Multilevel Models

Nora Wickelmaier

Last modified: April 24, 2025

Multilevel models

- are special cases of mixed-effects models
- are useful for data with a hierarchical structure
- usually contain nested random effects

We will look at an example from the book “Practical Regression and Anova” by Julian Faraway: the Junior School Project data

```
library(faraway)
data("jsp")
?jsp
str(jsp)
summary(jsp)
head(jsp)
```

Junior School Project

A data frame with 3236 observations on the following 9 variables

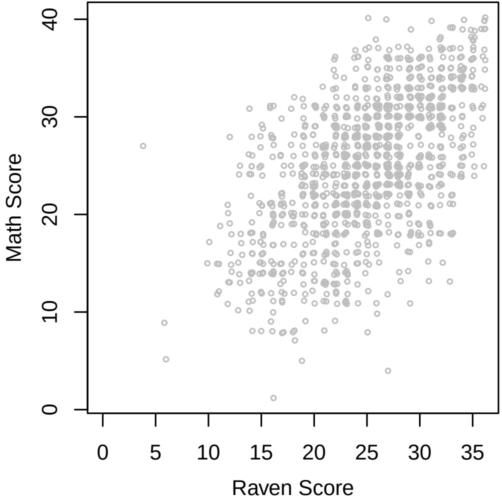
`school`	50 schools code 1–50
`class`	a factor with levels '1' '2' '3' '4'
`gender`	a factor with levels 'boy' 'girl'
`social`	class of the father I = 1; II = 2; III nonmanual = 3; III manual = 4; IV = 5; V = 6; Long-term unemployed = 7; Not currently employed = 8; Father absent = 9
`raven`	test score
`id`	student id coded 1–1402
`english`	score on English
`math`	score on Maths
`year`	year of school

https://en.wikipedia.org/wiki/Raven%27s_Progressive_Matrices

Nested structure of data

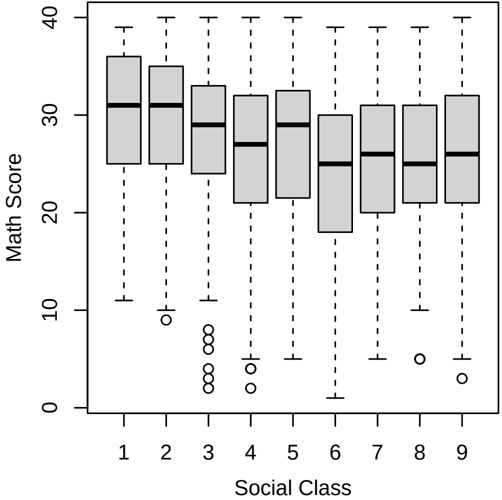
```
data("jsp")
str(jsp)
xtabs( ~ school + class, data = jsp, sparse = TRUE)

# Investigate nested structure of the data
xtabs( ~ school + factor(school:class), data = jsp, sparse = TRUE)

# Several data points per student
table(jsp$year, useNA = "ifany")
table(jsp$id, useNA = "ifany")

dat <- subset(jsp, year == 0)
```

Visualization of data




```r
plot(math ~ raven, data = dat,
  xlab = "Raven Score",
  ylab = "Math Score")
```

Visualization of data




```
plot(math ~ social, data = dat,
  xlab = "Social Class",
  ylab = "Math Score")
```

Centering variables

- Psychological variables often do not have a "natural" zero and linear transformations of the form $y = a \cdot x + b$, with a and b being constants, are allowed
- $x - \bar{x}$ is a linear transformation with $a = 1$ and $b = -\bar{x}$; this transformation is called centering of a variable (compare z transformation)
- By centering variables the interpretation of the intercept in a linear model changes
    - Uncentered intercepts represent the difference to a value of 0
    - Centered intercepts represent the difference to the mean

Centering variables

Options to center variables in multilevel models with two levels

- Level 1
  - Centering around group mean
  - Centering around grand mean
- Level 2
  - Centering around grand mean
- Interpretation of the intercepts needs to refer to the respective value of 0 (point of origin)
- Attention
  - Centering around the respective group mean might eliminate possible group differences between the group means
  - In order to avoid this pitfall a variable that contains the group means should be included

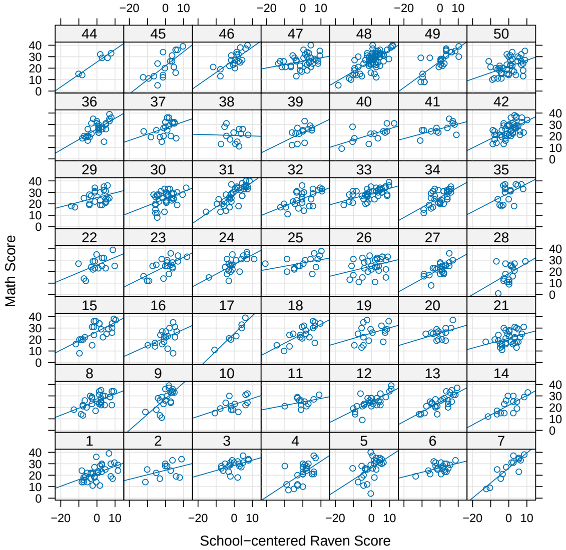
Centering of raven score

```
# Centering around grand mean
dat$craven <- dat$raven - mean(dat$raven)

lm1 <- lm(math ~ raven, data = dat)
lm2 <- lm(math ~ craven, data = dat)

# Visualization
plot(jitter(math) ~ jitter(raven), data = dat, cex = .7,
  xlab = "Raven score", ylab = "Math score",
  xlim = c(0, 36), col = "gray")
abline(v = 0, h = coef(lm1)[1], col = "darkgray")
abline(v = mean(dat$raven), h = coef(lm2)[1], col = "darkgray")
abline(lm1)
```

Centering of raven score

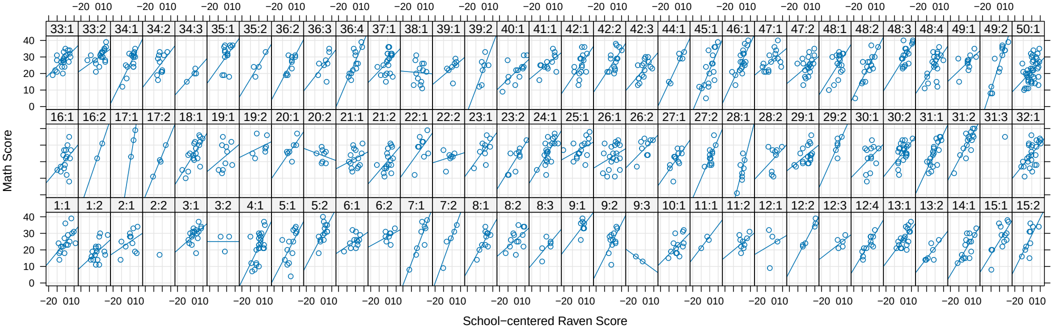
```r
# Centering around group mean for each school
## add mean raven score per school
dat$mraven <- with(dat, ave(raven, school))
dat$mraven <- dat$mraven - mean(dat$mraven)
## center raven score: mean = 0 for each school
dat$gcraven <- dat$craven - dat$mraven

aggregate(craven ~ school, data = dat, FUN = mean)
aggregate(gcraven ~ school, data = dat, FUN = mean)

# Visualization
plot(jitter(math) ~ jitter(craven), data = dat, cex = .7,
  xlab = "Centered raven score", ylab = "Math score", col = dat$school)
abline(v = unique(dat$mraven), col = unique(dat$school))
abline(v = mean(dat$craven), col = "darkblue", lwd = 4)
```

Visualization of data – Schools




```r
lattice::xyplot(
  math ~ gcraven | school,
  data = dat,
  xlab = "Raven Score",
  ylab = "Math Score",
  type = c("p", "g", "r")
  )
```

Visualization of data – Classes




```
lattice::xyplot(math ~ gcraven | school:class, data = dat,
                xlab = "Raven Score", ylab = "Math Score",
                type = c("p", "g", "r"))
```

Random intercept model

- The data actually consist of 3 levels: students in classes in schools
- Because there are no class-level predictors, the class effect ω enters on the same level as the school effect υ

$$
\begin{aligned}
\text{(Level 1)} \quad y_{ijk} &= b_{0ij} + b_{1i}\, gcraven_{ijk} + b_{2i}\, social_{ijk} \\
&\quad + b_{3i}\,(gcraven_{ijk} \times social_{ijk}) + \varepsilon_{ijk} \\
\text{(Level 2)} \quad b_{0ij} &= \beta_0 + \upsilon_{0i} + \omega_{0j} \\
b_{1i} &= \beta_1 \\
b_{2i} &= \beta_2 \\
b_{3i} &= \beta_3 \\
\text{(2) in (1)} \quad y_{ijk} &= \beta_0 + \beta_1\, gcraven_{ijk} + \beta_2\, social_{ijk} \\
&\quad + \beta_3\,(gcraven_{ijk} \times social_{ijk}) \\
&\quad + \upsilon_{0i} + \omega_{0j} + \varepsilon_{ijk}
\end{aligned}
$$

with $\upsilon_{0i} \sim N(0, \sigma^2_\upsilon)$ i.i.d., $\omega_{0j} \sim N(0, \sigma^2_\omega)$ i.i.d., $\varepsilon_{ijk} \sim N(0, \sigma^2)$ i.i.d.

Random intercept model

```
library(lme4)
m1 <- lmer(math ~ gcraven * social + (1 | school) + (1 | school:class),
           data = dat, REML = FALSE)
confint(m1)

# Significance tests
m0 <- lmer(math ~ 1 + (1 | school) + (1 | school:class),
           data = dat, REML = FALSE)
m0.1 <- m0 |> update(. ~ gcraven + .)
m0.2 <- m0 |> update(. ~ gcraven + social + .)
anova(m0, m0.1, m0.2, m1)

# Model diagnostics
plot(m1)
lattice::qqmath(m1)
```

Multilevel structure

We consider two levels:

- Level 1 refers to the students
- Level 2 refers to the schools

Level	Variable	Description
2	`school`	50 schools code 1–50
2	`mraven`	mean raven score of school (overall mean 0)
1	`social`	class of the father (categorical)
1	`raven`	test score
1	`gcraven`	centered test score (mean for each school 0)
1	`math`	score on Maths

Multilevel structure

$$
\begin{aligned}
\text{(Level 1)} \quad y_{ij} &= b_{0i} + b_{1i}\, gcraven_{ij} + b_{2i}\, social_{ij} \\
&\quad + b_{3i}\, (gcraven_{ij} \times social_{ij}) + \varepsilon_{ij} \\
\text{(Level 2)} \quad b_{0i} &= \beta_0 + \beta_4\, mraven_i + \upsilon_{0i} \\
b_{1i} &= \beta_1 + \beta_5\, mraven_i + \upsilon_{1i} \\
b_{2i} &= \beta_2 \\
b_{3i} &= \beta_3 \\
\text{(2) in (1)} \quad y_{ij} &= \beta_0 + \beta_1\, gcraven_{ij} + \beta_2\, social_{ij} \\
&\quad + \beta_3 (gcraven_{ij} \times social_{ij}) + \beta_4\, mraven_i \\
&\quad + \beta_5\, (gcraven_{ij} \times mraven_i) \\
&\quad + \upsilon_{0i} + \upsilon_{1i}\, gcraven_{ij} + \varepsilon_{ij}
\end{aligned}
$$

with $\boldsymbol{\upsilon} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\upsilon)$ i.i.d., $\varepsilon_{ij} \sim N(0, \sigma^2)$ i.i.d.

Fitting multilevel model

```
m2 <- lmer(math ~ mraven * gcraven + (gcraven | school),
           data = dat, REML = FALSE)

m3 <- lmer(math ~ mraven * gcraven + social + (gcraven | school),
           data = dat, REML = FALSE)

m4 <- lmer(math ~ mraven * gcraven + gcraven * social +
           (gcraven | school), data = dat, REML = FALSE)

anova(m2, m3, m4)

summary(m3)
confint(m3)
```

Interpretation of results

Fixed effects

- The mean math score for a student with mean intelligence in a mean intelligent school in the highest social class is 25.79
- By partitioning $craven = mraven + gcraven$, we can consider how intelligence affects math score on different levels
    - If the mean raven score per school increases by one point, the math score increases by 0.62 (95 % CI: [0.37, 0.88])
    - If the raven score for a student increases by one point, the math score increases by 0.69 (95 % CI: [0.61, 0.78])

Interpretation of results

Random effects

- We get an estimate of $\hat{\sigma}^2_{\upsilon_0} = 2.14$ for the variance of the mean math scores; this leaves room for improving the prediction by adding more school-level predictors
- There is hardly any variance for the dependence of math score on the raven score between schools ($\hat{\sigma}^2_{\upsilon_1} = 0.03$); this should be kept in mind when interpreting the correlation of $\hat{\rho}_{\upsilon_0\upsilon_1} = -0.06$ (95 % CI: [-0.56, 0.47])
- The corresponding covariance is $\hat{\sigma}_{\upsilon_0\upsilon_1} = -0.06 \cdot \hat{\sigma}_{\upsilon_0} \cdot \hat{\sigma}_{\upsilon_1} = -0.001$
- These results imply that a simpler model without random slopes for *gcraven* within schools might fit the data

References

Faraway, J. (2025). *Faraway: Datasets and functions for books by Julian Faraway* [R package version 1.0.9]. https://CRAN.R-project.org/package=faraway