Introduction to Hierarchical Linear Models/Multilevel Analysis

Edps/Psych/Soc 589

Carolyn J. Anderson

Department of Educational Psychology

ILLINOIS

©Board of Trustees, University of Illinois

January 13, 2021

Overview

- Hierarchical Linear Models
- Multilevel Analysis using Linear Mixed Models
- Variance Components Analysis
- Random coefficients Models
- Growth curve analysis

All are special cases of Generalized Linear Mixed Models (GLMMs)

Reading:
Snijders & Bosker (2012) — chapters 1 & 2

Definition of Multilevel Analysis

- Snijders & Bosker (2012):

  > *Multilevel analysis is a methodology for the analysis of data with complex patterns of variability, with a focus on nested sources of variability.*

- Wikipedia (Aug, 2014):

  Multilevel models (also hierarchical linear models, nested models, mixed models, random coefficient, random-effects models, random parameter models, or split-plot designs) are statistical models of parameters that vary at more than one level. These models can be seen as generalizations of linear models (in particular, linear regression), although they can also extend to non-linear models. These models became much more popular after sufficient computing power and software became available.[1]

- Today:
  - Data and examples
  - Range of applications
  - Multilevel Theories

Data and Examples

Children within families:

- Children with same biological parents tend to be more alike than children chosen at random from the general population.
- They are more a like because
  - Genetics
  - Environment
  - Both

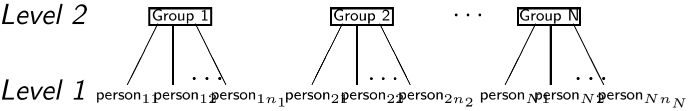
Data

Measurements on individuals (e.g., blood pressure: systolic & diastolic).

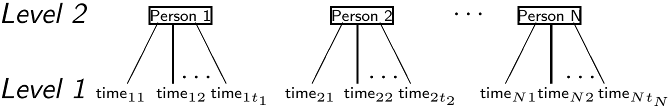
		Sources of Variability
I	Measured at the same time	Measurement error, between individuals
II	Members of same family	Measurement error, between members, between families
III	Under different conditions or over time	Measurement error, serial, between individuals
IV	Measures of members of a family over time (or different conditions	Measurement error, between individuals, between families, serial

Examples of Hierarchies

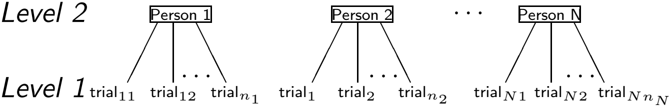
(a) Individuals within groups



(b) Longitudinal



(c) Repeated Measures



More Examples of Hierarchies

peer groups	schools	litters	companies
kids	students	animals	employees

neighborhoods	schools	clinics
families	classes	doctors
children	students	patients

A Little Terminology

Hierarchy	Levels	Labels/terminology
Schools	level 3	population, macro, primary units (first level sampled)
Classes	level 2	sub-population, secondary units, groups
Students	level 1	individuals, micro (last level sampled)

Sampling Designs

Structure of data obtained by the way data are collected.

- Observational Studies.
- Experiments.

Observational Studies

Multi-stage sampling is cost effective.

1. Take random sample from population

   e.g.(schools).
2. Take random sample from sub-population (classes).
3. Take random sample from sub-population (students).

Experiments

Hierarchies are created in the experiment.

Random assignment of individuals to treatments and create within group dependencies (compleletely randomized design).

- e.g., randomly assign patients to different clinics and due to grouping create within groups dependencies.
- e.g., randomly assign students to classes and due to grouping dependencies of individuals in the same group created.

Experiments (continue)

Grouping may initially be random but over the course of the experiment individuals become differentiated.

- Groups $\Longrightarrow$ members.
- Members $\Longrightarrow$ groups.

Analysis Must Incorporated Structure

Need to take structure of data into account because

- Invalidates most traditional statistical analysis methods (i.e., independent observations).
- Risk overlooking important group effects.
- Within group dependencies is interesting phenomenon.

People exist within social contexts and want to study and make inferences about individuals, groups, and the interplay between them.

Classic Example

- Bennett (1976): Statistically significant difference between ways of teaching reading (i.e., "formal" styles are better than others).
- Data analyzed using traditional multiple regression where students were the units of analysis.
- Atikin et al ('81): When the grouping of children into classes was accounted for, significant differences disappeared.

References

- Aitkin, M, Anderson, D, & Hinde, J. (1981). Statistical modelling of data on teaching styles. *Journal of the Royal Statistical Society, A, 144*, 419-461.
- Aitkin, M., & Longford, N. (1986). Statistical modeling issues in school effectiveness studies. *Journal of the Royal Statistical Society, A, 149*, 1-43. (with discussion).
- Goldstein, H. (1995). *Multilevel statistical models*, 2nd Edition. London: Arnold.

What happened?

- Children w/in a classroom tended to be more similar with respect to their performance.
- Each child provides less information than would have been the case if they were taught separately.
- Teacher should have been the unit of comparison.
- Students provide information regarding the effectiveness of teacher.

What Happened? (continued)

Students provide information regarding the effectiveness of teacher.

Increase the number of students per teacher,
Increase the precision of measurement of teacher.

Increase the number of teachers (with same or even fewer students),
Increase the precision of comparisons between teachers.

Unit of Analysis Problem

- Problems with ignoring hierarchical structure of data were well understood, but until recently, they were difficult to solve.
- Solution: Hierarchial linear models, along with computer software.

Hierarchical linear models are

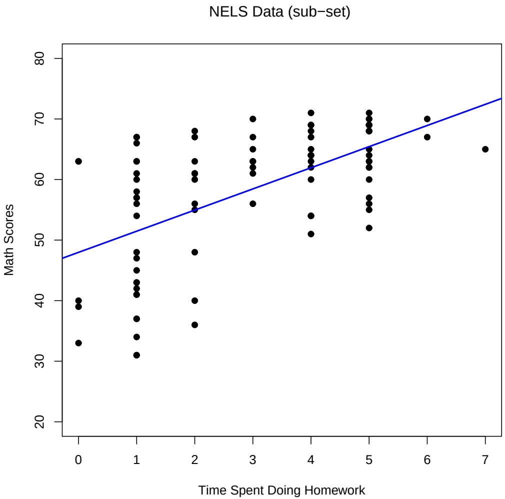
- Generalizations of traditional linear regression models.
- Special cases of them include random and mixed effects ANOVA and ANCOVA models.

A Little Example: NELS88 data

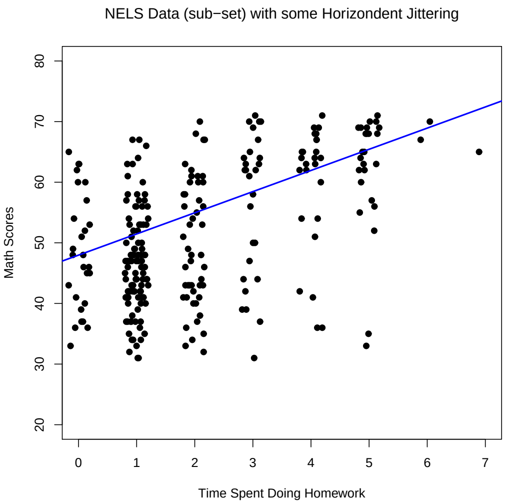
National Education Longitudinal Study — conducted by National Center for Education Statistics of the US department of Education.

- Data constitute the first in a series of longitudinal measurements of students starting in 8th grade. Data were collected Spring 1988.
- I obtained the data used here from www.stat.ucla.edu/∼deleeuw/sagebook
- From these data, we'll use 2 out of the 1003 schools.

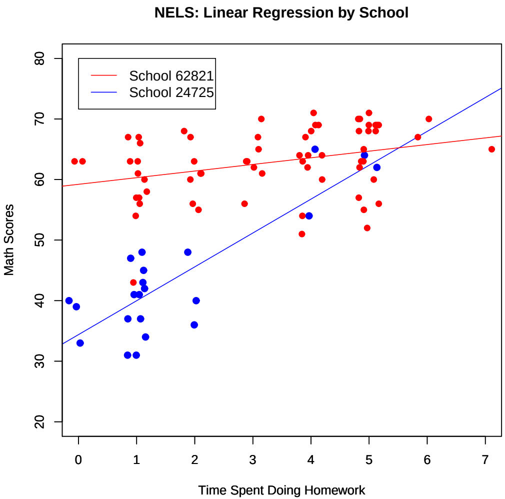
NELS88: Data from two schools



NELS88: Data from two schools with a Little Jittering



Schools 24725 and 62821 identified



Applications of Multilevel Models

An incomplete list of possibilities:

Sample survey	Measurement error
School/teacher effectiveness	Multivariate
Longitudinal	Structural Equation
Discrete responses	Event history
Random cross-classifications	Nonlinear patterns
Meta Analysis	IRT Models

Survey Samples

Multi-stage sampling often used to collect data.

geographical area (clustering of polticial attitudes)

neighborhoods (clustering of SES)

households

"Nuisance factor"

The population structure is not interesting. So, multilevel sampling is a way to collect and analyze data about higher level units.

School (teacher) Effectiveness

Students nested within schools.

- 1995 special issue *Journal of Educational and Behavioral Statistic, 20* (summer) on Hierarchical Linear Models: Problems and Prospects.
- Educational researchers interested in comparing schools w/rt student performance (measured by standardized achievement tests).
- Public accountability.
- What factors explain differences between schools.

Examples

Question: Does keeping gifted students in class or separate classes lead to better performance?

Measures available: Performance at beginning of year, performance at end of year, and aptitude.

Question: To what extent do differences in average exam results between schools accounted for by factors such as

- Organizational practices
- Characteristics of students

Advantages of multilevel approach

- Statistically efficient estimates of regression coefficients.
- Correct standard errors, confidence intervals, and significance tests.
- Can use covariates measured at any of the levels of the hierarchy.

Example with Data

Rank schools w/rt to quality (adjusting for factors such as student "intake")

Data: http://multilevel.ioe.ac.uk/

"The data come from the Junior School Project (Mortimore et al, 1988). There are over 1000 students measured over three school years with 3236 records included in this data set. Ravens test in year 1 is an ability measure."

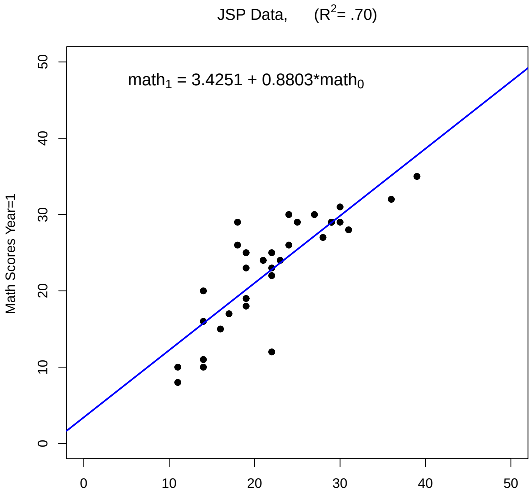
JSP Data:

Columns	Description	Coding
1-2	School	Codes from 1 to 50
14-15	Mathematics test	Score 1-40
16	Junior school year	One=0; Two=1; Three=2

Goldstein,H. (1987). *Multilevel Models in Educational and Social Research*. London, Griffin; New York, Oxford University Press.[1]
Mortimore,P.,Sammons,P.,Stoll,L.,Lewis,D. & Ecob,R. (1988). *School Matters, the Junior Years*. Wells, Open Books.
Prosser,R., Rasbash,J., and Goldstein,H.(1991). *ML3 Software for Three-level Analysis, Users' Guide for V.2*, Institute of Education, University of London.

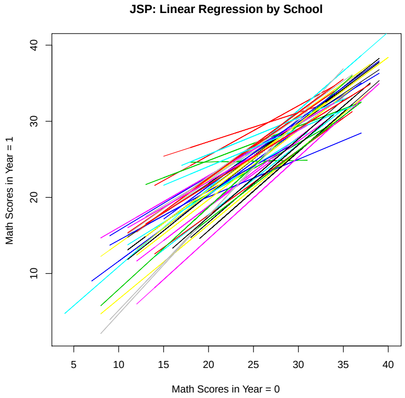
[1] The data used by Goldstein consist of measures on 728 students in 50 elementary

JSP: Level 1 Within School #1 Variation



JSP: Level 2 Between School Variation

Most R^2's between $.6$ and $.9$.



Different slopes and intercepts.

School/Teacher Effectiveness

May be OK to fit separate regressions, if

- Only a few schools each with a large number of students
- Only want to make inferences about these specific schools.

However, if view schools as random sample from a large population of schools, then need multilevel approach.

Longitudinal Data

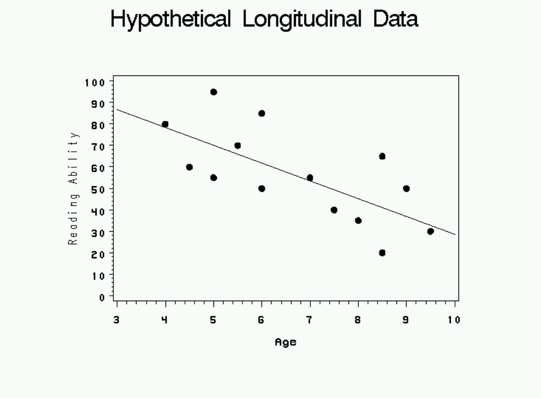
Same individuals measured on multiple occasions.

Individual
|
Occasions

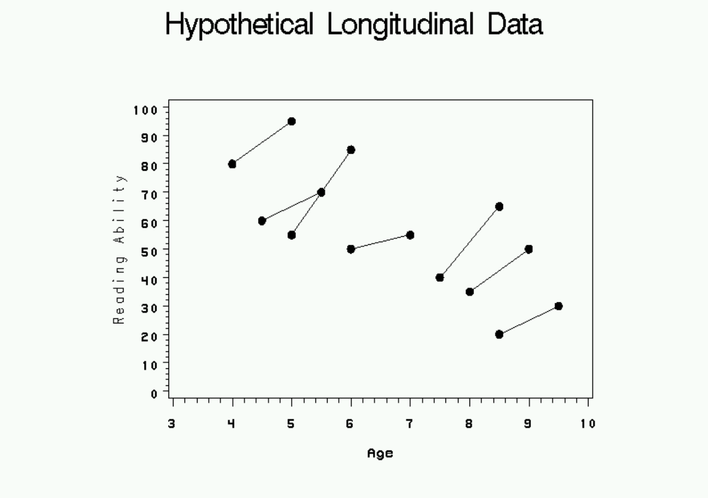
- Strong hierarchies.
- Much more variations between individuals than between occasions within individuals.

A Little (hypothetical) Example

- Response variable: reading ability
- Explanatory variable: Age
- Two measurement occasions



Hypothetical Example



Any explanations?

Longitudinal (continued)

- Traditional procedures:
    - Balanced designs (no missing data)
    - All measurement occasions the same for all individuals.
- Multilevel modeling allows:
    - Different occasions for different individuals.
    - Different number of observations per individual.
    - Build in particular error structures within individuals (eg, auto-correlated errors).
    - Others....later

Discrete Response Data

The response dependent variables are discrete rather than continuous.

- School's exam pass rate (proportions).
- Graduation rate as a function of ethnic class.
- Rate of arrest from 911 calls.

Generalized linear mixed models (SAS procedures NLMIXED, GLMMIX, MDC, MCMC).

Some common IRT models are generalized non-linear mixed models (e.g., Rasch, 2PL, others).

Multivariate Data

This is a variation of the use of hierarchical linear models for analyzing longitudinal data.

$$\begin{array}{ccccc} & & \text{individual} & & \\ & \swarrow & \swarrow & & \searrow \\ x_1 & x_2 & & \ldots & x_p \end{array}$$

Here we can have different variables and not every individual needs to have been measured on all of the variables...

Nonlinear Models

Nonlinear models that are not linear in the parameters (e.g., multiplicative).

Some kinds of growth models.

e.g., Growth spurts in children and when reach adulthood, growth levels off.

Some nonlinear patterns can be modeled by polynomials or splines, but not all (e.g., logistic, discontinuous).

Random cross-classifications

Subject X Stimuli

↘ ↙

trial

---

elementary school X high school

↘ ↙

individual

Raudenbush, S.W. (1993). A crossed random effects model for unbalanced data with applications in cross-sectional and longitudinal research. *Journal of Educational Statistics, 18*, 321–350.

Structural Equation Modeling

Including Factor analysis

group

individual

item1 . . . item20

If apply factor analysis to responses from group data, the resulting factors could represent

- Group differences
- Individual differences

Measurement Error

...in the explanatory variables at different levels.

e.g. Let Y_{ij} be measure on individual i within group/cluster j and x^*_{ij} be an explanatory variable measured with error.

$$
\begin{aligned}
Y_{ij} &= \beta_o + \beta_1 x^*_{ij} + \epsilon_{ij} \\
&= \beta_o + \beta_1 (x_i + u_j) + \epsilon_{ij} \\
&= (\beta_o + \beta_1 u_j) + \beta_1 x_i + \epsilon_{ij} \\
&= \beta^*_{oj} + \beta_1 x_i + \epsilon_{ij}
\end{aligned}
$$

See also Muthën & Asparouhov (2011) who take a latent variable approach.

Other Applications

- Image analysis (e.g., analysis of shapes, DNA patterns, computer scans).
- How is repeated measures different from longitudinal?
- How could you do a meta-analysis as a multilevel (HLM) analysis?
- For some examples of these, see http://www.dartmouth.edu/∼eugened (Demidenko, Eugene *Mixed Models: Theory and Applications*. NY: Wiley).

Multilevel Theories and Propositions

From Snijders & Bosker

Handy device:

Macro-level	Z	marco in capital letters
	$\ldots \searrow \ldots\ldots$	
Micro-level	$x \quad \rightarrow y$	micro in lower case

Micro-level propositions

No variables as the macro-level.
Dependency is a nuisance.

$$\begin{array}{ll} \ldots & \ldots\ldots \\ x & \to y \end{array}$$

e.g., At macro-level you've randomly sampled towns and within towns households.

$x =$occupational status,
$y =$income

Macro-level propositions

$Z \quad \rightarrow Y$

$\ldots \quad \ldots\ldots$

Z and Y are not directly observable, but are composites (averages, aggregates) of micro-level measurements, then we end up with multilevel structure.

e.g., $Z =$ wealth of area (average SES).
$Y =$ school performance (mean achievement test).

lower mean SES $\rightarrow$ lower mean achievement test scores. or $Z =$

student/teacher ratio.

Macro-Micro relations

Three basic possibilities:

- 1. Macro to micro.
- 2. Macro and micro to micro.
- 3. Macro–micro interaction.

1. Macro to Micro.

$$\begin{array}{lll} Z & & \\ \cdots & \searrow \cdots\cdots & \\ & \quad y & \end{array}$$

$y =$ math achievement
$Z =$ mean SES of students

Theory/proposition:

Higher average SES $\rightarrow$ higher math achievement

2. Macro and Micro to Micro

$$
\begin{array}{lll}
Z & & \\
\ldots & \searrow \ldots\ldots & \\
x & \rightarrow y &
\end{array}
$$

$x = \#$ of hours spent doing homework.

Theory/proposition:

- Given time spent doing homework, higher average SES $\rightarrow$ higher math achievement.
- Given average SES, more time spent doing homework $\rightarrow$ higher math achievement.

3. Macro-Micro Interaction

In the two macro-micro relations above, there is essentially a change in mean (random intercept). Here the relationship between x and y depends on Z.

$$\begin{array}{ccc} & Z & \\ \ldots & \downarrow\ldots & \ldots\ldots \\ x & \rightarrow & y \end{array}$$

$Z =$ no/ability grouping of children,

$x =$ aptitude or IQ, and $y =$ achievement.

Theory: Small effect of x when there is grouping but large effect when there is no grouping.

Emergent or micro-macro propositions

$$\begin{array}{lll} & & Z \\ \ldots & \nearrow & \ldots\ldots \\ x & & \end{array}$$

$Z =$ teacher's experience of stress.
$x =$ student achievement.

Another Example of Emergent

$$
\begin{array}{ccccccc}
 & W & & & & Z & \\
\cdots & \searrow & \cdots\cdots & & \nearrow & \cdots\cdots & \\
 & & x & \rightarrow & y & &
\end{array}
$$

W = teacher's attitude toward learning.
x = student's attitude toward learning.
y = student achievement.
Z = teacher's prestige.

Summary

Clustered/multilevel/hierarchically structured data are assumed to be

1. Random sample of macro-level units from population of macro-level units (or a representative sample).
2. Random sample of micro-level units from population of a (sampled) macro-level unit (or a representative sample).

Advantages of multilevel approach

- Takes care of dependencies in data and gives correct standard errors, confidence intervals, and significance tests.
- Statistically efficient estimates of regression coefficients.
- With clustered/multilevel/hierarchially structured data, can use covariates measured at any of the levels of the hierarchy.
- Model all levels simultaneously.
- Study contextual effects.
- Theories can be rich.

However,

- Need to modify tools used in normal linear regression.
- Models can become overwhelmingly complex.
- Estimation can be a problem.