Modeling two-level longitudinal data

03.16.22

STA 310

Click here for PDF of slides

Announcements

- Quiz 03: Wed, Mar 16 - Fri, Mar 18
  - Feb 21 - Mar 14 lectures
- DataFest: April 1 - 3 in Penn Pavilion
  - Click here to sign up

Learning goals

- Compare maximum likelihood (ML) and restricted maximum likelihood (REML) estimation approaches
- Describe general process for fitting and comparing multilevel models
- Fit and interpret multilevel models for longitudinal data

Fitting multilevel models

ML and REML

Maximum Likelihood (ML) and Restricted (Residual) Maximum Likelihood (REML) are the two most common methods for estimating the fixed effects and variance components

Maximum Likelihood (ML)

- Jointly estimate the fixed effects and variance components using all the sample data
- *Issue*: Fixed effects are treated as known values when estimating variance components
    - Results in biased estimates of variance components (especially when sample size is small)

See the post Maximum Likelihood (ML) vs. REML for details and illustration of ML vs. REML.

STA 310

ML and REML

Restricted Maximum Likelihood (REML)

- Estimate the variance components using the <u>sample residuals</u>, resulting in unbiased estimates of the variance components
- **Process:**
  - Fit regression model of fixed effects using ordinary least squares (OLS)
  - Take the residuals from this model and estimate the variance components by maximizing the likelihood of the residuals.
  - Obtain generalized least squares (GLS) estimates for fixed effects (estimates that take into account variance components from previous step). Retain the GLS estimates of the fixed effects. *Note: The GLS and OLS estimates can be equivalent*

Illustration estimating regression standard error in OLS.

ML vs. REML

Researchers have not definitively determined one method superior to the other

ML (`REML = FALSE`)

- Use ML estimates to compare models with different fixed effects
- Can get biased estimates of variance when number of groups is small

Source: Hierarchical Linear Modeling with Maximum Likelihood, Restricted Maximum Likelihood, and Fully Bayesian Estimation by Peter Boedeker

REML (`REML = TRUE`)

- Default in `lmer`
- Get more accurate estimates of variance components when there are a small number of groups or large number of parameters
- Can only compare models that have the same fixed effects and only differ in the variance components

STA 310

Comparing ML and REML estimates

Below are estimates for the fixed effects and variance components for the music model from the previous lecture.

	Maximum Likelihood		Restricted Maximum Likelihood	
Term	**Estimate**	**Std. Error**	**Estimate**	**Std. Error**
(Intercept)	15.924	0.623	15.930	0.641
orchestra1	1.696	0.919	1.693	0.945
large_ensemble1	-0.895	0.827	-0.911	0.845
orchestra1:large_ensemble1	-1.438	1.074	-1.424	1.099
sd__(Intercept)	2.286	NA	2.378	NA
cor__(Intercept).large_ensemble1	-1.00	NA	-0.635	NA
sd__large_ensemble1	0.385	NA	0.672	NA
sd__Observation	4.665	NA	4.670	NA

Modeling two-level longitudinal data

Data: Charter schools in MN

The data set `charter-long.csv` contains standardized test scores and demographic information for schools in Minneapolis, MN from 2008 to 2010. The data were collected by the Minnesota Department of Education. Understanding the effectiveness of charter schools is of particular interest, since they often incorporate unique methods of instruction and learning that differ from public schools.

- **`MathAvgScore`**: Average MCA-II score for all 6th grade students in a school (response variable)
- **`urban`**: urban (1) or rural (0) location school location
- **`charter`**: charter school (1) or a non-charter public school (0)
- **`schPctfree`**: proportion of students who receive free or reduced lunches in a school (based on 2010 figures).
- **`year08`**: Years since 2008

Data

```
charter <- read_csv("data/charter-long.csv")
```

schoolName	year08	urban	charter	schPctfree	MathAvgScore
RIPPLESIDE ELEMENTARY	0	0	0	0.363	652.8
RIPPLESIDE ELEMENTARY	1	0	0	0.363	656.6
RIPPLESIDE ELEMENTARY	2	0	0	0.363	652.6
RICHARD ALLEN MATH&SCIENCE ACADEMY	0	1	1	0.545	NA
RICHARD ALLEN MATH&SCIENCE ACADEMY	1	1	1	0.545	NA
RICHARD ALLEN MATH&SCIENCE ACADEMY	2	1	1	0.545	631.2

Assess missingness

Missing data is common in longitudinal data. Before starting the analysis, it is important to understand the missing data patterns. Use the **`skim`** function from the skimr R package to get a quick view of the missingness.

```
library(skimr)
charter %>% skim() %>% select(skim_variable, n_missing, complete_rate)
```

```
## # A tibble: 9 × 3
##   skim_variable n_missing complete_rate
##   <chr>             <int>         <dbl>
## 1 schoolid              0         1
## 2 schoolName            0         1
## 3 urban                 0         1
## 4 charter               0         1
## 5 schPctnonw            0         1
## 6 schPctsped            0         1
## 7 schPctfree            0         1
## 8 year08                0         1
## 9 MathAvgScore        121         0.935
```

Closer look at missing data pattern

Output | Code

MathAvgScore0_miss	MathAvgScore1_miss	MathAvgScore2_miss	n
0	0	0	540
0	0	1	6
0	1	0	4
0	1	1	7
1	0	0	25
1	0	1	1
1	1	0	35

STA 310

Closer look at missing data pattern

Output Code

```
charter %>%
  select(schoolid, schoolName, year08, MathAvgScore) %>%
  pivot_wider(id_cols = c(schoolid, schoolName), names_from = year08,
              names_prefix = "MathAvgScore.", values_from = MathAvgScore) %>%
  mutate(MathAvgScore0_miss = if_else(is.na(MathAvgScore.0), 1, 0),
         MathAvgScore1_miss = if_else(is.na(MathAvgScore.1), 1, 0),
         MathAvgScore2_miss = if_else(is.na(MathAvgScore.2), 1, 0)) %>%
  count(MathAvgScore0_miss, MathAvgScore1_miss, MathAvgScore2_miss) %>%
  kable()
```

Dealing with missing data

- **Complete case analysis**: Only include schools with complete data for all three years.
  - Would remove 12.6% of observations in this data.
- **Last observation carried forward**: Keep the last observation from each group (school in this case) and conduct analysis for independent observations.
- **Impute missing observations**: "Fill in" values of missing observations using the typical observed trends from groups with similar covariates.
- **Apply multilevel methods**: Estimate patterns using available data recognizing that trends for groups with complete data re more precise than for those with fewer measurements. This is under the condition that the probability of missingness does not depend on unobserved predictors or the response.

What is an advantage of each method? What is an disadvantage?

03:00

Strategy for building multilevel models

- Conduct exploratory data analysis for Level One and Level Two variables.
- Fit model with no covariates to assess variability at each level.
- Create Level One models. Start with a single term, then add terms as needed.
- Create Level Two models. Start with a single term, then add terms as needed. Start with equation for intercept term.
- Begin with the full set of variance components, then remove variance terms as needed.

Alternate model building strategies in BMLR Section 8.6

STA 310

Exploratory data analysis

Given the longitudinal structure of the data, we are able to answer questions at two levels

Level One (within school): How did average math scores within a school change over time?

- **Level Two (between schools)**: What is the effect of school-specific covariates on the average math scores in 2008 and the rate of change from 2008 to 2010?

We can conduct exploratory data analysis at both levels, e.g.,

- Univariate and bivariate EDA
- Lattice plots
- Spaghetti plots

Open **`lecture-18.Rmd`**.

See BMLR Section 9.3 for full exploratory data analysis.

Unconditional means model

Start with the **unconditional means model** a model with no covariates at any level

- Also called the random intercepts model

Level One : $Y_{ij} = a_i + \epsilon_{ij}, \quad \epsilon_{ij} \sim N(0, \sigma^2)$

Level Two: $a_i = \alpha_0 + u_i, \quad u_i \sim N(0, \sigma_u^2)$

Write the composite model.

Intraclass correlation

The **intraclass correlation** is the relative variability between groups

$$\hat{\rho} = \frac{\text{Between variability}}{\text{Within variability}} = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + \hat{\sigma}^2}$$

- What is the meaning of $\hat{\rho}$ close to 0?
- What is the meaning of $\hat{\rho}$ close to 1?

Fit the unconditional means model and calculate the intraclass correlation.

Unconditional growth model

A next step in the model building is the **unconditional growth model**, a model with Level One predictors but no Level Two predictors.

The composite model is

$$Y_{ij} = \alpha_0 + \beta_0 Year08_Pij + u_i + v_i Year08_{ij} + \epsilon_{ij}$$

$$\epsilon_{ij} \sim N(0, \sigma^2) \text{ and } \begin{bmatrix} u_i \\ v_i \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{bmatrix} \right)$$

- Write the Level One and Level Two models.
- What can we learn from this model?
- Fit the unconditional growth model.

STA 310

Pseudo R^2

We can use **Pseudo R^2** to explain changes in variance components between two models

- **Note**: This should only be used when the definition of the variance component is the same between the two models

Pseudo R^2 = \frac{\hat{\sigma}_.(Model 1) - \hat{\sigma}_.(Model 2)}{\frac{\hat{\s

Calculate the $Pseudo R^2$ to estimate the change of within school variance between the unconditional means and unconditional growth models.

Fit a model

Model in section 9.6.3

- Write it out
- Notice difference in parameters use for sloeps and intercepts - what does this mean
- Interpret the coeffcients for the model

Acknowledgements

The content in the slides is from

- BMLR: Chapter 8 - Introduction to Multilevel Models
    - Sections 8.5 - 8.12
    - Singer, J. D. & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford university press.