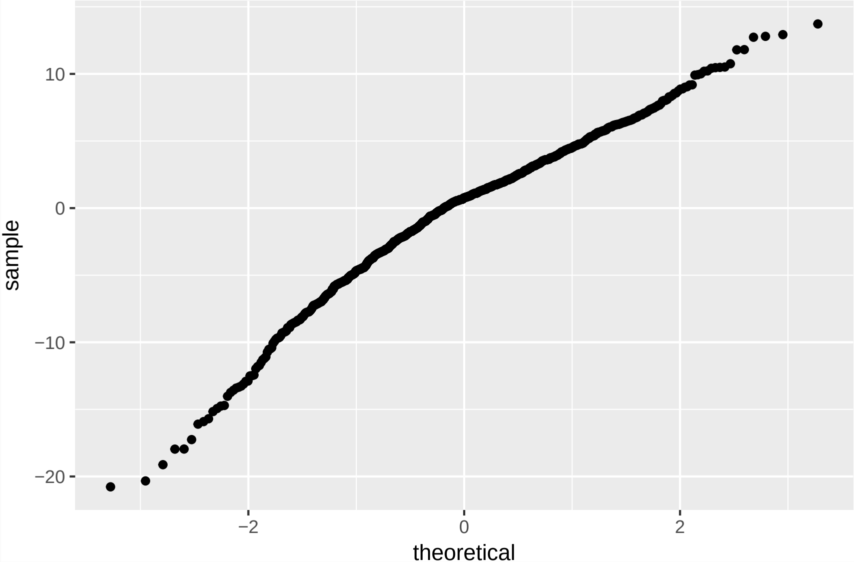
```r
anova(all3, all2, notrs, notrg, notsg, onlyrs, all1, nogen)
```

```
##        Df    AIC    BIC  logLik
## nogen  13 5954.3 6017.5 -2964.2
## all1   14 5955.6 6023.6 -2963.8
## onlyrs 22 5950.1 6057.0 -2953.1
## notrs  23 5961.6 6073.4 -2957.8
## notsg  23 5952.0 6063.8 -2953.0
## notrg  30 5956.1 6101.9 -2948.1
## all2   31 5957.8 6108.4 -2947.9
## all3   39 5966.7 6156.2 -2944.3
```

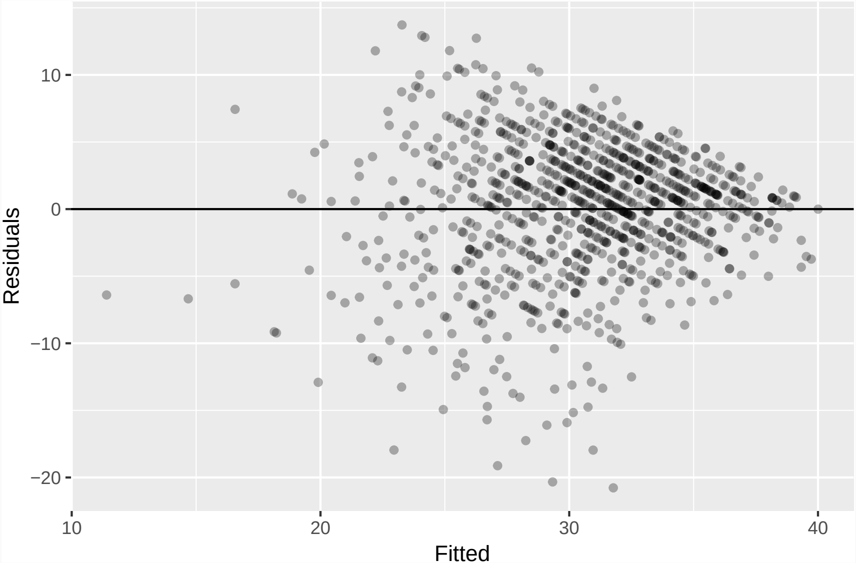
Final model

```
jspr$craven <- jspr$raven-mean(jspr$raven)
mmod <- lmer(math ~ craven*social +
          (1|school) +
          (1|school:class), jspr)
sumary(mmod)
```

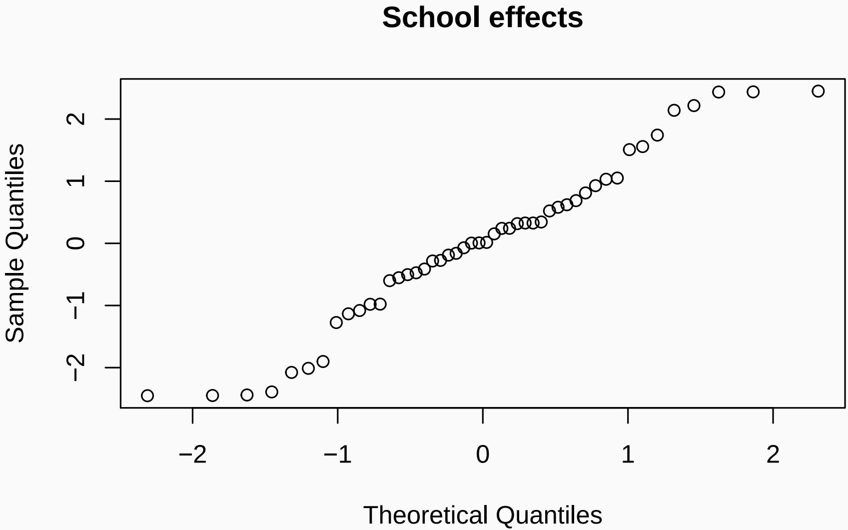
Model diagnosis: fixed effects



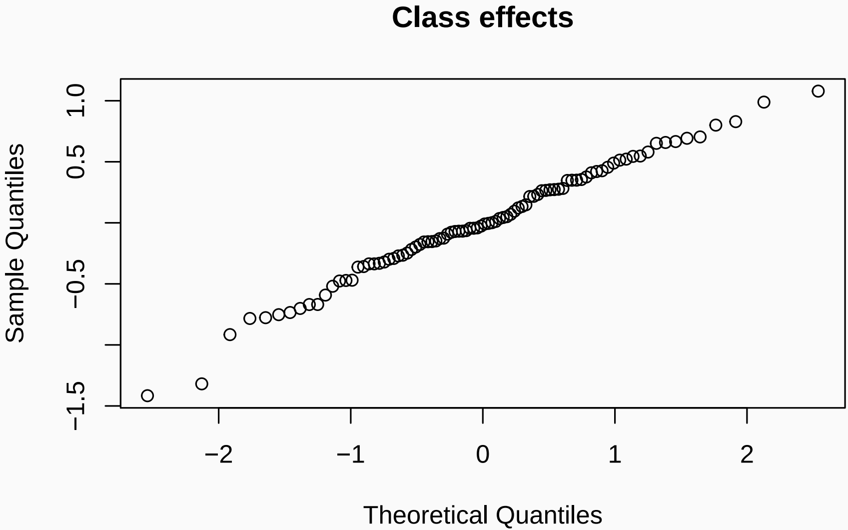
Model diagnosis: fixed effects



Model diagnosis: random effects



Model diagnosis: random effects



Outline

Longitudinal data

- Repeated measurements on each unit taken over time.
- Called “panel data” in econometrics. Called “longitudinal data” in every other discipline.
- Individuals treated as random effects
- Differs from time series in having many units (e.g., people) but often not many observations per person.
- i.e., Longitudinal data has large N, small T; Time series data has small N, large T.

Longitudinal data

For unit (individual) i, $\boldsymbol{y}_i$ is a T-vector such that

$$\boldsymbol{y}_i | \gamma_i \sim N(\boldsymbol{X}_i \beta + \gamma_i, \sigma^2 \boldsymbol{\Lambda}_i)$$

- $\gamma_i \sim N(0, \sigma^2 D)$ is effect of ith unit
- $\boldsymbol{X}_i$ contains predictors for fixed effects
- $\boldsymbol{\Lambda}_i$ handles autocorrelations within units
- $\boldsymbol{y}_i \sim N(\boldsymbol{X}_i \beta, \boldsymbol{\Sigma}_i)$ where $\boldsymbol{\Sigma}_i = \sigma^2(\boldsymbol{\Lambda}_i + \boldsymbol{D})$
- Assume individuals are independent, and random effects and errors are uncorrelated.

Combining individuals (assuming independence):

$$\boldsymbol{y} = \begin{bmatrix} \boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_N \end{bmatrix} \qquad \boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 \\ \vdots \\ \boldsymbol{X}_N \end{bmatrix}$$

$$\boldsymbol{\Sigma} = \text{diag}(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \ldots, \boldsymbol{\Sigma}_N),$$

$$\boldsymbol{y} \sim N(\boldsymbol{X}\beta, \boldsymbol{\Sigma})$$

Longitudinal data

Combining individuals (assuming independence):

$$\boldsymbol{y} = \begin{bmatrix} \boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_N \end{bmatrix} \qquad \boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 \\ \vdots \\ \boldsymbol{X}_N \end{bmatrix}$$

$$\boldsymbol{\Sigma} = \mathrm{diag}(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \ldots, \boldsymbol{\Sigma}_N),$$

$$\boldsymbol{y} \sim N(\boldsymbol{X}\beta, \boldsymbol{\Sigma})$$

- Only additional complication is choosing correlation structure
- Other random effects can be added; then γ_i becomes a vector.