

SW864¹: Multilevel and Longitudinal Data Analysis
Winter 2023

“Why can’t numbers be beautiful too?

We all talk of beautiful words, art, buildings
and they’re not part of the natural world, either.
An x in Algebra is no more abstract than
an idea in philosophy, just more useful.

But it can’t be use that makes the difference.
Keats found beauty in a Grecian urn,
surely practical at some time
and no one is blind to the beauty of symmetry.

We all get Blake’s awe of the tiger’s stripes.
Why not awe at Gaussian curves? Of course,
I know there is no great beauty in a single number,
in a four or a seven or an eight, but it is the same

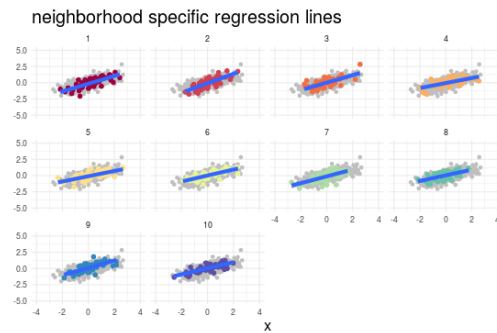
with the alphabet. Where is the wonder in a b
or a k or a t ? It is only the combinations,
the meanings, the relationships between
the letters that make the words and sounds we
love.

– And so, why can’t my numbers be beautiful to
me?
Why the scorn, the doubt in your face? Do you
think I am brittle and dusty as old paper?
Look again. See the numbers shine in my eyes.”

— Eveline Pye

“Research is formalized curiosity. It is poking and
prying with a purpose.”

— Zora Neale Hurston



“What is Imagination? First: it is the Combining
Faculty. It seizes points in common, between
subjects having no very apparent connection....
Secondly: it conceives and brings into mental
presence that which is far away, or invisible, or
which in short does not exist within our physical &
conscious cognizance....

Mathematical Science shows us what is. It is the
language of unseen relations between things. But
to use & apply that language we must be able fully
to appreciate, to feel, to seize, the unseen, the
unconscious. Imagination too shows us what is, the
is that is beyond the senses.”

— Ada Lovelace

“Mathematics is the art of giving the same name to
different things.”

— Henri Poincare

“The best thing about being a statistician is that you
get to play in everyone’s backyard.”

— John Tukey

¹ I have taught this course for over 10 years now, and up until recently, the course was numbered 832. I will no doubt refer to the course as 832 from time to time. When I refer to the course as 832, please do not simply kindly ignore my mistake, but please do offer a friendly correction. This applies more generally to material in the course: please always feel free to request clarification and correction when something seems not clearly explained.

Credit Hours: 3
Prerequisites: None
Instructor: Andy Grogan-Kaylor, MA, MSSW, PhD,
Sandra K. Danziger Collegiate Professor of Social Work
Pronouns: he, his, him
Office: 3846 School of Social Work
Phone: 734-615-3369 (seldom used)
Email: agrogan@umich.edu (best way to get in touch with me)
(please put "multilevel modeling" in the subject line)
Office Hours: TBD

Course Web site at <http://canvas.umich.edu>

Course Description and Content

Multilevel models have become a standard statistical tool for quantitative research on neighborhoods, communities and schools. Perhaps surprisingly, the multilevel model for cross-sectional data can easily accommodate longitudinal data where study participants are observed repeatedly. While this is sometimes not recognized, multilevel models for longitudinal data are closely related to other important longitudinal data models, such as fixed effects regression, an important technique for controlling for unobserved variables, and obtaining causally robust estimates.

This course focuses on the use of multilevel and longitudinal data analysis for social work research. The course is conceptualized as covering the following topics:

- 1) The multilevel model for cross-sectional data.
- 2) The extension of multilevel modeling to repeated measures in longitudinal research (i.e. multilevel growth trajectory models).
- 3) Possible additional topics based on student interest and availability of time.
 - a. Models for dichotomous outcomes.
 - b. Other panel data models such as fixed effects and random effects models.
 - c. The multilevel model for meta-analysis
 - d. Bayesian approaches: implications for small samples, for hypothesis testing, and cumulative science.

Although there are multilevel models for dichotomous outcomes, the models discussed for the majority of this course are primarily intended, and best developed, for continuous outcomes, so are most appropriate for students with research questions that can in some way be conceptualized as continuous outcomes. The course is intended for students who have solid familiarity with ordinary least squares regression, since the OLS model will be used as a starting point for discussion.

The treatment of these topics involves discussion of the underlying statistical theory, but is also focused on the application of these models with real data. Stata is intended to be the primary statistical software used in this course. Other software packages may be discussed as student interest suggests, and time allows.

Method of Instruction

Instruction is conceived as a mixture of in class lecture and discussion and time spent in the classroom working through actual data problems with statistical software in order to convey relevant content. In general, the course will have the following structure: 9-10: Asynchronous viewing of recorded materials; 10-11:30: Lecture, discussion and group activities.

Required Readings

My presentation of the material for the course will draw on articles and chapters from several different books. Given the high price of many statistical books, there is no one book that I am drawing on consistently enough to justify the purchase of a text.

Therefore, all readings for this course will be made available using <http://canvas.umich.edu>.

Course Requirements

- 1) **Two 8-10 page “mini-papers” focusing upon the statistical analysis on a research question of interest using secondary or other available data.** The first of these papers is intended to be a cross-sectional multilevel analysis and the second paper is to be a longitudinal repeated measures analysis. At a minimum, each paper should include some basic demographic controls in addition to one or two independent variables of theoretical or substantive interest. At times, your demographic variables may in fact be the relevant independent variables of theoretical interest. The two papers will be due at different points over the term. Consultation with the professor and other students will be available. Quality of writing, and clarity of presentation, will be one of the criteria used in grading. Please make sure that you cite other’s work properly and avoid plagiarism. Plagiarism²—when discovered—will be dealt with severely. Each paper should include: a review of at least 5 key pieces of background literature³; descriptive statistics; a multivariate multilevel analysis⁴; a well-documented and nicely formatted and labeled graph illustrating some bivariate aspect of your data; and a discussion of your findings and their implications. **Nota Bene:** The analysis in each paper should include testing and discussion of the meaning of a random intercept, and at least 2 random slope components, even though your final model does not need to have either a random intercept or random slope. In the second, longitudinal paper, you should further test and report upon the possible correlation of the random intercept and at least one random slope component. Further details of these assignments will be forthcoming and will also appear on the class website.

- a. **Cross-Sectional Multilevel Mini-Paper.** How are multiple factors associated with a continuous outcome of interest? This paper should make use of data appropriate for multilevel analysis, generally data containing information on individual subjects nested inside neighborhoods, countries or schools that contains some type of identifier (often anonymous) of the neighborhood, region or school. Please note that because of the way our presentation in this course is structured, it is easiest to focus on a continuous outcome. **[30% of grade]**

² Please note that for purposes of this course, plagiarism consists of six or more consecutive words, taken from another source without proper attribution. Failure upon my part to detect plagiarism does not imply approval of plagiarism.

³ Please note that this literature review is not intended to be exhaustive, merely to pave the way for the essential argument of the paper.

⁴ While I have no set format for presentation of tables of results, please pay some attention to formatting your results, and do not simply paste unformatted statistical output directly into your paper.

- b. **Longitudinal Repeated Measures Mini-Paper.** How are multiple factors associated with a continuous outcome of interest? This paper should make use of data appropriate for a longitudinal, repeated measures, analysis, based in the multilevel tradition. Data that are appropriate for this kind of analysis will contain repeated observations on multiple subjects and are often called *panel* data, that contains some type of identifier (often anonymous) of each individual in the study. Please note that because of the way our presentation in this course is structured, it is easiest to focus on a continuous outcome. **[30% of grade]**
- 2) **Class participation.** You will learn more the more you participate. Because this is a doctoral course it is important to participate actively, and especially to ask questions. I recognize that there are many variations that affect the rate of participation in this course, but a portion of your grade will depend on class participation. **[20% of grade]**
- 3) **Attendance.** You are expected to attend each class and lab session. If you miss a class session for any reason, please see me to make sure that you are grasping the material. In particular, you will be expected to complete all lab assignments, unless we make other arrangements.
- 4) **Lab Work.** We will spend some time most weeks doing lab work⁵. The purpose of the lab is to give you hands-on experience with the building blocks of multilevel and repeated measures longitudinal analysis as well as working with the specific kinds of data structures pertaining to this type of analysis. Most of the learning in lab occurs in the doing of lab. Lab is intended to be a collaborative endeavor where students work with the instructor and with each other. I will not directly grade your lab work, or be able to provide written feedback on your lab work. I am always happy to discuss lab work during lab, during office hours, or over e-mail.

A note on work handed in late: Most students turn in work in accordance with class deadlines. In order to be fair to the majority of students, I have developed the following policy: late work will be graded down by half a grade a day unless prior arrangements for an extension have been made with me. I very much understand that extenuating circumstances may arise which make it difficult to turn in work on time. All I am asking you to do is to communicate with me if you need some kind of extension so that we can work out an arrangement that is mutually agreeable.

Grades

Class Participation	20%
Choosing First Data Set	10%
First Mini-Paper	30%
Choosing and Preparing Second Data Set	10%
Second Mini-Paper	30%

Grading Scale⁶

⁵ You may use your own laptop to connect to <http://virtualsites.umich.edu>. NB that while virtualsites has considerably improved over the years, it can be a “buggy” experience. Let me know if you need help with this.

⁶ Adapted from the MSW student guide

"A" grades are given for exceptional individual performance and mastery of the material. The use of "A+", "A", and "A-" distinguishes the degree of superiority. "B" grades are given to students who demonstrate mastery of the material, at the level expected for the course. "B+" is used for students who perform just above the mastery level but not in an exceptional manner. "B-" is used for students just below the mastery level. "C" grades are given when mastery of the material is minimal. A "C-" is the lowest grade which carries credit. "D" grades indicate deficiency and carry no credit. "E" grades indicate failure and carry no credit.

Grading Criteria for Written Assignments⁷

The two mini-papers will be graded using the following criteria:

- 1) The quality of the analysis and depth of understanding of the concepts, ideas, and information presented.
- 2) The clarity of expression and organization of the paper – i.e., a logical order to the presentation of its major themes.
- 3) The appropriate use of references and resources, and the variety/range of resources referenced.
- 4) The use of proper grammar and the overall professional presentation of the paper.

Note: Please take time to proofread your writing and make sure you submit a second or third draft.

Data Sets

Both of the written paper projects for this course require you to make use of previously collected data, usually publicly available secondary data, for your analyses. I will work with you to identify suitable data. Below are some suggested data sets for the two major types of analyses that we will discuss in the course.

It may be the case that you have access to data that is not publicly available, for example data from a research project, or agency, with which you are involved. It is perfectly acceptable to use such data, but you must have a copy of this data in hand at the time that you turn in your one paragraph on each study.

Multilevel Cross-Sectional Analyses (data appropriate for these projects are generally data on individuals nested inside schools, neighborhoods or regions; such data have an identifier for the school, neighborhood or region in which the individuals are nested).

Longitudinal Repeated Measures Analyses (data appropriate for these projects are generally data on the same individuals observed repeatedly over time; such data have an identifier for each individual).

Software (you need to purchase software for this course)

- Stata (available in most campus computer labs and via virtualsites) (in class exercises will focus on the use of Stata although the statistical concepts covered transcend any one statistical software package)

⁷ Adapted from a statement by Michael Reisch.

- If time and interest permit, we may make some use of SAS Proc Mixed as well as R procedures for multilevel modeling to conduct multilevel analysis.

Suggested Ancillary Texts

Although all course readings are available on Canvas, you may find some of the following texts to be useful in your study of multilevel models. **None of the following texts are required**; they are simply listed as an aid in your education. Should you have questions, I am happy to discuss what I see as the merits of various texts at any time.

The classic text on HLM is:

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage Publications.

In my mind, despite its age, the Raudenbush and Bryk text remains one of the most pedagogically accessible introductory texts on hierarchical or multilevel models.

To explain some of the fundamentals of MLM, I will be piloting the use of two chapters from this recently updated book:

Rabe-Hesketh, S. & Skrondal (2022) *Multilevel and Longitudinal Modeling Using Stata (Fourth Edition)*. College Station, TX: Stata Press.

A few years ago, I read two very short introductory texts that turn out to provide very accessible & shorter introductions to multilevel modeling:

Luke, D. (2004). *Multilevel Modeling*. Thousand Oaks, CA: Sage Publications.

Kreft, I.G.G. & de Leeuw, J. (1998). *Introducing Multilevel Modeling*. Thousand Oaks, CA: Sage Publications.

The HLM Program Manual is also a very accessible guide to HLM.

Raudenbush, S. W., Bryk, A. S. & Congdon, R. (2001). *HLM: Hierarchical linear and nonlinear modeling*.

For longitudinal applications of multilevel modeling, the best introductory text is the first half of:

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford ; New York : Oxford University Press.

Lastly, as one moves forward in one's understanding of multilevel modeling, certain technical questions sometimes arise. I've found the best sources of answers to very technical questions to be the Raudenbush and Bryk text, or the more recent book by Rabe-Hesketh and Skrondal.

Week	Date	Topic	Readings	Lab	Assignments
1	January 5, 2023	welcome and introduction to the class			
2	January 12, 2023	Instructor @ SSWR Conference. No class			1 paragraph on cross-sectional multilevel analysis project; template to be provided on Canvas.
3	January 19, 2023	introduction to multilevel (hierarchical) modeling: the multiple equation model	Chs. 2 & 4 from Raudenbush, S. W., & Bryk, A. S. (2002). Hierarchical linear models: Applications and data analysis methods. Thousand Oaks, CA: Sage Publications.	Instructor may provide output for simple multiple equation models in HLM	
4	January 26, 2023	introduction to MLM single equation approach	Stata <i>mixed</i> documentation from StataCorp (2017). <i>Stata 15 Multilevel Mixed-Effects Linear Regression Manual</i> . College Station, TX: Stata Press. Section of Ch. 3 from Skrondahl and Rabe-Hesketh.	running simple single equation models in Stata	
5	February 2, 2023	introduction to MLM single equation approach (continued)	Ch. 4 from Skrondahl and Rabe-Hesketh.		
6	February 9, 2023	Lab Day 1	Aarts et al. (2014). A solution to dependency: using multilevel analysis to accommodate nested data, <i>Nature Neuroscience</i> , 17(4), 491-496.		

7	February 16, 2023	MLM for longitudinal applications	<p>Review <i>mixed</i> documentation</p> <p>Ch. 3 from Singer, J. D., & Willett, J. B. (2003). <i>Applied longitudinal data analysis : Modeling change and event occurrence</i>. Oxford ; New York : Oxford University Press.</p> <p>p. 62-72 from Luke, D. A. (2004). <i>Multilevel modeling</i>. Thousand Oaks, CA : Sage.</p>	Longitudinal applications	first minipaper due
8	February 23, 2023	MLM for longitudinal applications 2	<p>Ma, J., & Grogan-Kaylor, A. (2017). Longitudinal associations of neighborhood collective efficacy and maternal corporal punishment with behavior problems in early childhood. <i>Developmental Psychology</i>, 53(6), 1027–1041. https://doi.org/10.1037/dev0000308</p> <p>(focus on section on longitudinal modeling)</p>	Longitudinal applications	1 paragraph on longitudinal repeated measures multilevel analysis project
9	March 2, 2023	WINTER BREAK			
10	March 9, 2023	MLM Longitudinal 3 and/or Fixed Effects Regression and/or Visualizing Multilevel Models	TBD	Review and extension of the concepts	

11	March 16, 2023	Categorical Outcomes	Stata documentation on <code>melogit</code>		
12	March 23, 2023	Video lecture on statistical power	Video lecture	Power Analysis submitted to Instructor via e-mail	Brief paragraph conducting a power analysis of a question of interest.
13	March 30, 2023	Meta-Analysis	Walker, E., Hernandez, A. V., & Kattan, M. W. (2008). Meta-analysis: Its strengths and limitations. <i>Cleveland Clinic Journal of Medicine</i> , 75(6), 431–439.	Meta-Analysis is with ESP data	
14	April 6, 2023	Bayesian Multilevel Models or Fixed Effects Regression	Stata documentation on <code>bayes: mixed</code> https://agrogan1.github.io/Bayes/accepting-H0/accepting-H0.html https://agrogan.shinyapps.io/Thinking-Through-Bayes/ or Fixed Effects Readings TBD		
16	April 13, 2022	Lab Day 2			second minipaper due one week after last day of class.