AMA3602
Applied Linear Models
Department of Applied Mathematics

Lecturer : Dr. Catherine Liu

THE HONG KONG POLYTECHNIC UNIVERSITY
香港理工大學

03/2024

Chapter 4 Linear Random-Effects Models

Reference:

[1] Boehmke, B., and Greenwell, B. M. (2019). Hands-on machine learning with R. CRC Press.

[2] Raudenbush, S. W., and Bryk, A. S. (2002). Hierarchical linear models: Applications and data analysis methods (Vol. 1). SAGE.

[3] Fox, John (2014) Lecture Notes of Sociology 761. Statistical Applications in Social Research https://socialsciences.mcmaster.ca/jfox/Courses/soc761/

Outline

Review

Introduction

LMM (=Linear Random-effects Model)----slide 23

Modeling Hierarchical Data

Formulating a Mixed Model

Random-Effects One-Way Analysis of Variance Random-

Coefficients Regression Model Coefficients-as-Outcomes

Model

Review

Introduction

LMM

Modeling Hierarchical Data

Formulating a Mixed Model

Random-Effects One-Way Analysis of Variance

Random-Coefficients Regression Model

Coefficients-as-Outcomes Model

LM vs GLM

- Linear Models (lm in R) $\geq$ Generalized linear models (glm in R)
  - Model error for LM: normal
  - Model error for GLM: exponential family
- That is, response may be
  - Binary or countable (Logistic regression or Poisson regression)
  - Normal, exponential, Gamma distributed
- In our course, we focus on LM $\geq$ MLR ($\geq$ SLR).

From perspective of data

- Single-level linear regression: classical MLR ($>$SLR),
  - fitting and graphical display;
  - C.I./P.I.;
  - Hypothesis testing;
  - Diagnostics.
- Multilevel linear regression: Varying intercepts and slopes
  - Regression with potentially large numbers of coefficients that are themselves modeled
  - Regressions with coefficients that can vary by groups

Statistical models

- A Statistical Model is the use of statistics to build a representation of the data and then conduct analysis to infer any relationships between variables or discover insights.
- A statistical model will have sampling, probability spaces, assumptions and diagnostics etc, to make inferences.
- We use statistical models to find insights given a particular set of data. We can conduct modelling on a relatively small set of data just to try and understand the underlying nature of the data.
- However, inherently all statistical models are wrong or not perfect. They are used to approximate reality. Sometimes the underlying assumptions of the model are far too strict and not representative of reality.

Why Hierarchical Model

In the linear models SLR/MLR, that we've looked at so far, we've assumed that

the observations are **at least uncorrelated or independent of each other** given the predictor variables.

However, there are many situations in which
that type of uncorrelation or independence does NOT hold

One major type of situation violating these uncorrelated assumptions is

- **cluster-level attributes:**
  when observations belong to different clusters and each cluster has its own properties (different response mean, different sensitivity to each predictor).
- **hierarchical** (also called **multi-level** and, in some cases **mixed-effects) models**, are designed to handle this type of **mutual dependence among datapoints**.

Fundamentals of Hierarchcal (Linear) Modeling

To handle data in the way correctly modeling correlated error where observations are not uncorrelated but instead cluster by one or more grouping variables.

Variant terms

- Hierarchical linear models
- Multilevel models
- Linear mixed models (LMM)

Uncorrelated error

Uncorrelated error is an important but often violated assumption of statistical procedures in the general *linear model family*.

- analysis of variance
- correlation
- regression
- factor analysis

Linear mixed modeling can lead to substantially different conclusions compared to conventional regression analysis.

Real-world problems

- predicted student test scores and errors in predicting them may cluster by classroom, school, and municipality.
- When clustering occurs due to a grouping factor (the rule, not the exception), then the standard errors computed for prediction parameters will be wrong.
- effects of predictor variables may be misinterpreted, not only in magnitude but even in direction.

Review

Introduction

LMM (Random-Effects Model)

Modeling Hierarchical Data

Formulating a Mixed Model

Random-Effects One-Way Analysis of Variance

Random-Coefficients Regression Model

Coefficients-as-Outcomes Model

Introduction (and Review)

The standard linear model

$$y_i = \beta_1 + \beta_2 x_{2i} + \cdots + \beta_p x_{pi} + \varepsilon_i$$
$$\varepsilon_i \sim \mathrm{NID}\left(0, \sigma^2\right)$$

assumes independently sampled observations, and hence independent errors ε_i. In matrix form,

$$\mathsf{y} = \mathsf{X}\boldsymbol{\beta} + \varepsilon$$
$$\boldsymbol{\varepsilon} \sim \mathsf{N}_n\left(0, \sigma^2 \mathsf{I}_n\right)$$

where $-\mathsf{y} = (y_1, y_2, \ldots, y_n)'$ is the response vector; $-\mathsf{X}$ is the model matrix, with typical row $\mathsf{x}_i' = (1, x_{2i}, \ldots, x_{pi})$ $-\beta = (\beta_1, \beta_2, \ldots, \beta_p)'$ is the vector of regression coefficients; $-\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)'$ is the vector of errors; $-\mathsf{N}_n$ represents the n -variable multivariate-normal distribution;

estimation of σ^2

0 is an $n \times 1$ vector of zeroes; and $-\mathrm{I}_n$ is the order- n identity matrix.
The standard linear model has one random effect, the error term ε_i, and one variance component, $\sigma^2 = \mathrm{Var}\,(\varepsilon_i)$.
When the assumptions of the standard linear model hold, ordinaryleast-squares (OLS) regression provides maximum-likelihood estimates of the regression coefficients,

$$\widehat{\boldsymbol{\beta}} = \left(\mathrm{X}'\mathrm{X}\right)^{-1}\mathrm{X}'\mathrm{y}$$

estimation of σ^2

- The MLE of the error variance σ^2 is

$$\widehat{\sigma}^2 = \frac{(\mathrm{y} - \mathrm{X}\widehat{\boldsymbol{\beta}})'(\mathrm{y} - \mathrm{X}\widehat{\boldsymbol{\beta}})}{n}$$

  $\widehat{\sigma}^2$ is a biased estimator of σ^2; usually, the unbiased estimator

$$s^2 = \frac{(\mathrm{y} - \mathrm{X}\widehat{\boldsymbol{\beta}})'(\mathrm{y} - \mathrm{X}\widehat{\boldsymbol{\beta}})}{n - p}$$

  is preferred. The standard linear model and OLS regression are generally inappropriate for dependent observations.
- Dependent (or clustered) data arise in many contexts, the two most common of which are hierarchical data and longitudinal data.

Hierarchical data

Hierarchical data are collected when sampling takes place at two or more levels, one nested within the other. Some examples:

Students within schools (two levels).
Students within classrooms within schools (three levels).

Individuals within nations (two levels).
Individuals within communities within nations (three levels).

Patients within physicians (two levels).
Patients within physicians within hospitals (three levels).

There can also be non-nested multi-level data - for example, highschool students who each have multiple teachers.

Longitudinal data

- Longitudinal data are collected when individuals (or other units of observation) are followed over time. Some examples:
- Annual data on vocabulary growth among children. Biannual data on weight-preoccupation and exercise among adolescent girls.
- Data collected at irregular intervals on recovery of IQ among coma patients.
- Annual data on employment and income for a sample of adult Canadians.

In all of these cases, it is not generally reasonable to assume that observations within the same higher-level unit, or longitudinal observations within the same individual, are independent of one-another.

Mixed-effect

- Mixed-effect models make it possible to take account of dependencies in hierarchical, longitudinal, and other dependent data. Unlike the standard linear model, mixed-effect models include more than one source of random variation – i.e., more than one random effect.
- Mixed-effects models have been developed in a variety of disciplines, with varying names and terminology: random-effects models (statistics, econometrics), variance and covariance-component models (statistics), hierarchical linear models (education), multi-level models (sociology), contextual-effects models (sociology), random-coefficient models (econometrics), repeated-measures models (statistics, psychology).
- Mixed-effects models have a long history, dating to Fisher and Yates's work on "split-plot" agricultural experiments.
- What distinguishes modern mixed models from their predecessors is generality: for example, the ability to accommodate irregular and missing observations.

Principal and Topics

- Principal sources for these lectures on mixed models:
  - J. Fox, Applied Regression Analysis and Generalized Linear Models, Third Edition, Sage, 2002 , Chapters 23 and 24 .
  - J. Fox and S. Weisberg, An *R* Companion to Applied Regression, Second Edition, Sage, 2011, "Mixed-Effects Models in R" (Appendix, draft).
- Topics:
  - The linear mixed-effects model.
  - Modeling hierarchical data.
  - Modeling longitudinal data.

Review

Introduction

LMM

Modeling Hierarchical Data

Formulating a Mixed Model

Random-Effects One-Way Analysis of Variance

Random-Coefficients Regression Model

Coefficients-as-Outcomes Model

The Linear Mixed-Effects Model

This section introduces a very general linear mixed model, which we will adapt to particular circumstances.
The Laird-Ware form of the linear mixed model:

$$
\begin{aligned}
y_{ij} =& \beta_1 + \beta_2 x_{2ij} + \cdots + \beta_p x_{pij} + b_{1i} z_{1ij} + \cdots + b_{qi} z_{qij} + \varepsilon_{ij} \\
b_{ki} \sim& N\left(0, \psi_k^2\right), \mathrm{Cov}\left(b_{ki}, b_{k'i}\right) = \psi_{kk'} \\
& b_{ki}, b_{k'i'} \text{ are independent for } i \neq i' \\
\varepsilon_{ij} \sim& N\left(0, \sigma^2 \lambda_{ijj}\right), \mathrm{Cov}\left(\varepsilon_{ij}, \varepsilon_{ij'}\right) = \sigma^2 \lambda_{ijj'}
\end{aligned}
$$

$\varepsilon_{ij}, \varepsilon_{i'j'}$ are independent for $i \neq i'$ where - y_{ij} is the value of the response variable for the j th of n_i observations in the i th of M groups or clusters. $\beta_1, \beta_2, \ldots, \beta_p$ are the fixed-effect coefficients, which are identical for all groups.

LMM

- $x_{2ij}, \ldots, x_{pij}$ are the fixed-effect regressors for observation j in group i; there is also implicitly a constant regressor, $x_{1ij} = 1$.
- $b_{1i}, \ldots, b_{qi}$ are the random-effect coefficients for group i, assumed to be multivariately normally distributed, independent of the random effects of other groups. The random effects, therefore, vary by group.
  - The b_{ik} are thought of as random variables, not as parameters, and are similar in this respect to the errors ε_{ij}.
- $z_{1ij}, \ldots, z_{qij}$ are the random-effect regressors.
  - The z s are almost always a subset of the x s (and may include all of the x s).
- When there is a random intercept term, $z_{1ij} = 1$ ψ_k^2 are the variances and $\psi_{kk'}$ the covariances among the random effects, assumed to be constant across groups.
  - In some applications, the ψ s are parametrized in terms of a smaller number of fundamental parameters.

LMM

- ε_{ij} is the error for observation j in group i
  - The errors for group i are assumed to be multivariately normally distributed, and independent of errors in other groups.
- $\sigma^2\lambda_{ijj'}$ are the covariances between errors in group i
  - Generally, the $\lambda_{ijj'}$ are parametrized in terms of a few basic parameters, and their specific form depends upon context.
  - When observations are sampled independently within groups and are assumed to have constant error variance (as is typical in hierarchical models), $\lambda_{ijj} = 1, \lambda_{ijj'} = 0$ (for $j \neq j'$), and thus the only free parameter to estimate is the common error variance, σ^2.
  - If the observations in a "group" represent longitudinal data on a single individual, then the structure of the λ s may be specified to capture serial (i.e., over-time) dependencies among the errors.

Laird-Ware model

The Laird-Ware model in matrix form:

$$\begin{aligned} \mathrm{y}_i &= \mathrm{X}_i \boldsymbol{\beta} + \mathrm{Z}_i \mathrm{b}_i + \boldsymbol{\varepsilon}_i \\ \mathrm{b}_i &\sim \mathrm{N}_q(0, \boldsymbol{\Psi}) \end{aligned}$$

$\mathrm{b}_i, \mathrm{b}_{i'}$ are independent for $i \neq i'$

$$\boldsymbol{\varepsilon}_i \sim \mathrm{N}_{n_i}\left(0, \sigma^2 \boldsymbol{\Lambda}_i\right)$$

$\varepsilon_i, \varepsilon_{i'}$ are independent for $i \neq i'$ where y_i is the $n_i \times 1$ response vector for observations in the i th group.

- X_i is the $n_i \times p$ model matrix for the fixed effects for observations in group i β is the $p \times 1$ vector of fixed-effect coefficients.
- Z_i is the $n_i \times q$ model matrix for the random effects for observations in group i
- b_i is the $q \times 1$ vector of random-effect coefficients for group i.
- ε_i is the $n_i \times 1$ vector of errors for observations in group i

Laird-Ware model

- Ψ is the $q \times q$ covariance matrix for the random effects.
- $\sigma^2\Lambda_i$ is the $n_i \times n_i$ covariance matrix for the errors in group i, and is $\sigma^2 \mathrm{I}_{n_i}$ if the within-group errors are homoscedastic and independent. Notice that there are two things that distinguish the linear mixed model from the standard linear model:
  (a) There are structured random effects b_i in addition to the errors ε_i.
  (b) The model can accommodate heteroscedasticity and dependencies among the errors.

Review

Introduction

LMM

Modeling Hierarchical Data

Formulating a Mixed Model

Random-Effects One-Way Analysis of Variance

Random-Coefficients Regression Model

Coefficients-as-Outcomes Model

Modeling Hierarchical Data

- Applications of mixed models to hierarchical data have become common in the social sciences, and nowhere more so than in research on education.
- We'll restrict this course to two-level models, but three or more levels can also be handled through an extension of the Laird-Ware model.
- The following example is borrowed from Raudenbush and Bryk, and has been used by others as well (though we will learn some things about the data that apparently haven't been noticed before).
  - The data are from the 1982 "High School and Beyond" survey, and pertain to 7185 U.S. high-school students from 160 schools – about 45 students on average per school.
  -70 of the high schools are Catholic schools and 90 are public schools.

Data analysis

- The object of the data analysis is to determine how students' math achievement scores are related to their family socioeconomic status.
  - We will entertain the possibility that the level of math achievement and the relationship between achievement and SES vary among schools.
  - If there is evidence of variation among schools, we will examine whether this variation is related to school characteristics - in particular, whether the school is a public school or a Catholic school and the average SES of students in the school.
- A good place to start is to examine the relationship between math achievement and SES separately for each school.
  - 160 schools are too many to look at individually, so I sampled 20 Catholic school and 20 public schools at random.

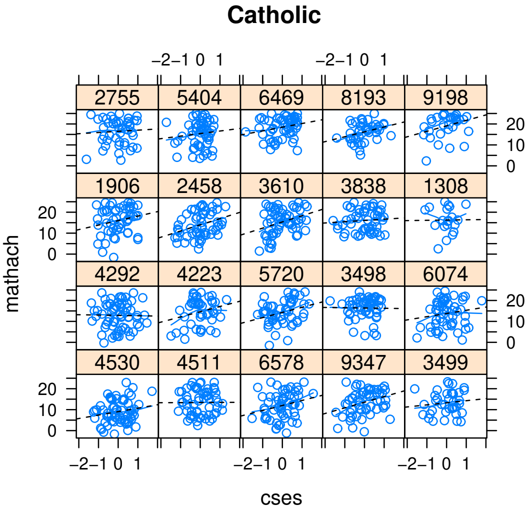

Figure: Math achievement by SES for students in 20 randomly selected Catholic schools. SES is centred at the mean of each school.

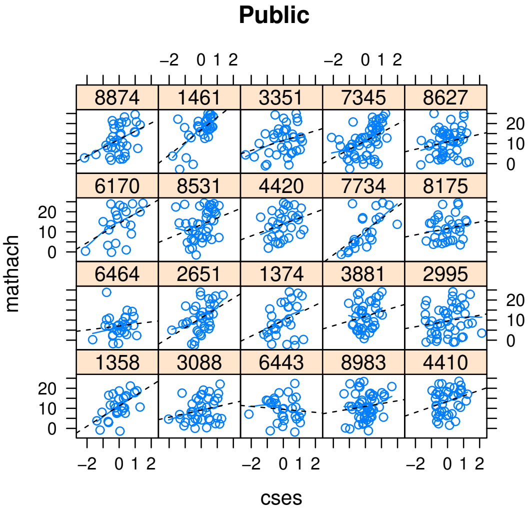

Figure: Math achievement by centred SES for students in 20 randomly selected public schools.

Data analysis

- In each scatterplot, the broken line is the linear least-squares fit to the data, while the solid line gives a nonparametric-regression fit. The number at the top of each panel is the ID number of the school.
- Particularly given the relatively small numbers of students in individual schools, the linear regressions seem to do a reasonable job of summarizing the relationship between math achievement and SES within schools.
- Although there is substantial variation in the regression lines among schools, there also seems to be a systematic difference between Catholic and public schools: The lines for the public schools appear to have steeper slopes on average.
- "SES" in these scatterplots is expressed as deviations from the school mean SES. That is, the average SES for students in a particular school is subtracted from each individual student's SES.

- Centering SES in this manner makes the within-school intercept from the regression of math achievement on SES equal to the average math achievement score in the school:
− In the i th school we have the regressing equation

$$\text{mathach}_{ij} = \alpha_{0i} + \alpha_{1i}\left(\text{ses}_{ij} - \overline{\text{ses}}_{i.}\right) + \varepsilon_{ij}$$

where

$$\overline{\text{ses}}_i = \frac{\sum_{j=1}^{n_i} \text{ses}_{ij}}{n_i}$$

- Then the least-squares estimate of the intercept is

$$\widehat{\alpha}_{0i} = \overline{\text{mathach}}_i = \frac{\sum_{j=1}^{n_i} \text{mathach}_{ij}}{n_i}$$

- A more general point is that it is helpful for interpretation of hierarchical (and other!) models to scale the explanatory variables so that the parameters of the model represent quantities of interest.

- Having made it satisfied that linear regressions reasonably represent the within-school relationship between math achievement and SES, we fit this model by least squares to the data from each of the 160 schools.
- Here are three displays of these coefficients:
  - Figures 3 and 4 shows confidence intervals for the intercept and slope estimates for Catholic and public schools.
  - Figure 5 shows boxplots of the intercepts and slopes for Catholic and public schools.
  — It is apparent that the individual slopes and intercepts are not estimated very precisely, and there is also a great deal of variation from school to school.
  - On average, however, Catholic schools have larger intercepts (i.e., a higher average level of math achievement) and lower slopes (i.e., less of a relationship between math achievement and SES).

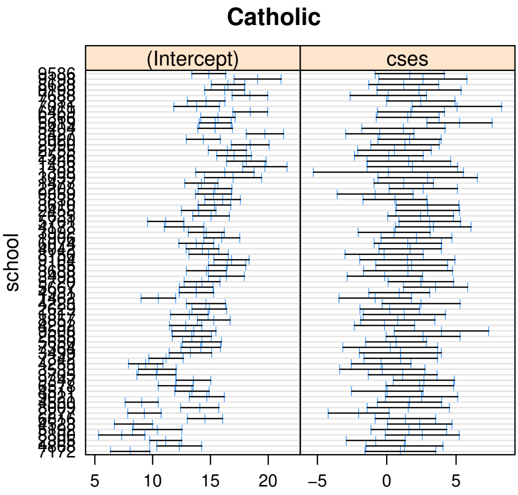

Figure: Confidence intervals for least-squares intercepts (left) and slopes (right) for the within-school regressions of math achievement on centered SES: 70 Catholic high schools.

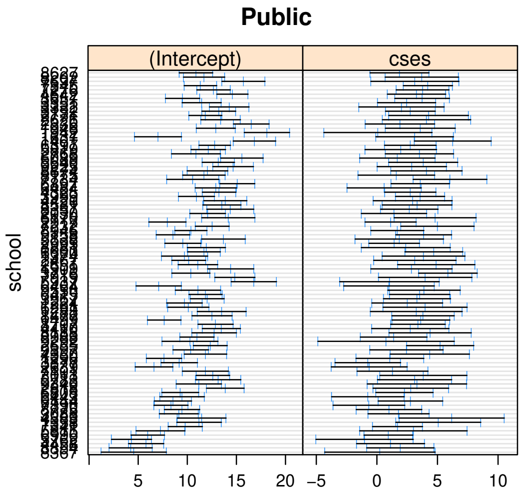

Figure: Confidence intervals for least-squares intercepts (left) and slopes (right) for the within-school regressions of math achievement on centered SES: 90 public high schools.

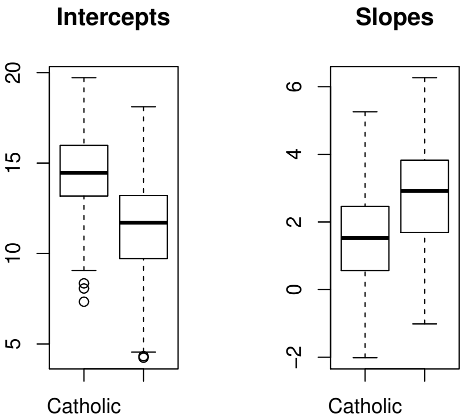

Figure: Boxplots of within-school coefficients for the least-squares re- gression of math achievement on school-centered SES, for 70 Catholic and 90 public schools.

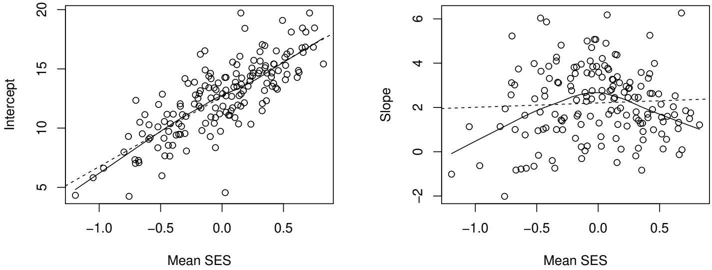

Figure: Within-school intercepts and slopes by mean SES. In each panel, the broken line is the linear least-squares fit and the solid line is from a nonparametric regression.

Review

Introduction

LMM

Modeling Hierarchical Data

Formulating a Mixed Model

Random-Effects One-Way Analysis of Variance

Random-Coefficients Regression Model

Coefficients-as-Outcomes Model

Formulating a Mixed Model

We already have a "level-1" model for math achievement:

$$\text{mathach}_{ij} = \alpha_{0i} + \alpha_{1i}\,\text{cses}_{ij} + \varepsilon_{ij}$$

where $\text{cses}_{ij} = \text{ses}_{ij} - \overline{\text{ses}}_i$ A "level-2" model relates the coefficients in the "level-1" model to characteristics of schools. - Our exploration of the data suggests the following level-2 model:

$$\alpha_{0i} = \gamma_{00} + \gamma_{01}\overline{\text{ses}}_i + \gamma_{02}\,\text{sector}_i + u_{0i}$$
$$\alpha_{1i} = \gamma_{10} + \gamma_{11}\overline{\text{ses}}_i + \gamma_{12}\overline{\text{ses}}^2_{i.} + \gamma_{13}\,\text{sector}_i + u_{1i}$$

where sector is a dummy variable, coded 1 (say) for Catholic schools and 0 for public schools.

Formulating a Mixed Model

Substituting the school-level equation into the individual-level equation produces the combined or composite model:
Except for notation, this is a mixed model in Laird-Ware form, as we can see by replacing γ s with βs and u s with b s:

$$\begin{aligned} y_{ij} = &\beta_1 + \beta_2 x_{2ij} + \beta_3 x_{3ij} + \beta_4 x_{4ij} \\ &+ \beta_5 x_{5ij} + \beta_6 x_{6ij} + \beta_7 x_{7ij} \\ &+ b_{1i} + b_{2i} z_{2ij} + \varepsilon_{ij} \end{aligned}$$

Remark

- Note that all explanatory variables in the Laird-Ware form of the model carry subscripts i for schools and j individuals within schools, even when the explanatory variable in question is constant within schools.
  - Thus, for example, $x_{2ij} = \overline{\text{ses}}_{i\cdot}$ (and so all individuals in the same school share a common value of school-mean SES).
- There is both a data-management issue here and a conceptual point:
  - With respect to data management, software that fits the Laird-Ware form of the model (such as the **lme** or **lmer** functions in **R**) requires that level-2 explanatory variables (here sector and school-mean SES, which are characteristics of schools) appear in the level-1 (i.e., student) data set $-$ much as the person $\times$ time-period data set that we employed in survival analysis with time-varying covariates required that time-constant covariates appear on the data record for each time period.

Remark

- The conceptual point is that the model can incorporate contextual effects - characteristics of the level-2 units can influence the level-1 response variable.
- Such contextual effects are of two kinds:
- Compositional effects, such as the effect of school-mean SES, which are composed from characteristics of individuals within a level-2 unit.
- Effects of characteristics of the level-2 units, such as school sector, that do not pertain to the level-1 units.
- Rather than proceeding with this relatively complicated model, let us first investigate some simpler mixed-effects models derived from it.

Review

Introduction

LMM

Modeling Hierarchical Data

Formulating a Mixed Model

Random-Effects One-Way Analysis of Variance

Random-Coefficients Regression Model

Coefficients-as-Outcomes Model

Random-Effects One-Way Analysis of Variance

- Consider the following level-1 and level-2 models:

$$\text{mathach}_{ij} = \alpha_{0i} + \varepsilon_{ij}$$
$$\alpha_{0i} = \gamma_{00} + u_{0i}$$

- The combined model is

$$\text{mathach}_{ij} = \gamma_{00} + u_{0i} + \varepsilon_{ij}$$

- In Laird-Ware form:

$$y_{ij} = \beta_1 + b_{1i} + \varepsilon_{ij}$$

This is a random-effects one-way **ANOVA** model with one fixed effect, β_1, representing the general population mean of math achievement, and two random effects:

Random-Effects One-Way Analysis of Variance

- b_{1i}, representing the deviation of math achievement in school i from the general mean; that is, $\mu_i = \beta_1 + b_{1i}$ is mean math achievement in school i.
- ε_{ij}, representing the deviation of individual j 's math achievement in school i from the school mean.
  - Two observations y_{ij} and $y_{ij'}$ in school i are not independent because they share the random effect, b_{1i} There are also two variance components for this model:
- $\psi_1^2 = \text{Var}(b_{1i})$ is the variance among school means.
- $\sigma^2 = \text{Var}(\varepsilon_{ij})$ is the variance among individuals in the same school.
  -Because the b_{1i} and ε_{ij} are assumed to be independent, variation in math scores among individuals can be decomposed into these two variance components:

$$\text{Var}(y_{ij}) = E\left[(b_{1i} + \varepsilon_{ij})^2\right] = \psi_1^2 + \sigma^2$$

[since $E(b_{1i}) = E(\varepsilon_{ij}) = 0$, and hence $E(y_{ij}) = \beta_1$]

- The intra-class correlation coefficient is the proportion of variation in individuals' scores due to differences among schools:

$$\rho = \frac{\psi_1^2}{\mathrm{Var}\left(y_{ij}\right)} = \frac{\psi_1^2}{\psi_1^2 + \sigma^2}$$

- ρ may also be interpreted as the correlation between the math scores of two individuals from the same school. That is,

$$\begin{aligned}
\mathrm{Cov}\left(y_{ij}, y_{ij'}\right) &= E\left[\left(b_{1i} + \varepsilon_{ij}\right)\left(b_{1i} + \varepsilon_{ij'}\right)\right] = E\left(b_{1i}^2\right) = \psi_1^2 \\
\mathrm{Var}\left(y_{ij}\right) &= \mathrm{Var}\left(y_{ij'}\right) = \psi_1^2 + \sigma^2 \\
\mathrm{Cor}\left(y_{ij}, y_{ij'}\right) &= \frac{\mathrm{Cov}\left(y_{ij}, y_{ij'}\right)}{\sqrt{\mathrm{Var}\left(y_{ij}\right) \times \mathrm{Var}\left(y_{ij'}\right)}} = \frac{\psi_1^2}{\psi_1^2 + \sigma^2} = \rho
\end{aligned}$$

The **lme** function in the **nlme** package in **R** provides two methods to estimate mixed-effects models (as does the **lmer** function in the **lme4** package):

- Full maximum-likelihood (ML) estimation of the model maximizes the likelihood with respect to all of the parameters of the model simultaneously (i.e., both the fixed-effects parameters and the variance components).
- Restricted (or residual) maximum-likelihood (REML) estimation integrates the fixed effects out of the likelihood and estimates the variance components; given the estimates of the variance components, estimates of the fixed effects are recovered.
- A disadvantage of ML estimates of variance components is that they are biased downwards in finite samples (much as the ML estimate of the error variance in the standard linear model is biased downwards).
- The REML estimates, in contrast, correct for loss of degrees of freedom due to estimating the fixed effects.

- The difference between the ML and REML estimates can be important when the number of "clusters" (i.e., level-2 units) in the data is small. ML and REML estimates for the current example, where there are 160 schools (level-2 units), are nearly identical:

Parameter	ML Estimate	REML Estimate
β_1	12.637	12.637
ψ_1	2.925	2.935
σ	6.257	6.257

- Note that the standard deviations (rather than the variances) of the random effects are shown.
- The estimated intra-class correlation coefficient is $\widehat{\rho} = \frac{2.935^2}{2.935^2+6.257^2} = 0.180$ and so 18 percent of the variation in students' math-achievement scores is "attributable" to differences among schools.

Review

Introduction

LMM

Modeling Hierarchical Data

Formulating a Mixed Model

Random-Effects One-Way Analysis of Variance

Random-Coefficients Regression Model

Coefficients-as-Outcomes Model

Random-Coefficients Regression Model

- Let us introduce school-centered SES into the level-1 model as an explanatory variable,

$$\text{mathach}_{ij} = \alpha_{0i} + \alpha_{1i}\,\text{cses}_{ij} + \varepsilon_{ij}$$

and allow for random intercepts and slopes in the level-2 model:

$$\begin{aligned}\alpha_{0i} &= \gamma_{00} + u_{0i}\\ \alpha_{1i} &= \gamma_{10} + u_{1i}\end{aligned}$$

- The combined model is now

$$\begin{aligned}\text{mathach}_{ij} &= (\gamma_{00} + u_{0i}) + (\gamma_{10} + u_{1i})\,\text{cses}_{ij} + \varepsilon_{ij}\\ &= \gamma_{00} + \gamma_{10}\,\text{cses}_{ij} + u_{0i} + u_{1i}\,\text{cses}_{ij} + \varepsilon_{ij}\end{aligned}$$

- In Laird-Ware form:

$$y_{ij} = \beta_1 + \beta_2 x_{2ij} + b_{1i} + b_{2i} z_{2ij} + \varepsilon_{ij}$$

This model is a random-coefficients regression model.

- The fixed-effects coefficients β_1 and β_2 represent, respectively, the average within-schools population intercept and slope.
- Because SES is centered within schools, the intercept β_1 represents the general level of math achievement in the population (in the sense of the average within-school mean).
- The model has four variance-covariance components:
- $\psi_1^2 = \text{Var}(b_{1i})$ is the variance among school intercepts (i.e., school means, because SES is school-centered).
- $\psi_2^2 = \text{Var}(b_{2i})$ is the variance among within-school slopes.
- $\psi_{12} = \text{Cov}(b_{1i}, b_{2i})$ is the covariance between within-school intercepts and slopes.
- $\sigma^2 = \text{Var}(\varepsilon_{ij})$ is the error variance around the within-school regressions.

The composite error tor individual j in school i Is

$$\zeta_{ij} = b_{1i} + b_{2i} z_{2ij} + \varepsilon_{ij}$$

- The variance of the composite error is $\operatorname{Var}(\zeta_{ij}) = E\left(\zeta_{ij}^2\right) = E\left[(b_{1i} + b_{2i} z_{2ij} + \varepsilon_{ij})^2\right] = \psi_1^2 + z_{2ij}^2 \psi_2^2 + 2 z_{2ij} \psi_{12} + \sigma^2$ - And the covariance of the composite errors for two individuals j and j' in the same school is

$$\begin{aligned} \operatorname{Cov}\left(\zeta_{ij}, \zeta_{ij'}\right) &= E\left(\zeta_{ij} \times \zeta_{ij'}\right) = E\left[(b_{1i} + b_{2i} z_{2ij} + \varepsilon_{ij})\left(b_{1i} + b_{2i} z_{2ij'} + \varepsilon_{ij'}\right]\right] \\ &= \psi_1^2 + z_{2ij} z_{2ij'} \psi_2^2 + \left(z_{2ij} + z_{2ij'}\right) \psi_{12} \end{aligned}$$

- Consequently the composite errors are heteroscedastic, and errors for individuals in the same school are correlated.
- But the composite errors for two individuals in different schools are independent.

ML and REML estimates for the model are as follows:

Parameter	ML Estimate	Std. Error	REML Estimate	Std. Error
β_1	12.636	0.244	12.636	0.245
β_2	2.193	0.128	2.193	0.128
ψ_1	2.936		2.946	
ψ_2	0.823		0.833	
ψ_{12}	0.041		0.042	
σ	6.058		6.058	

Again, the ML and REML estimates are very close.
Note that I've given standard errors only for the fixed effects.
- Standard errors for variance and covariance components can be obtained in the usual manner from the inverse of the information matrix, but tests and confidence intervals based on these standard errors tend not to be accurate.

- We can, however, test variance and covariance components by a likelihood-ratio test, contrasting the (restricted) log-likelihood for the fitted model with that for a model removing the random effects in question. For example, for the current model (say model 1), removing $b_{i2}z_{2ij}$ from the model (producing, say, model 0) implies that the SES slope is identical across schools.

Removing $b_{i2}z_{2ij}$ from the model gets rid of two variance-covariance parameters, ψ_2 and ψ_{12}.

A likelihood-ratio test for these parameters (using REML) suggests that they should not be omitted from the model:

$$\log_e L_1 = -23,357.12$$
$$\log_e L_0 = -23,362.00$$
$$G^2 = 2\left(\log_e L_1 - \log_e L_0\right) = 9.76, df = 2, p = .008$$

Cautionary Remarks:

- Because REML estimates are calculated integrating out the fixed effects, one cannot legitimately perform likelihood-ratio tests across models with different fixed effects when the models are estimated by REML.
- Likelihood-ratio for variance-covariance components across nested models with identical fixed effects are perfectly fine, however.
- The null hypothesis for the likelihood-ratio test of a variance (here ψ_2^2) sets the variance to 0, which is on the boundary of the parameter space; the p -value should be adjusted for this constraint (as explained in the reading).
- A common source of estimation difficulties in mixed models is the specification of overly complex random effects.
- Interest usually centers in the fixed effects, and it often pays to try to simplify the random-effect part of the model.

Review

Introduction

LMM

Modeling Hierarchical Data

Formulating a Mixed Model

Random-Effects One-Way Analysis of Variance

Random-Coefficients Regression Model

Coefficients-as-Outcomes Model

Coefficients-as-Outcomes Model

The regression-coefficients-as-outcomes model introduces explanatory variables at level 2 to account for variation among the level-1 regression coefficients. This returns us to the model that we originally formulated for the math-achievement data: at level 1 ,

$$\text{mathach}_{ij} = \alpha_{0i} + \alpha_{1i} \text{ cses}_{ij} + \varepsilon_{ij}$$

at level 2 ,

$$\begin{aligned}\alpha_{0i} &= \gamma_{00} + \gamma_{01}\overline{\text{ses}}_i + \gamma_{02} \text{ sector}_i + u_{0i} \\ \alpha_{1i} &= \gamma_{10} + \gamma_{11}\overline{\text{ses}}_i + \gamma_{12}\overline{\text{ses}}^2_{i.} + \gamma_{13} \text{ sector}_i + u_{1i}\end{aligned}$$

- The combined model:

$$+$$

$$\text{mathach}_{ij} = \gamma_{00} + \gamma_{01}\overline{\text{ses}}_i + \gamma_{02}\text{sector}_i + \gamma_{10}\text{cses}_{ij}$$
$$\gamma_{11}\overline{\text{ses}}_i \cdot \times \text{cses}_{ij} + \gamma_{12}\overline{\text{ses}}^2_i \times \text{cses}_{ij} + \gamma_{13} \text{ sector}_i \times \text{cses}_{ij}$$
$$+\text{u}_{0i} + u_{1i} \text{ cses}_{ij} + \varepsilon_{ij}$$

- The combined model in Laird-Ware form:

$$y_{ij} = \beta_1 + \beta_2 x_{2ij} + \beta_3 x_{3ij} + \beta_4 x_{4ij}$$
$$\beta_5 x_{5ij} + \beta_6 x_{6ij} + \beta_7 x_{7ij}$$
$$+b_{1i} + b_{2i} z_{2ij} + \varepsilon_{ij}$$

- This model has more fixed effects than the preceding randomcoefficients regression model, but the same random effects and variance components: $\psi_1^2 = \text{Var}(b_{1i}), \psi_2^2 = \text{Var}(b_{2i}), \psi_{12} = \text{Cov}(b_{1i}, b_{2i})$ and $\sigma^2 = \text{Var}(\varepsilon_{ij})$. After fitting this model to the data by REML, I tested to check whether random intercepts and slopes are still required:

Model	Omitting	$\log_e L$
1	—	-23,247.70
2	ψ_1^2, ψ_{12}	-23,357.86
3	ψ_2^2, ψ_{12}	-23,247.93

Thus, the test for random intercepts is highly statistically significant,

$$G^2 = 219.86, df = 2, p \simeq 0$$

But the test for random slopes is not, $G^2 = 0.46, df = 2, p = .80$:
Apparently, the level-2 explanatory variables do a sufficiently good job of accounting for differences in slopes that the variance component for slopes is no longer needed.
- The same caveat as before applies: These p -values should be adjusted for constraining $\psi_1^2 = 0$ and $\psi_2^2 = 0$

Refitting the model removing $b_{i2}z_{2ij}$ produces the following REML estimates:

Parameter	Term	REML Estimate	Std. Error
β_1	intercept	12.128	0.199
β_2	ses $_i$	5.337	0.369
β_3	sector $_i$	1.225	0.306
β_4	cses $_{ij}$	3.140	0.166
β_5	$\overline{\text{ses}}_{i.} \times \text{cses}_{ij}$	0.755	0.308
β_6	$\overline{\text{ses}}^2_{i.} \times \text{cses}_{ij}$	-1.647	0.575
β_7	sector $_i \times \text{cses}_{ij}$	-1.516	0.237
ψ_1	(intercept)	1.541	
σ	(ε_{ij})	6.060	

These estimates, all of which are statistically significant, have the following interpretations:
$-\widehat{\beta}_1 = 12.128$ is the estimated general level of math achievement in public schools (where the dummy variable sector is coded 0) at mean school SES. The interpretation of this coefficient depends upon the fact that ses $_i$. (school SES) is centered to a mean of 0 across schools.
$-\widehat{\beta}_2 = 5.337$ is the estimated increase in mean math achievement associated with a one-unit increase in school SES.
$-\widehat{\beta}_3 = 1.225$ is the estimated difference in mean math achievement between Catholic and public schools at fixed levels of school SES.
$-\widehat{\beta}_1, \widehat{\beta}_2$, and $\widehat{\beta}_3$, therefore, describe the between-schools regression of mean math achievement on school characteristics.

- Figure 7 shows how the coefficients $\widehat{\beta}_4, \widehat{\beta}_5, \widehat{\beta}_6$, and $\widehat{\beta}_7$ combine to produce the level-1 (i.e., within-school) coefficient for SES.
· At fixed levels of school SES, individual SES is more positively related to math achievement in public than in Catholic schools.
· The maximum positive effect of individual SES is in schools with a slightly higher than average SES level; the effect declines at low and high levels of school SES, and becomes negative at the lowest levels of school SES.
-An alternative, and more intuitive representation of the fitted model is shown in Figure 8, which graphs the fitted within-school regression of math achievement on centered SES for Catholic and public schools and for three levels of school SES: -0.7 (the approximate 5th percentile of school SES), 0 (the median), and 0.7 (the 95 th percentile).

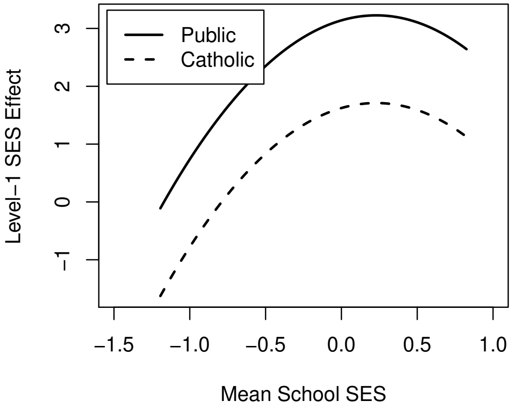

Figure: The level-1 effect of SES as a function of type of school (Catholic or public) and mean school SES.

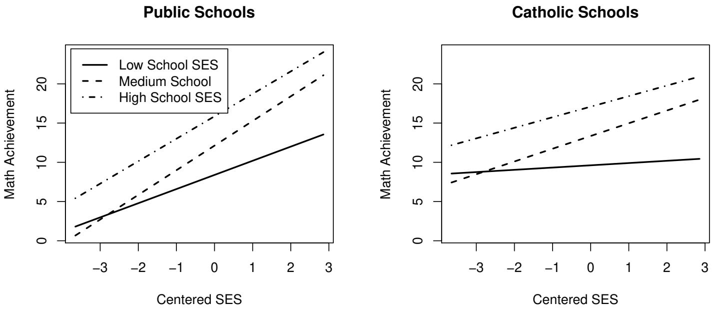

Figure: Fitted within-school regressions of math achievement on centred SES for public and Catholic schools at three levels of mean school SES