

# Modeling two-level longitudinal data

Inference

03.21.22

[Click here for PDF of slides](#)

# Announcements

- Mini-project 02
  - draft due **Thu, Mar 24 at noon** (peer review in lab)
  - presentations **Mon, Mar 28** in class
  - report due **Mon, Mar 28** at 11:59pm
- Looking ahead
  - HW 04 due Fri, Apr 1 (assigned later this week)
  - Final project - optional draft due **Fri, Apr 15**, final report due **Wed, Apr 27**
- DataFest: April 1 - 3 in Penn Pavilion
  - [Click here](#) to register

# Learning goals

- Compare multilevel models
- Conduct inference for random effects
- Conduct inference for fixed effects

# Comparing multilevel models

# 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<br>MATH&SCIENCE ACADEMY | 0      | 1     | 1       | 0.545      | NA           |
| RICHARD ALLEN<br>MATH&SCIENCE ACADEMY | 1      | 1     | 1       | 0.545      | NA           |
| RICHARD ALLEN<br>MATH&SCIENCE ACADEMY | 2      | 1     | 1       | 0.545      | 631.2        |

# Unconditional means model

Start with the **unconditional means model**, a model with no covariates at any level. This is 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

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

$\hat{\rho} = 0.798$ . This means...

- About 79.8% of the variability in math scores can be attributed to differences between schools (school-to-school variability). About 20.2% of the variability can be attributed to changes over time.
- The average correlation between any two responses from the same school is about 0.798.
- The **effective sample size** (number of independent pieces of information available for modeling) is closer to the number of schools ( $\rho$  close to 1) than the number of

# 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.

**Level One:**  $Y_{ij} = a_i + b_i Year08_{ij} + \epsilon_{ij}$

**Level Two:**

- $a_i = \alpha_0 + u_i$
- $b_i = \beta_0 + v_i$

- Write the composite model.
- What can we learn from this model?

# Pseudo $R^2$

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

$$\text{Pseudo } R^2 = \frac{\hat{\sigma}^2(\text{Model 1}) - \hat{\sigma}^2(\text{Model 2})}{\hat{\sigma}^2(\text{Model 1})}$$

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

- The Pseudo  $R^2$  estimating the change in within school variance between the unconditional means and unconditional growth model is about 0.17. What does this value mean?
- Why can't we use Pseudo  $R^2$  to understand change in  $\hat{\sigma}_u^2$  between the two models?

# Model with school-level covariates

Fit a model with school-level covariates that takes the following form:

## Level One

$$Y_{ij} = a_i + b_i Year08_{ij} + \epsilon_{ij}$$

## Level Two

$$a_i = \alpha_0 + \alpha_1 Charter_i + \alpha_2 urban_i + \alpha_3 schpctfree_i + u_i$$
$$b_i = \beta_0 + \beta_1 Charter_i + \beta_2 urban_i + v_i$$

- Write out the composite model.
- Fit the model in R.

# Consider a simpler model

Would a model without the effects  $v_i$  and  $\rho_{uv}$  be preferable?

## Level One

$$Y_{ij} = a_i + b_i Year08_{ij} + \epsilon_{ij}$$

## Level Two

$$a_i = \alpha_0 + \alpha_1 Charter_i + \alpha_2 urban_i + \alpha_3 schpctfree_i + u_i$$
$$b_i = \beta_0 + \beta_1 Charter_i + \beta_2 urban_i$$

In this model, the effect of year is the same for all schools with a given combination of **Charter** and **urban**

# Full and reduced models

```
full_model <- lmer(MathAvgScore ~ charter + urban + schPctfree +  
                  charter:year08 + urban:year08 + year08 +  
                  (year08|schoolid), REML = T, data = charter)
```

```
reduced_model <- lmer(MathAvgScore ~ charter + urban + schPctfree +  
                     charter:year08 + urban:year08 + year08 +  
                     (1 | schoolid), REML = T, data = charter)
```

# Issues with the drop-in-deviance test

```
anova(full_model, reduced_model, test = "Chisq") %>%  
  kable(digits = 3)
```

|               | npar | AIC      | BIC      | logLik    | deviance | Chisq | Df | Pr(>Chisq) |
|---------------|------|----------|----------|-----------|----------|-------|----|------------|
| reduced_model | 9    | 9952.992 | 10002.11 | -4967.496 | 9934.992 | NA    | NA | NA         |
| full_model    | 11   | 9953.793 | 10013.83 | -4965.897 | 9931.793 | 3.199 | 2  | 0.202      |

- $\chi^2$  test conservative, i.e., the p-values are larger than they should be, when testing random effects at the boundary ( e.g.,  $\sigma_v^2 = 0$ ) or those with bounded ranges (e.g.,  $\rho_{uv}$  )
- If you observe small p-values, you can feel relatively certain the tested effects are statistically significant
- Use bootstrap methods to obtain more accurate p-values



# Parametric bootstrapping

- **Bootstrapping** is from the phrase "pulling oneself up by one's bootstraps"
- Accomplishing a difficult task without any outside help
- **Task:** conduct inference for model parameters (fixed and random effects) using only the sample data

# Parametric bootstrapping for likelihood ratio test

- 1 Fit the null (reduced) model to obtain the fixed effects and variance components (*parametric* part).
- 2 Use the estimated fixed effects and variance components to generate a new set of response values with the same sample size and associated covariates for each observation as the original data (*bootstrap* part).
- 3 Fit the full and null models to the newly generated data.
- 4 Compute the likelihood test statistic comparing the models from the previous step.
- 5 Repeat steps 2 - 4 many times ( $\sim 1000$ ).

# Parametric bootstrapping for likelihood ratio test

- 6 Create a histogram of the likelihood ratio statistics to get the distribution of likelihood ratio statistic under the null hypothesis.
- 7 Get the p-value by calculating the proportion of bootstrapped test statistics greater than the observed statistic.

Go to **lecture-19.Rmd** to calculate the bootstrapped p-value for the likelihood ratio test statistic testing whether  $v_i$  and  $\rho_{uv}$  can be removed from the model.

# Inference for fixed effects

- The output for multilevel models do not contain p-values for the coefficients of fixed effects
- The exact distribution of the test statistic under the null hypothesis (no fixed effect) is unknown, because the exact degrees of freedom are unknown
  - Finding suitable approximations is an area of ongoing research
- We can use likelihood ratio test with an approximate  $\chi^2$  distribution to test these effects, since we're not testing on the boundary and fixed effects do not have limited ranges
  - Some research suggests the p-values are too low but approximations are generally pretty good
  - Can also calculate the p-values using parametric bootstrap approach

Go to **lecture-19.Rmd** to test whether **schPctFree** should be included in the current model using likelihood ratio test with the  $\chi^2$  distribution and the parametric bootstrap.

# Inference for individual coefficients

- Use the t-value  $\left( \frac{estimate}{std.error} \right)$  in the model output
  - General rule: Coefficients with  $|t\text{-value}| > 2$  considered to be statistically significant, i.e., different from 0
- Calculate confidence intervals using **nonparametric bootstrapping**
  - Take a sample, with replacement, of size  $n$  (the size of the original data) (called *case resampling*)
  - Fit the model to obtain estimates of the coefficients
  - Repeat these two steps many times ( $\sim 1000$ ) to obtain the bootstrap distribution
  - Get the coefficients for the 95% confidence interval by taking the middle 95% of the bootstrap distribution

Go to **lecture-19.Rmd** for an example of nonparametric bootstrapping to estimate the 95% confidence interval for the model coefficients.



# Bootstrapped CI for coefficients

```
confint(reduced_model, method = "boot", level = 0.95, oldNames = F) %>%  
  kable(digits = 3)
```

|                           | 2.5 %   | 97.5 %  |
|---------------------------|---------|---------|
| sd_(Intercept)   schoolid | 4.282   | 4.846   |
| sigma                     | 2.857   | 3.121   |
| (Intercept)               | 659.538 | 661.388 |
| charter                   | -4.489  | -1.525  |
| urban                     | -2.116  | -0.251  |
| schPctfree                | -19.289 | -16.129 |
| year08                    | 1.216   | 1.801   |
| charter:year08            | 0.319   | 1.689   |

# Summary: Model comparisons

## Methods to compare models with different fixed effects

- Likelihood ratio tests based on  $\chi^2$  distribution
- Likelihood ratio test based on parametric bootstrapped p-values
- AIC or BIC

## Methods to compare models with different variance components

- Likelihood ratio test based on parametric bootstrapped p-values
- AIC or BIC

# Summary: Understanding the model

## Methods to understand individual coefficients

- Bootstrap confidence intervals
- Pseudo  $R^2$  for variance components (if meaning is unchanged between models)

## Methods to understand data structure

- Calculate intraclass correlation coefficient using unconditional means model

# Acknowledgements

The content in the slides is from

- BMLR: Chapter 9 - Two-level Longitudinal Data
  - Sections 9.1 - 9.6
- Hierarchical Linear Modeling with Maximum Likelihood, Restricted Maximum Likelihood, and Fully Bayesian Estimation by Peter Boedeker
- *Applied longitudinal data analysis: Modeling change and event occurrence.* by J.D. Singer and J.B. Willett
  - Online copy available through Duke library

# Acknowledgements

- *Extending the linear model with R.* by Julian Faraway
  - Online copy available through Duke library
- More mathematical details in *Computational methods for mixed models* by Douglas Bates (a vignette for the **lme4** R package)