Lecture 5: Time-Invariant Predictors in Longitudinal Models

- **Predictors with Missing Data**
- Definition, Centering, and Roles in a Multilevel Model
- Review of Interpreting Interactions
- Model Specification
- Model-Building Strategies and Assessing Significance
- Time-Invariant Predictors Examples in SAS, SPSS, and STATA

Taxonomy of Missingness

(see Enders, 2010 for a thorough, readable treatment)

- **Missing Completely At Random** (**MCAR**; non-informative missingness)
  - Probability of missing Y is **unrelated** to the observed data or the missing Y values (i.e., dropout is unrelated to participant characteristics)
  - e.g., Copy machine eats the questionnaire, so questions about income are not administered to some participants
- **Missing At Random** (**MAR** ; non-informative missingness)
  - Probability of missing Y does depend on what the missing Y responses would have been, but **not after controlling for** other predictors
  - e.g., Participants of lower income choose not to answer questions about income BUT they did choose to answer questions about education, neighborhood, etc. that are related to income (MCAR after controlling for those observed variables)
- **Not Missing At Random** (**NMAR**; informative missingness)
  - Probability of missing Y depends on what the missing Y responses would have been **even after controlling for** other predictors
  - e.g., Participants of lower income will not tell you anything related to income

What does this imply for missing longitudinal data?

- If data are missing from some occasions, all is not lost!
- Missingness predictors: Person-level variables, **previous occasions**:
  - **MCAR**: probability of missingness does not depend on the persons' predictors or their outcomes for the unobserved occasions
  - **MAR**: probability of missingness does depend on the persons' predictors or their outcomes for the unobserved occasions, but you can draw correct inferences based on (controlling for) their other data
  - **MNAR**: probability of missingness does depend on their outcomes for the unobserved occasions in a way that is not predictable from the person's predictors or their previous occasions (even after controlling for)
- You will likely get different estimates from models with complete cases (survivors) only… so use *all* the data you have if at all possible!
- Now, the bad news…

Missing Data in MLM

- Common misconceptions about how MLM "handles" missing data
- Most MLM programs (e.g., MIXED) analyzes all COMPLETE CASES
  - Does NOT require listwise deletion of whole persons
  - DOES delete any incomplete cases (occasions within a person)
- Observations missing predictors OR outcomes are not included!
  - **Time** is (probably) measured for **everyone**
  - **Predictors may NOT be measured for everyone**
  - N may change due to missing data for different predictors across models
- You need to think about what predictors you want to examine PRIOR to model building, and pre-select your sample accordingly
  - Models and model fit statistics are only **directly comparable** if they include the **exact same observations**
  - Will have less statistical power as a result of removing incomplete cases!

Be careful of missing predictors!

ID	T1	T2	T3	T4	Person Pred	T1 Pred	T2 Pred	T3 Pred	T4 Pred
100	5	6	8	12	50	4	6	7	.
101	4	7	.	11	.	7	.	4	9

Row	ID	Time	DV	Person Pred	Time Pred
1	**100**	1	5	50	4
2	**100**	2	6	50	6
3	**100**	3	8	50	7
4	**100**	4	12	50	.
5	**101**	1	4	.	7
6	**101**	2	7	.	.
7	**101**	3	.	.	4
8	**101**	4	11	.	9

Only rows with complete data get used – for each model, which rows get used in MIXED?

Model with Time → DV: 1-6, 8

Model with Time, Time Pred → DV: 1-3, 5, 8

Model with Time, Person Pred → DV: 1-4

Model with Time, Time Pred, & Person Pred → DV: 1-3

So what does this mean for missing data in MLM?

- **Missing outcomes are assumed MAR**
  - Because the likelihood function is for predicted Y, just estimated on whatever Y responses a person does have (can be incomplete)
- **Missing time-varying predictors are assumed MAR-to-MCAR**
  - Would be MCAR because X is not in the likelihood function (is Y given X instead), but other occasions may have those predictors (MAR-ish)
- **Missing time-invariant predictors are assumed MCAR**
  - Because the predictor would be missing for all occasions, whole people will be deleted (may lead to bias)

- Missingness on predictors can be accommodated:
  - In Multilevel SEM with certain assumptions (≈ outcomes then)
  - Via multilevel multiple imputation – new in Mplus v 6.0+ (but careful!)
    - Must preserve all effects of potential interest in imputation model, including random effects; -2ΔLL tests are not done in same way

Lecture 5: Time-Invariant Predictors in Longitudinal Models

- Predictors with Missing Data
- **Definition, Centering, and Roles in a Multilevel Model**
- Review of Interpreting Interactions
- Model Specification
- Model-Building Strategies and Assessing Significance
- Time-Invariant Predictors Examples in SAS, SPSS, and STATA

Modeling Time-Invariant Predictors

What independent variables can be time-invariant predictors?

- Also known as 'person-level' or 'level-2' predictors (referred to as '**Z**')
- Include substantive predictors, controls, and predictors of missingness
- Can be anything that **does not change across time** (e.g., Gender)
- Can be anything that **is not likely to change across the study –** may have to argue for this (e.g., Parenting Strategies, SES)
- Can be anything that **does change across the study**…
  - But you have **only measured once**
    - Limit conclusions to variable's status at time of measurement
    - e.g., "Parenting Strategies at age 10"
  - Or **is perfectly correlated within-persons with time** (age, time to event)
    - Would use Age at Baseline, or Time to Event from Baseline instead

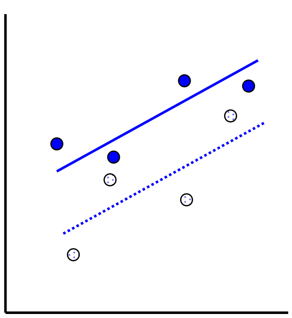
Centering Time-Invariant Predictors

- <u>Very useful to **center** all predictors so that 0 is a meaningful value:</u>
  - Same significance level of main effect, different interpretation of intercept
  - Different (more interpretable) main effects within higher-order interactions
    - With interactions, main effects → simple effects when other predictor = 0
- <u>Choices for centering **continuous** predictors:</u>
  - <u>At Mean</u>: Reference point is *average level of predictor within the sample*
    - Useful if predictor is on arbitrary metric (e.g., unfamiliar test)
  - <u>Better → At Meaningful Point</u>: Reference point is *chosen level of predictor*
    - Useful if predictor is already on a meaningful metric (e.g., age, education)
- <u>Choices for centering **categorical** predictors:</u>
  - Re-code group so that your chosen reference group = **highest category!** (which is the default in SAS and SPSS mixed models)
  - I do not recommend mean-centering categorical predictors (because who is at the mean of a categorical variable ?!?)

Fixed Effects of Level-2 Predictors in the Model for the Means

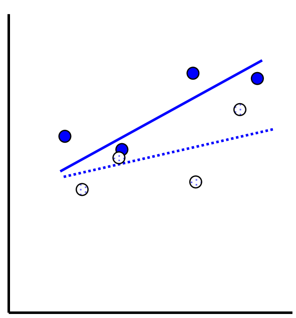
- **In Within-Person Change Models** → Adjust parts of growth curve

Main effect of X, No interaction with time



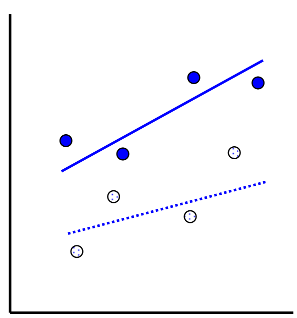
← Time →

Interaction with time, Main effect of X?



← Time →

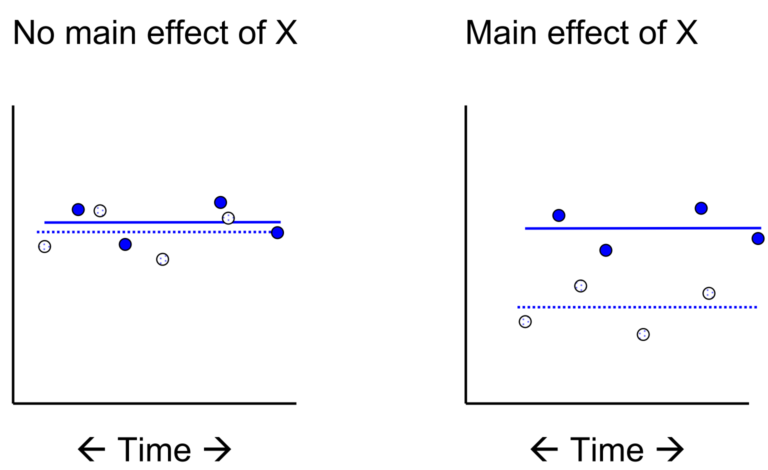
Main effect of X, and Interaction with time



← Time →

Fixed Effects of Level-2 Predictors in the Model for the Means

- In **Within-Person Fluctuation Models** → Adjust mean level

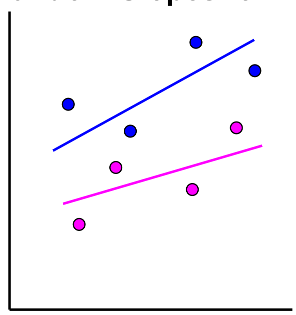


The Role of Time-Invariant Predictors in the Model for the Variance

- In addition to fixed effects in the model for the means, time-invariant predictors can allow be used to allow **heterogeneity of variance** at their level or below
- e.g., Sex as a predictor of heterogeneity of variance:
  - **At level 2**: amount of individual differences in intercepts/slopes varies across boys and girls (i.e., one group is more variable)
  - **At level 1**: amount of within-person residual variation varies across boys and girls
    - In within-person **fluctuation** model: differential fluctuation over time
    - In within-person **change** model: differential fluctuation/variation remaining after controlling for fixed and random effects of time
- These models are harder to estimate (i.e., in NLMIXED)

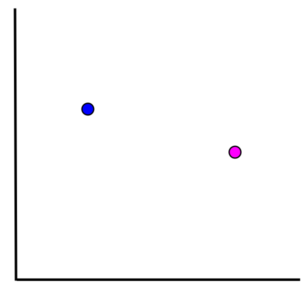
Why Level-2 Predictors Cannot Have Random Effects in 2-Level Models

Random Slopes for Time



Time
(or Any Level-1 Predictor)

Random Slopes for Sex?



Sex
(or any Level-2 Predictor)

You cannot make a line out of a dot, so level-2 effects cannot vary randomly over persons.

Lecture 5: Time-Invariant Predictors in Longitudinal Models

- Predictors with Missing Data
- Definition, Centering, and Roles in a Multilevel Model
- **Review of Interpreting Interactions**
- Model Specification
- Model-Building Strategies and Assessing Significance
- Time-Invariant Predictors Examples in SAS, SPSS, and STATA

Interpreting Interactions

<u>Interaction Effect = Moderation+ Effect</u>

- Indicates that the magnitude/direction of the effect of one predictor on an outcome "depends on" the level of another predictor
- Mathematically, either predictor can be the 'predictor' or the 'moderator' (substantive interpretation)
- Can be evaluated for any combination of categorical and continuous predictors
- Do NOT need to categorize continuous predictors:
  - In order to include their interactions with other predictors
  - In order to interpret the interaction

+ Note that "moderation" is NOT the same as "mediation"

Building the Model for the Means: Interpreting Interactions

- Main effects must be in the model with the interaction term EVEN IF THEY ARE NOT SIGNIFICANT
- Conventional wisdom has suggested main effects should not be interpreted in the presence of an interaction….
- New rule: We CAN and SHOULD interpret main effects, so long as we do so CORRECTLY…
  - As **simple effects** conditional on the level of the other variable
  - *Effect of X on Y when Z=0* OR *Effect of Z on Y when X=0*
  - The idea (and the term) of a "main effect" no longer applies
- DO NOT report "main" effects from one model and interactions from another – report simple main effects and interactions from SAME model (the APA regression table example is WRONG)

Interaction Example

- Y = Student achievement (GPA as percentage grade)
  X = Parent attitudes about education (mean on 1-5 scale)
  Z = Parent education level (years of education)

- **Model:** $GPA_i = \beta_0 + \beta_1 * Att_i + \beta_2 * Ed_i + \beta_3 * Att_i * Ed_i + e_i$
  $GPA_i = 30 + 2 * Att_i + 1 * Ed_i + 0.5 * Att_i * Ed_i + e_i$

- **Interpret β_0: GPA when Att=0 and Ed=0**
- **Interpret β_1: β_0 up by 2 for a 1-unit ΔAtt *when Ed=0***
- **Interpret β_2: β_0 up by 1 for a 1-unit ΔEd *when Att=0***
- **Interpret β_3: #1: β_1 more positive by 0.5 for a 1-unit ΔEd**
  #2: β_2 more positive by 0.5 for a 1-unit ΔAtt

- Predicted **GPA** for **attitude of 3** and **Ed of 12**?
  $66 = 30 + 2 * (3) + 1 * (12) + 0.5 * (3) * (12)$

What if we centered the predictors?
Attitudes: 3=0, Ed: 12=0

Old Equation using uncentered predictors:

$GPA_i = 30 + 2 * Att_i + 1 * Ed_i + 0.5 * Att_i * Ed_i + e_i$

Parts of new equation:

- **β_0: expected value of GPA when Att=3 and Ed=12**
  (calculated from old equation by filling values) → **66**
- **β_1: effect of Att when Ed=12**
  (start with old effect, then add interaction): **2 + (0.5*12) → 8**
- **β_2: effect of Ed when Att=3**
  (start with old effect, then add interaction): **1 + (0.5*3) → 2.5**
- **β_3: two-way interaction of Att and Ed: always 0.5 → 0.5**

New and equivalent full equation:

$GPA_i = 66 + 8 * Att_i + 2.5 * Ed_i + 0.5 * Att_i * Ed_i + e_i$

More Generally…

- Can decompose an interaction by testing simple effect of X at different levels of Z (and vice-versa)
  - You can use the model parameters to calculate what simple effects would be at another point of the other predictor…
  - But re-centering will also give you significance tests at those points
- General rules, given a three-way interaction:
  - *Simple (main) effects move the intercept*
    - 1 possible interpretation for each simple effect
    - Each simple effect is conditional on other two variables = 0
  - *The 2-ways (3 of them in a 3-way model) move the simple effects*
    - 2 possible interpretations for each two-way interaction
    - Each two-way interaction is conditional on third variable = 0
  - *The 3-way moves each of the 2-ways*
    - 3 possible interpretations of the three-way interaction
    - Is highest-order term in model, so is unconditional (applies always)

Lecture 5: Time-Invariant Predictors in Longitudinal Models

- Predictors with Missing Data
- Definition, Centering, and Roles in a Multilevel Model
- Review of Interpreting Interactions
- **Model Specification**
- Model-Building Strategies and Assessing Significance
- Time-Invariant Predictors Examples in SAS, SPSS, and STATA

Education as a Time-Invariant Predictor
Example Random Quadratic Model

- Main Effect of Education = *Education * Intercept Interaction*
  - Bump in intercept → Increase or decrease in expected Y at time 0 for every year of education
- Effect of Education on Linear Time = *Education * Linear Interaction*
  - Bump in linear slope → Increase or decrease in expected linear rate of change at time 0 for every year of education
- Effect of Education on Quadratic Time = *Education * Quad Interaction*
  - Bump in quadratic slope → Increase or decrease in half of expected acceleration/deceleration of linear rate of change for every year of education

Conditional Random Quadratic Model:
Adding a Level 2 Predictor of Ed (0=12yrs)

$$y_{ti} = \beta_{0i} + \beta_{1i}Time_{ti} + \beta_{2i}Time_{ti}^2 + e_{ti}$$

$$\beta_{0i} = \gamma_{00} + \gamma_{01}Ed_i + U_{0i}$$

β_{0i}: **Intercept** person i; γ_{00}: Sample **Intercept** for Ed=12; γ_{01}: Sample effect of Ed on the **intercept**; U_{0i}: Random **Intercept** Deviation

$$\beta_{1i} = \gamma_{10} + \gamma_{11}Ed_i + U_{1i}$$

β_{1i}: **Linear Slope** person i; γ_{10}: Sample **Linear Slope** for Ed=12; γ_{11}: Sample effect of Ed on the **linear slope**; U_{1i}: Random **Linear Slope** Deviation

$$\beta_{2i} = \gamma_{20} + \gamma_{21}Ed_i + U_{2i}$$

Fixed Effects of Time-Invariant Predictors

- Question of interest: Why do people change differently?
  - We're trying to predict individual differences in intercepts and slopes (i.e., reduce level-2 random effects variances)
  - So level-2 random effects (error) variances become '**conditional**' on predictors → actually random effects variances left over

$$\begin{aligned} \beta_{0i} &= \gamma_{00} + U_{0i} \\ \beta_{1i} &= \gamma_{10} + U_{1i} \\ \beta_{2i} &= \gamma_{20} + U_{2i} \end{aligned} \longrightarrow \begin{aligned} \beta_{0i} &= \gamma_{00} + \gamma_{01}Ed_i + U_{0i} \\ \beta_{1i} &= \gamma_{10} + \gamma_{11}Ed_i + U_{1i} \\ \beta_{2i} &= \gamma_{20} + \gamma_{21}Ed_i + U_{2i} \end{aligned}$$

  - Can calculate pseudo-R^2 for each level-2 variance pile as:

$$\text{Pseudo } R^2_{U0} = \frac{\text{intercept variance}_{\text{fewer}} - \text{intercept variance}_{\text{more}}}{\text{intercept variance}_{\text{fewer}}}$$

$$\text{Pseudo } R^2_{U1} = \frac{\text{slope variance}_{\text{fewer}} - \text{slope variance}_{\text{more}}}{\text{slope variance}_{\text{fewer}}}$$

Fixed Effects of Time-Invariant Predictors

- What about predicting effects with no random variance?
  - If the random linear effect of time is n.s., can I test interactions with time?

This should be ok to do…

$$\begin{aligned} \beta_{0i} &= \gamma_{00} + \gamma_{01}Ed_i + U_{0i} \\ \beta_{1i} &= \gamma_{10} + \gamma_{11}Ed_i + U_{1i} \\ \beta_{2i} &= \gamma_{20} + \gamma_{21}Ed_i + U_{2i} \end{aligned}$$

Is this still ok to do?

$$\begin{aligned} \beta_{0i} &= \gamma_{00} + \gamma_{01}Ed_i + U_{0i} \\ \beta_{1i} &= \gamma_{10} + \gamma_{11}Ed_i \\ \beta_{2i} &= \gamma_{20} + \gamma_{21}Ed_i \end{aligned}$$

  - **YES**, but….
  - **In theory**, if a growth effect does not vary randomly over individuals, then predictors of that effect are not needed – it has no real variance to predict
  - However, because power to detect random effects is often lower than power to detect fixed effects, **fixed effects of predictors can still be significant even if there is 'no' (≈0) variance for them to predict**
  - Small (≈0) random variance → harder to find significant interactions
  - SMEP 2011 talk result: the random effect *really* needs to not be there for Type 1 error rates for these cross-level interactions to still be ok

3 Types of Effects: Fixed, Random, and Systematically (Non-Randomly) Varying

Let's say we have a significant fixed linear effect of time. What happens after we test a sex*time interaction?

	Non-Significant Sex*Time?	Significant Sex*Time?
Random time slope initially **not** significant	Linear effect of time is **FIXED**	Linear effect of time is **systematically varying**
Random time initially sig, **not** sig. after sex*time	---	Linear effect of time is **systematically varying**
Random time initially sig, **still** sig. after sex*time	Linear effect of time is **RANDOM**	Linear effect of time is **RANDOM**

The effects of level-1 predictors (time-level) can be fixed, random, or systematically varying. The effects of level-2 predictors (person-level) can only be fixed or systematically varying (nothing to be random over…yet).

Variance Accounted For By Level-2 Time-Invariant Predictors

- **Fixed effects of level 2 predictors *by themselves*:**
  - L2 (BP) main effects (e.g., sex) reduce L2 (BP) random intercept variance
  - L2 (BP) interactions (e.g., sex by ed) also reduce L2 (BP) random intercept variance
- **Fixed effects of *cross-level interactions* (level 1* level 2):**
  - If the interacting level 1 predictor is random, any cross-level interaction with it will reduce its corresponding L2 BP random slope variance
    - e.g., if *time* is random, then sex**time,* ed**time*, and sex*ed**time* can each reduce the random linear time slope variance
  - If the interacting level 1 predictor not random, any cross-level interaction with it will reduce the L1 WP residual variance instead
    - e.g., if $time^2$ is fixed, then sex*$time^2$, ed*$time^2$, and sex*ed*$time^2$ will reduce the L1 (WP) residual variance → Different quadratic slopes from sex and ed will allow better trajectories, reduce the variance around trajectories

Lecture 5: Time-Invariant Predictors in Longitudinal Models

- Predictors with Missing Data
- Definition, Centering, and Roles in a Multilevel Model
- Review of Interpreting Interactions
- Model Specification
- **Model-Building Strategies and Assessing Significance**
- Time-Invariant Predictors Examples in SAS, SPSS, and STATA

Model-Building Strategies

- **Build UP: Start with lowest-level fixed effect**, add higher-order fixed effect interactions IF the lower-lever fixed effects are significant
  - Example: Sex predicting growth over time
    - Start with sex main effect; IF significant, then sex*time, then sex*time^2….
  - Problem: May miss higher-order interactions
    - Example: Even if sex*time^2 is significant, the effects of sex on the intercept and linear terms may not be significant, and thus you may stop too soon
- **Build DOWN: Start with highest-level fixed effect**, drop higher-order fixed effect interactions IF they are not significant
  - Example: Sex predicting growth over time
    - Start with sex*time^2, drop if non-significant, then go to sex*time, drop if non-significant, then go to main effect of sex only (→ sex*intercept)
  - Problem: Where to start?!?
    - Example: 3 predictors in a quadratic growth model: Start with X_1*X_2*X_3*time^2
    - Requires 5 main effects, 10 two-ways, 6 three-ways, 2 four-ways

Evaluating Statistical Significance of New Fixed Effects

- Fixed effects can be tested via **Wald** tests: the ratio of its estimate/SE forms a statistic we then compare to a distribution…

	Denominator DF is assumed **infinite**	Denominator DF is **estimated** instead
Numerator DF = 1	use ***z*** distribution (Mplus, STATA)	use ***t*** distribution (SAS, SPSS)
Numerator DF > 1	use χ^2 distribution (Mplus, STATA)	use ***F*** distribution (SAS, SPSS)
Denominator DF (DDFM) options	not applicable, so DDF is not given	SAS: BW and KR SAS and SPSS: Satterthwaite

Denominator DF (DDF) Methods

- **Between-Within** (DDFM=BW in SAS, not in SPSS):
  - Total DDF (T) comes from total number of observations, separated into level-2 for *N* persons and level-1 for *n* occasions
    - **Level-2 DDF** = N – #level-2 fixed effects
    - **Level-1 DDF** = Total DDF – Level-2 DDF – #level-1 fixed effects
    - Level-1 effects with random slopes still get level-1 DDF
- **Satterthwaite** (DDFM=Satterth in SAS, default in SPSS):
  - More complicated, but analogous to two-group *t*-test given unequal residual variances and unequal group sizes
  - Incorporates contribution of variance components at each level
    - Level-2 DDF will resemble Level-2 DDF from BW
    - Level-1 DDF will resemble Level-1 DDF from BW if the level-1 effect is not random, but will resemble level-2 DDF if it is random

Denominator DF (DDF) Methods

- **Kenward-Roger** (DDFM=KR in SAS, not in SPSS):
  - Adjusts the sampling covariance matrix of the fixed effects and variance components to reflect the uncertainty introduced by using large-sample techniques of ML/REML in small N samples
  - This creates different (larger) SEs for the fixed effects
  - Then uses Satterthwaite DDF, new SEs, and *t* to get *p*-values

- In an unstructured variance model, all effects use level-2 DDF
- Differences in inference not likely to matter often in practice
  - e.g., critical *t*-value at DDF=20 is 2.086, at infinite DDF is 1.960
- When in doubt, use KR (is overkill at worst, becomes Satterth)
  - I use Satterthwaite instead to maintain comparability across programs

Evaluating <u>Statistical Significance</u> of Multiple New Fixed Effects at Once

- Compare nested models with **ML -2ΔLL test**
- Useful for 'borderline' cases - example:
  - Ed*time^2 interaction at p = .04, with nonsignificant ed*time and ed*Intercept (main effect of ed) terms?
  - Is it worth keeping a marginal higher-order interaction that requires two (possibly non-significant) lower-order terms?
  - ML -2ΔLL test on 3 df: -2ΔLL must be > 7.82
  - **REML is WRONG for -2ΔLL tests for models with different fixed effects, regardless of nested or non-nested**
  - Because of this, it may be more convenient to switch to ML when focusing on modeling fixed effects of predictors

- Compare non-nested models with **ML** AIC & BIC instead

Summary…

- MLM uses ONLY rows of data that are COMPLETE – both predictors AND outcomes must be there!
  - Using whatever data you do have for each person will likely lead to better inferences and more statistical power than using only complete persons (listwise deletion)
- Time-invariant predictors modify the level-1 created growth curve → predict individual intercepts and slopes
  - Tested via fixed effect *p*-values or **ML** -2ΔLL comparisons
  - They account for random effect variances (the predictors are the reasons WHY people need their own intercepts and slopes)
  - If a level-1 effect is not random, it can still be moderated by a cross-level interaction with a time-invariant predictor…
    - … but then it will predict L1 residual variance instead

Lecture 5: Time-Invariant Predictors in Longitudinal Models

- Predictors with Missing Data
- Definition, Centering, and Roles in a Multilevel Model
- Review of Interpreting Interactions
- Model Specification
- Model-Building Strategies and Assessing Significance
- **Time-Invariant Predictors Examples in SAS, SPSS, and STATA**