Introduction to this Course: Multilevel Models for Longitudinal (and Other Repeated Measures) Data

- Topics:
  - Course overview
  - Features of longitudinal data
  - Features of longitudinal models
  - What can MLM do for you?

What To Expect This Semester…

- You will expand your knowledge of **longitudinal models** → analysis of *repeated* observations from same sampling *unit*
  - "Units" can be anything: persons, schools, countries, animals…
  - "Repeated" can span any length of time (milliseconds to days to years)
  - "Repeated" can also include trials: across items, conditions, etc.
- This will NOT require anxiety-provoking behaviors like:
  - Calculating things by hand—computers are always better, and more advanced statistical methods cannot be implemented by hand anyway
  - Deriving formulas or results—it's ok to trust the people who specialize in these areas to have gotten it right and use their work (for now, at least)
  - Memorizing formulas—it's ok to trust the computer programmers who have implemented various statistical techniques (for now, at least)
- It WILL require learning and implementing **new language and decision guidelines** for matching data, questions, and models
  - There will be A LOT of new vocabulary and model equations!

What To Expect This Semester…

- I will NOT:
  - Use infrequent high-stakes testing to assess your level of learning
  - Assume you know anything more than the GLM to begin with!
    - But I will make connections to other statistical approaches (MLM=SEM)

- I WILL:
  - Use **formative assessments** to help you figure out what you need to review (6 planned; 12 points for **completing them at all**)
  - Require **online homework assignments** that give you real-world practice (6 planned; 88 points for **completing them accurately**)
    - All canned data this semester, but questions about your own analyses are welcome—and appreciated!—during class or office hours
  - Link research designs, data, questions, and models explicitly
    - If we don't cover exactly the combination you need in class, ask me—odds are good I have an example from elsewhere that I can send you!

Our Responsibilities

- My job (over Zoom exclusively this semester):
  - Provide custom **lecture materials and examples** that are accurate, comprehensive, and with the necessary scaffolding for your use
  - **Answer questions** via email, in individual meetings, or in group-based office hours—you are each invited to work on homework during office hours and get immediate assistance if you want it
- Your job:
  - **Ask questions**—preferably in class, but any time is better than none!
  - **Review** the class material **frequently**, focusing on mastering the vocabulary (words and model notation), logic, and procedural skills
  - **Read**—I wrote the textbook for you! (and other articles as time permits)
  - Don't wait until the last minute to start **homework**, and don't be afraid to **ask for help** if you get stuck on one thing for more than 15 minutes
  - After your HW is done, **practice** using the techniques you are learning **on data you care about**—this will help you so much more!

Class-Sponsored Statistical Software

- **SAS** will be the primary software emphasized, as its MIXED and GLIMMIX packages are the most comprehensive and stable
- PilesOfVariance.com currently has examples of every model in my textbook using SAS, SPSS, STATA, and Mplus (and some using R)
  - Some textbook examples will be used in class, along with unique examples
  - **SPSS MIXED:** syntax is very similar to that of SAS MIXED, and it has largely the same functionality (so you can use it for course homework), but I will not be adding it to any unique course examples
  - **STATA MIXED:** can used for homework and I will try to provide syntax and output for unique class examples as my time permits
  - **R:** I am slowly (begrudgingly) learning, and I will try to provide example syntax for our class as time permits, but I don't know yet if I will be able to duplicate results in R (so that it could be used in class homework)
  - **Mplus:** syntax is very different and does not have REML estimation, so you cannot use it for class homework (but see Lecture/Example 9abc)

SAS vs. STATA: My Opinion

Activity	Winner	Commentary
Working with raw files or multiple datasets	SAS, hands down	As of STATA 15, only one dataset can be open at once—problematic for messy data management
Within-dataset manipulations	Tie, but STATA for some tasks	STATA wins for group-mean-centering, stacking, and unstacking (stay tuned)
Data analysis	Tie, but SAS for some tasks	I've had estimation problems in STATA for some MLM variants (unstructured R, crossed random effects)
Post-estimation (get predictions or simple effects)	STATA, hands down	STATA has simple yet powerful options for doing these tasks in bulk that SAS doesn't have
Automating data tasks (i.e., loops)	Tie	Both programs have ways to do this, but I only know how in SAS…

This Semester: Longitudinal Multilevel Models

- Background
  - Concepts in longitudinal modeling (and why use MLMs)
  - Review of concepts to carry forward from single-level linear models
  - Within-person analysis via Repeated Measures ANOVA

- Unconditional models: describing variance and covariance over time
  - Due to within-person fluctuation via alternative covariance structures
  - Due to within-person change via fixed and random effects of time (polynomial, piecewise, and exponential models)

- Conditional longitudinal models: adding predictors
  - Time-invariant (level-2) cross-sectional predictors
  - *If time permits, crossed random effects models for other kinds of repeated measures designs (subjects by items)*
  - Time-varying (level-1) predictors and multivariate MLMs ("M-SEM")

The Big Picture: What is this class about?

- "Longitudinal" data
  - Same individual units of analysis measured at different occasions (which can range from milliseconds to days to years)
- "Repeated measures" data (*as time permits*)
  - Same individual units of analysis measured via different items, using different stimuli, or under different conditions
- Both data types fall under a more general category of "multivariate" data of varying complexity
  - The link between them is the use of **random effects** to describe covariance of residuals from the same sampling unit

Data Requirements for Our Models

- A useful outcome variable:
  - Has an interval scale*
    - A one-unit difference means the same thing across all scale points
    - In subscales, each contributing item has an equivalent scale
    - **Other kinds of outcomes can be analyzed using generalized versions of multilevel models instead, but estimation will be more challenging*
  - Has scores with the same meaning over observations
    - Includes meaning of construct
    - Includes how items relate to the scale
    - Implies measurement invariance
- FANCY MODELS STILL CANNOT SAVE BADLY MEASURED VARIABLES OR CONFOUNDED RESEARCH DESIGNS.

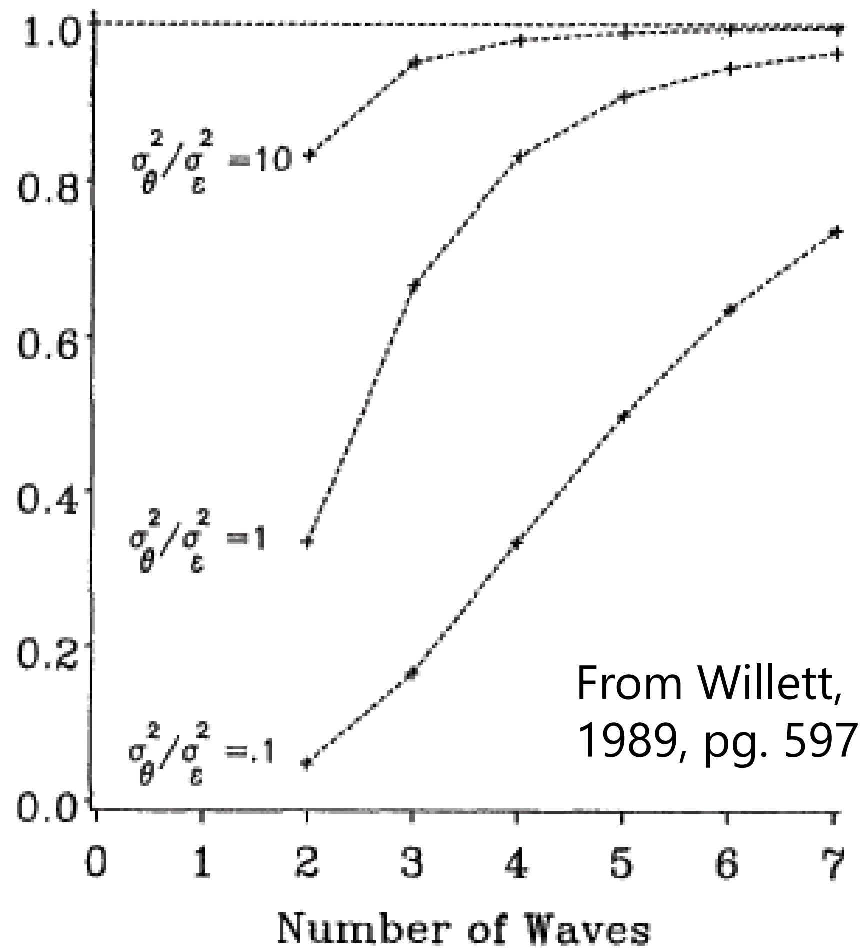
Requirements for Longitudinal Data

- Multiple **occasions** from same sampling unit (=person)
  - 2 occasions is the minimum, but just 2 can lead to problems:
    - Only 1 kind of change is observable (1 difference)
    - Can't distinguish "real" individual differences in change from error
    - Repeated measures ANOVA is just fine for 2 (complete) observations
      - Necessary assumption of "sphericity" is satisfied with only 2 observations even if compound symmetry doesn't hold
  - More data is better (with diminishing returns)
    - More occasions → better description of the form of change; more power to show effects of time-varying predictors (measured repeatedly)
    - More units (persons) → better estimates of the amount of individual differences in change; better prediction of those individual differences
    - More items/stimuli/groups → more power to show effects of differences between items/stimuli/groups

Power in Longitudinal Data

- **More occasions are better!**
  - Can examine more complex growth functions
  - Can get more reliable individual growth parameters
- **More units (people) are better!**
  - Can get more reliable estimates and prediction of individual differences

Reliability of Slopes (y-axis) by Signal-to-Noise Ratio and # Occasions



Levels of Analysis in Longitudinal Data

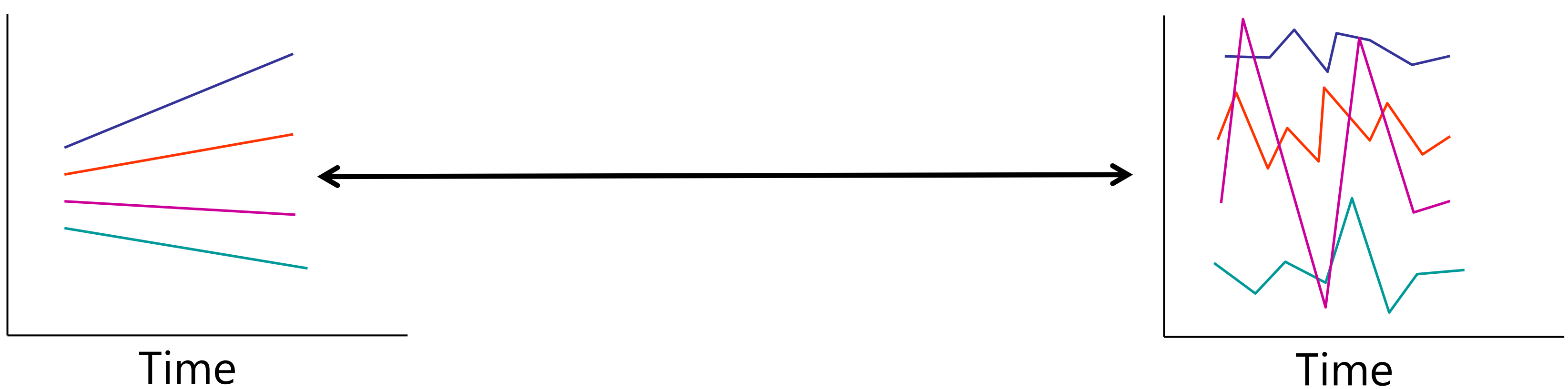
- Between-Person (BP) Variation:
  - Macro – **Level-2** – "**INTER**-individual Differences" – Time-Invariant
  - All longitudinal studies begin as cross-sectional studies
- Within-Person (WP) Variation:
  - Micro – **Level-1** – "**INTRA**-individual Differences" – Time-Varying
  - Only longitudinal studies can provide this extra information
- Longitudinal studies allow examination of both types of relationships simultaneously (and their interactions)
  - Any variable measured over time usually has both BP and WP variation
  - BP = more/less than other people; WP = more/less than one's average
- Btw, I will use **person**, but level-2 units can be anything that is measured repeatedly (like animals, schools, countries…)

A Longitudinal Data Continuum

- **Within-Person Change**: Systematic change
  - Magnitude or direction of change can be different across individuals
  - "Growth curve models" → Time is meaningfully chosen to show change
- **Within-Person Fluctuation**: No systematic change
  - Outcome just varies/fluctuates over time (e.g., emotion, stress)
  - e.g., "Ecological momentary assessments" or "daily diaries" or "intensive longitudinal data" → Time is just a way to get lots of data per individual

Pure WP Change

Pure WP Fluctuation



So what is a Multilevel Model (MLM)?

- Same as other terms you have heard of:
  - **General Linear Mixed-Effects Model** (if you are from statistics)
    - *Mixed Effects* = Fixed and Random effects
  - **Random Coefficients Model** (also if you are from statistics)
    - *Random coefficients* = Random effects = latent variables/factors
  - **Hierarchical Linear Model** (if you are from education)
    - Btw, is not the same as *hierarchical regression*

- Special cases of MLM:
  - Random Effects ANOVA or Repeated Measures ANOVA
  - (Latent) Growth Curve Model (where "Latent" implies SEM software)
  - Within-Person Fluctuation Model (e.g., for EMA or daily diary data)
  - Clustered/Nested Observations Model (e.g., for kids in schools)
  - Cross-Classified Models (e.g., teacher "value-added" models)
  - Psychometric Models (e.g., factor analysis, item response theory, SEM)

The Two Sides of Any Model

- **Model for the Means:**
  - *Aka* **Fixed Effects**, Structural Part of Model
  - What you are used to **caring about for testing hypotheses**
  - How the expected outcome for a given observation varies as a weighted function of its values of the predictor variables
    - Fixed slopes are **estimated constants** that multiply predictors

- **Model for the Variance:**
  - *Aka* **Random Effects and Residuals**, Stochastic Part of Model
  - What you are used to **making assumptions about** instead
  - How residuals are distributed and related across observations (persons, groups, time, etc.) → these relationships are called "dependency" and ***this is the primary way that multilevel models differ from general linear models*** *(e.g., regression)*

Dimensions for Organizing Models

- Outcome type: General (normal) vs. General*ized* (not normal)
- Dimensions of sampling: One (so one variance term per outcome) vs. **Multiple** (so multiple variance terms per outcome) → **OUR WORLD**

Note: Least Squares is only for GLM

- **General Linear Models:** conditionally normal outcome distribution, **fixed effects** (identity link; only one dimension of sampling)
- **General*ized* Linear Models:** **any conditional outcome distribution**, **fixed effects** through **link functions**, no random effects (one dimension)
- **General Linear Mixed Models:** conditionally normal outcome distribution, **fixed and random effects** (identity link, but multiple sampling dimensions)
- **General*ized* Linear Mixed Models:** **any conditional outcome distribution**, **fixed and random effects** through **link functions** (multiple dimensions)
  - Same concepts as for this course, but with more complexity in estimation

- "Linear" means fixed effects predict the *link-transformed* conditional mean of DV in a linear combination of (slope*predictor) + (slope*predictor)…

Options for Longitudinal Models

- Although models and software are logically separate, longitudinal data can be analyzed via multiple analytic frameworks (which are then usually tied to software):
  - "**Multilevel/Mixed-Effects Models**"
    - Dependency over time, persons, groups, etc. is modeled via random effects (multivariate through "levels" using stacked/long data)
    - Builds on GLM, generalizes easier to additional levels of analysis
  - "**Structural Equation Models**"
    - Dependency over time *only* is modeled via latent variables (single-level analysis using multivariate/wide data)
    - Generalizes easier to analysis of latent constructs, multivariate, mediation
  - Because random effects and latent variables are the same thing, many longitudinal models can be specified/estimated either way
    - And now "Multilevel Structural Equation Models" can do it all (maybe)…

What can MLM do for you?

1. **Model dependency across observations**
   - Longitudinal, clustered, and/or cross-classified data? No problem!
   - Tailor your model of sources of correlation to your data

2. **Include categorical or continuous predictors at any level**
   - Time-varying, person-level, group-level predictors for each variance
   - Explore reasons for dependency, don't just control for dependency

3. **Does not require same data structure for each person**
   - Unbalanced or missing data? No problem!

4. **You already know how (or you will soon)!**
   - Use SPSS Mixed, **SAS Mixed**, Stata, Mplus, R, HLM, MlwiN…
   - What's an intercept? What's a slope? What's a pile of variance?

1. Model Dependency

- Sources of dependency depend on the sources of **variation** created by your sampling design: residuals for outcomes from the same unit are likely to be related, which violates the GLM "independence" assumption

- **"Levels" for dependency** = "levels of random effects"
  - Sampling dimensions can be **nested** in each other
    - e.g., time within person, time within group, trial within person
  - If you can't figure out the direction of your nesting structure, odds are good you have a **crossed sampling design** instead
    - e.g., persons crossed with items, raters crossed with targets
  - To have a "level", there must be random outcome variation due to sampling that **remains** after including the model's fixed effects
    - e.g., treatment vs. control does not create another level of "group" (but it would if you had multiple treatment and multiple control groups)

Longitudinal dependency comes from…

- **Mean differences** across sampling units (e.g., persons)
  - Creates constant dependency (residual correlation) over time
  - Will be represented by a **random intercept** in our models (or could be addressed by fixed effects of person ID)

- Individual **differences in effects** (slopes) of predictors
  - Individual differences in change over time, stress reactivity
  - Creates non-constant dependency, the size of which depends on the value of the predictor at each occasion
  - Will be represented by **random slopes** in our models

- Non-constant within-person correlation for other unknown reasons (time-dependent autocorrelation)
  - Can add other patterns of correlation as needed for this

Why should we care about dependency?

- In other words, what happens if we have the wrong model for the variance (i.e., assume independent residuals instead)?
- **Validity of the tests of the predictors** depends on having the "most right" model for the variance
  - Estimates will usually be ok → come from model for the means
  - **Standard errors** (and thus ***p*-values**) can be **incorrect**
- The sources of variation that exist in your outcome will dictate **what kinds of predictors** will be useful
  - Between-Person variation needs Between-Person predictors
  - Within-Person variation needs Within-Person predictors
  - Between-whatever variation needs Between-whatever predictors…
  - Bottom line: If it don't vary, GAME OVER

2. Include categorical or continuous predictors at any level of analysis

- "ANOVA" test differences among discrete categories measured once (Between) or repeatedly measured (RM)
- "Regression" tests slopes for quantitative predictors measured once on outcomes measured once per person
- What if a predictor is measured repeatedly but can't be characterized by discrete "conditions"?
  - ANOVA or Regression won't work → you need MLM

- Some things don't change over time → time-invariant
- Some things do change over time → time-varying
- Some things are measured at higher levels
- Interactions are possible at same level or across levels

3. Does not require same data structure per person (by accident or by design)

RM ANOVA: uses **multivariate** (wide) data structure:

ID	Sex	T1	T2	T3	T4
100	0	5	6	8	12
101	1	4	7	.	11

People missing any data are excluded (data from ID 101 are not included at all)

MLM: uses **stacked** (long) data structure:

Only rows missing data are excluded

100 uses 4 cases

101 uses 3 cases

ID	Sex	Time	Y
100	0	1	5
100	0	2	6
100	0	3	8
100	0	4	12
101	1	1	4
101	1	2	7
101	1	3	.
101	1	4	11

Time can also be **unbalanced** across people such that each person can have their own measurement schedule: Time = "0.9" "1.4" "3.5" "4.2"…

4. You already know how (or you will soon)!

- If you can do GLMs, you can do MLMs
  (*and if you can do generalized linear models,*
  you can do generalized multilevel models, too)

- How do you interpret an estimate for…
  - the intercept?
  - the effect of a quantitative variable?
  - the effect of a categorical variable?
  - a variance component ("pile of variance")?