Time-Invariant Predictors in Longitudinal Models

- Topics:
  - What happens to missing predictors
  - Effects of time-invariant predictors
  - Fixed vs. systematically varying vs. random effects
  - Model building strategies and assessing significance

What happens to missing predictors?

- **Incomplete data patterns in longitudinal studies**
  - Sparse missingness (within occasion)
  - Differential attrition (monotonic dropout)
  - Measurements obtained at different intervals ("unbalanced data")
  - "Planned" missing data (yes, you can do this on purpose, but carefully)
  - Often unrecognized selection bias at beginning of all studies, too

- **The goal is to make valid inferences** about population parameters despite bias introduced by attrition
  - The goal is ***not*** to recover the missing data values

- Methods used to do analyses in the presence of missing data require assumptions about the causes associated with the missingness process as well as the variables' distributions

Methods of Analysis Given Missing Data

- **What NOT to do:**
  - **NEVER EVER:** Single mean replacement or regression imputation
  - **PREFERABLY NOT:** Listwise deletion (all available whole people)

- **What to do: FIML or multiple imputation**
  - FIML = Full-information maximum likelihood → uses all the original data in estimating model, not just a summary thereof
  - MIXED and Mplus use FIML by default for missing responses (REML and ML as we know them are both Full-Information)
  - Asymptotically equivalent results given the same missingness model, but FIML is more direct than multiple imputation (and is more readily available for non-normal variables)
  - Both of these assume **Missing at Random**, though…

Categorizations of Missing Data

- If data are missing from some occasions, all is not lost!
- Missingness predictors: Person-level variables, outcomes at other observed occasions:
  - Missing Completely at Random (MCAR): probability of missingness is unrelated to what those missing responses would have been
  - **Missing at Random (MAR):** probability of missingness depends on the persons' predictors or their other observed outcomes, but you can draw correct inferences after including (controlling for) their other data
  - Missing Not At Random (MNAR): probability of missingness is systematic but is not predictable based on the information you have (everything will be some shade of wrong)

- You will likely get different estimates from models with complete cases only… so use all the data you have if possible!
- Now, the bad news…

Missing Data in MLM Software

- Common misconceptions about how MLM "handles" missing data
- Most MLM programs (e.g., MIXED) analyze only 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 may 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 −2LL, AIC, and BIC are only **directly comparable** if they include the **exact same observations (LL is sum of each height)**
  - Will have less statistical power as a result of removing incomplete cases

Be Careful of Missing Predictors!

Multivariate (wide) data → stacked (long) data

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

Beware of Missing Predictors

- Any cases missing model predictors (that are not part of the likelihood*) will not be used in that model
  - Bad for time or time-varying predictors (MARish)
  - Really bad for time-invariant predictors (listwise MCAR)

- Options for solving the problem:
  - *Bring the predictor into the likelihood (only possible in software for multivariate models, such as Mplus or SEM programs)
    - Its mean, variance, and covariances "get found" as model parameters
    - Predictor then has distributional assumptions (default multivariate normal), which may not be plausible for all predictors
  - Multiple imputation (and analysis of *each* imputed dataset)
    - Imputation also makes distributional assumptions!
    - Also requires all parameters of interest for the analysis model are in the imputation model, too (problematic for interactions or random effects)

Modeling Time-Invariant Predictors

<u>What independent variables can be time-invariant predictors?</u>

- Also known as "person-level" or "level-2" predictors
- Include substantive predictors, controls, and predictors of missingness
- Can be anything that **does not change across time** (e.g., Biological Sex)
- Can be anything that **is not likely to change across the study,** but you 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 with time** (age, time to event)
    - Would use Age at Baseline, or Time to Event *from Baseline* instead

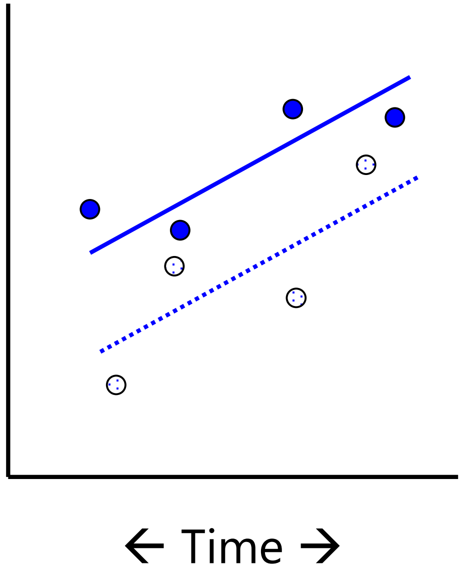
Centering Time-Invariant Predictors

- Very useful to center all predictors such that 0 is a meaningful value:
  - 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
- Choices for centering **continuous** predictors:
  - At Mean: Reference point is *average level of predictor within the sample*
    - Useful if predictor is on arbitrary metric (e.g., unfamiliar test)
  - Better → At Meaningful Point: Reference point is *chosen level of predictor*
    - Useful if predictor is already on a meaningful metric (e.g., age, education)
- Choices for centering **categorical** predictors:
  - 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 ?!?)

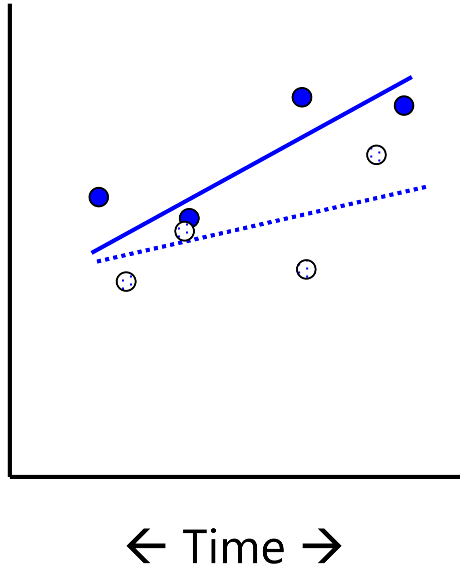
The Role of Time-Invariant Predictors in the **Model for the Means**

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

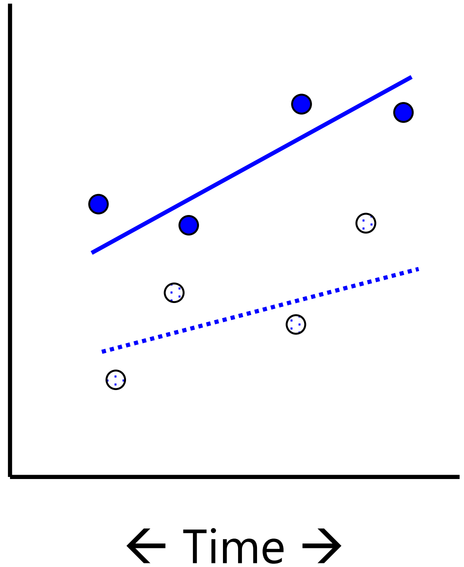
Main effect of X, No interaction with time



Interaction with time, Main effect of X?



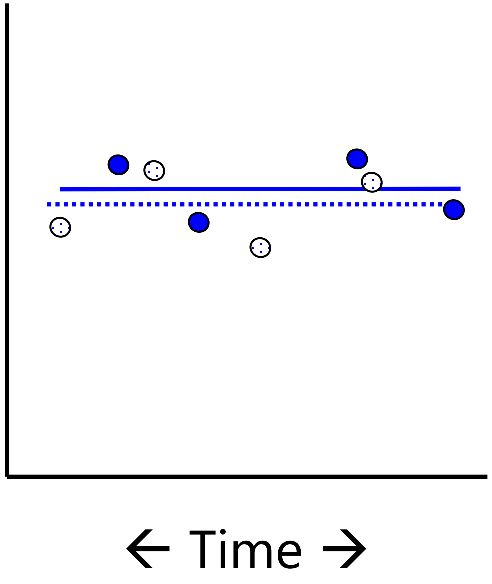
Main effect of X, and Interaction with time



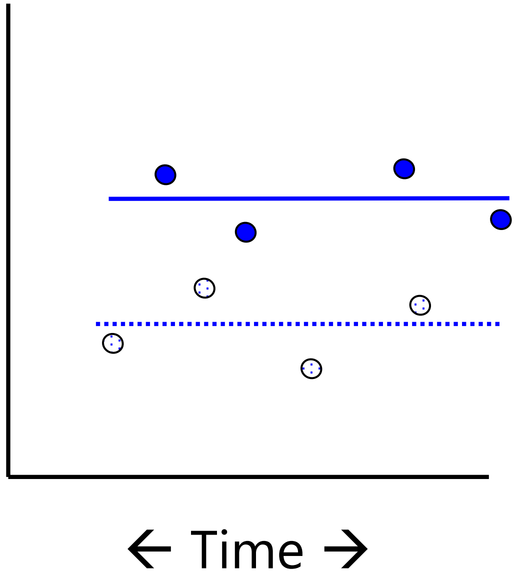
The Role of Time-Invariant Predictors in the **Model for the Means**

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

No main effect of X



Main effect of X

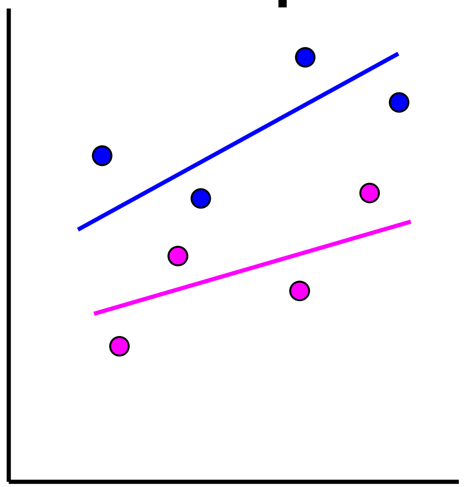


The Role of Time-Invariant Predictors in the Model for the Variance

- 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 differs between boys and girls (i.e., one group is more variable)
  - **At level 1**: amount of within-person residual variation differs between 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 and may require custom algorithms (e.g., SAS NLMIXED, also now in Mplus v 8)

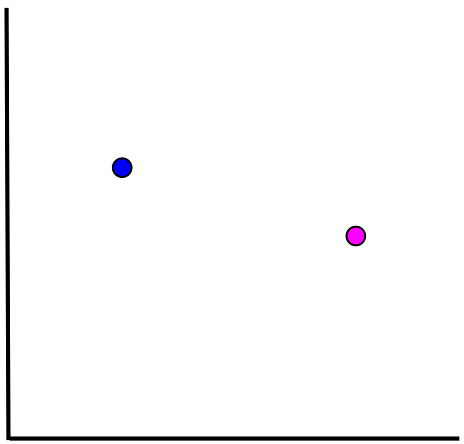
Why Level-2 Predictors Cannot Have Random Effects in 2-Level Models

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.

Education as a Time-Invariant Predictor: Example using a Random Quadratic Time Model

- Main Effect of Education = Education*Intercept Interaction
  - Moderates the intercept → Increase or decrease in expected outcome at time 0 for every year of education
- Effect of Education on Linear Time = Education*Time Interaction
  - Moderates the linear time slope → Increase or decrease in expected rate of change at time 0 for every year of education
- Effect of Education on Quadratic Time = Education*Time^2 Interaction
  - Moderates the quadratic time slope → Increase or decrease in half of expected acceleration/deceleration of linear rate of change for every year of education

Education (12 years = 0) as a Time-Invariant Predictor: Example using a Random Quadratic Time Model

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

Level 2 Equations (one per β):

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

- β_{0i}: **Intercept** for person *i*
- γ_{00}: **Fixed Intercept when Time=0 and Ed=12**
- γ_{01}: **Δ in Intercept per unit Δ in Ed**
- U_{0i}: **Random (Deviation) Intercept after controlling for Ed**

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

- β_{1i}: **Linear Slope** for person *i*
- γ_{10}: **Fixed Linear Time Slope when Time=0 and Ed=12**
- γ_{11}: **Δ in Linear Time Slope per unit Δ in Ed (=Ed*time)**
- U_{1i}: **Random (Deviation) Linear Time Slope after controlling for Ed**

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

- β_{2i}: **Quad Slope** for person *i*
- γ_{20}: **Fixed Quad Time Slope when Ed = 12**
- γ_{21}: **Δ in Quad Time Slope per unit Δ in Ed (=Ed*time^2)**
- U_{2i}: **Random (Deviation) Quad Time Slope after controlling for Ed**

Education (12 years = 0) as a Time-Invariant Predictor: Example using a Random Quadratic Time Model

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

Level 2 Equations (one per β):

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

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

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

γ_{11} and γ_{21} are known as "**cross-level**" interactions (level-1 predictor by level-2 predictor)

- Composite equation:
- $y_{ti} = (\gamma_{00} + \gamma_{01}Ed_i + U_{0i}) +$
  $(\gamma_{10} + \gamma_{11}Ed_i + U_{1i})Time_{ti} +$
  $(\gamma_{20} + \gamma_{21}Ed_i + U_{2i})Time_{ti}^2 + e_{ti}$

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 variances become 'conditional' on predictors → actually random effects variances *left over*

$$\beta_{0i} = \gamma_{00} + U_{0i} \qquad \beta_{0i} = \gamma_{00} + \gamma_{01}Ed_i + U_{0i}$$
$$\beta_{1i} = \gamma_{10} + U_{1i} \quad \longrightarrow \quad \beta_{1i} = \gamma_{10} + \gamma_{11}Ed_i + U_{1i}$$
$$\beta_{2i} = \gamma_{20} + U_{2i} \qquad \beta_{2i} = \gamma_{20} + \gamma_{21}Ed_i + U_{2i}$$

  - Can calculate pseudo-R^2 for each level-2 random effect variance between models with *fewer* versus *more* parameters as:

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

Fixed Effects of Time-Invariant Predictors

- What about predicting level-1 effects with no random variance?
  - If the random linear time slope is n.s., can I test interactions with time?

This should be ok to do…

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

Is this still ok to do?

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

  - **"NO"**: If a level-1 effect does not vary randomly over individuals, then it has "no" variance to predict (so cross-level interactions with that level-1 effect are not necessary); its SE and DDF could be inaccurate SE if $\tau^2_{U_1} \neq 0$
  - **"YES":** Because power to detect random effects is lower than power to detect fixed effects (especially with small L2n), cross-level interactions can still be significant even if there is "no" (≈0) variance to be predicted
  - Saying yes requires new vocabulary…

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 group*time interaction?

	Non-Significant Group*Time effect?	Significant Group*Time effect?
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 group*time	---	Linear effect of time is **systematically varying**
Random time initially sig, **still** sig. after group*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).

Are Systematically Varying Effects ok?

- **YES**, so long as you haven't accidentally omitted a "sizeable" random slope variance (i.e., made a Type II error)
- How to know? Consider significance of slope variance AND **Slope Reliability** (see Hoffman & Templin, under review)

$\tau^2_{U_1}$ = random slope variance
σ^2_e = residual variance
$L1n$ = L1 sample size per L2 unit
σ^2_{L1} = variance of L1 predictor

$$\textbf{SR} = \frac{\tau^2_{U_1}}{\tau^2_{U_1} + \dfrac{\sigma^2_e}{L1n * \sigma^2_{L1}}}$$

- Simulation examining L2n = 10 to 50 and L1n = 3 to 10 suggests keeping nonsignificant random slope variances with SR > .20 when using REML or SR > .15 when using ML maintains acceptable Type I errors for cross-level interactions

Variance Accounted For By Level-2 Time-Invariant Predictors

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

Variance Accounted for… For Real

- **Pseudo-R^2** is named that way for a reason… piles of variance can shift around, such that it can actually be negative
  - Sometimes a sign of model mis-specification
  - Hard to explain to readers when it happens!

- **One last simple alternative: Total R^2**
  - Generate model-predicted y's from fixed effects only (NOT including random effects) and correlate with observed y's
  - Then square correlation → total R^2
  - Total R^2 = total reduction in overall variance of y across levels
  - Can be "unfair" in models with large unexplained sources of variance

- MORAL OF THE STORY: Specify EXACTLY which kind of pseudo-R^2 you used—give the formula and the reference!!

Model-Building Strategies

- It may be helpful to examine predictor effects in **separate** models at first, including interactions with all growth terms to see the **total** pattern of effects for a single predictor
  - Question: Does age matter at all in predicting change over time?
  - e.g., random quadratic model + age, age*time, age*time^2

- Then predictor effects can be **combined** in layers in order to examine **unique** contributions (and interactions) of each
  - Question: Does age *still* matter after considering reasoning?
  - random quadratic + age, age*time, age*time^2,
    + reason, reason*time, reason*time^2
  - Potentially also + age*reason, age*reason*time, age*reason*time^2

- Sequence of predictors should be guided by theory and research questions—there may not be a single "best model"
  - One person's "control" is another person's "question", so may not end up in the same place given different orders of predictor inclusion

Evaluating Statistical Significance of Multiple New Fixed Effects at Once

- Can always do **multivariate Wald test** in REML or ML (using CONTRAST, available in all programs)
- Can only compare nested models via −2ΔLL test **in ML**
- Either is useful for 'borderline' cases—for 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?
- **REML is WRONG for −2ΔLL, AIC, or BIC comparisons for models with different fixed effects**
  - Because of this, many books (including mine) switch to ML when focusing on modeling fixed effects of predictors

Statistical Significance of Fixed Effects: What letters will I get?

Fixed effects can be tested via **Wald** tests: the ratio of its estimate/SE forms a statistic we 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 Stata 14: BW, Satt, and KR

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=Satterthwaite 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 Satterthwaite)
  - I used Satterthwaite in the book to maintain comparability across programs

Wrapping Up…

- 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 (so avoid listwise deletion if you can)

- Time-invariant predictors modify the level-1 created growth curve → predict individual intercepts and slopes
  - 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