Time-Invariant Predictors in Longitudinal Models

- Topics:
  - **Missing predictors in MLM**
  - Effects of time-invariant predictors
  - Fixed, systematically varying, and random level-1 effects
  - Model building strategies and assessing significance

Summary of Steps in Unconditional Longitudinal Modeling

For all outcomes:

1. Empty Model; Calculate ICC
2. Decide on a metric of time
3. Decide on a centering point
4. Estimate means model and plot individual trajectories

If your outcome shows systematic change:

5. Evaluate fixed and random effects of time
6. Still consider possible alternative models for the residuals (**R** matrix)

If your outcome does NOT show ANY systematic change:

5. Evaluate alternative models for the variances (**G**+**R**, or **R**)

Random Effects Models for the Variance

- Each source of correlation or dependency goes into a new variance component (or pile of variance) until each source meets the usual assumptions of GLM: normality, independence, constant variance
- **Example 2-level longitudinal model:**

Level 1 (one source of) Within-Person Variation: gets accounted for by time-level predictors

Residual Variance (σ_e^2)

FIXED effects make variance go away (explain variance).

RANDOM effects just make a new pile of variance.

Level 2 (two sources of) Between-Person Variation: gets accounted for by person-level predictors

BP Int Variance ($\tau_{U_0}^2$)

BP Slope Variance ($\tau_{U_1}^2$)

$\tau_{U_{01}}$ **covariance**

Now we get to add predictors to account for each pile!

Missing Data in MLM Software

- Common misconceptions about how MLM "handles" missing data
- Most MLM programs 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	Rows
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 MAR-to-MCAR ish**
  - Would be MCAR because X is not in the likelihood function (is Y given X instead), but other occasions may have predictors (so 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 in Mplus v 6.0+ (but careful!)
    - Must preserve all effects of potential interest in imputation model, including random effects; $-2\Delta LL$ tests are not done in same way

Time-Invariant Predictors in Longitudinal Models

- Topics:
  - Missing predictors in MLM
  - **Effects of time-invariant predictors**
  - Fixed, systematically varying, and random level-1 effects
  - Model building strategies and assessing significance

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 = **reference (0) category!** (highest is the default in SAS and SPSS; lowest is default in STATA)
  - I do not recommend mean-centering categorical predictors (because who is at the mean of a categorical variable ?!?)

Main Effects of Predictors within Interactions

- Main effects of predictors within interactions should remain in the model regardless of whether or not they are significant
  - An interaction is an over-additive (enhancing) or under-additive (dampening) effect, so *what it is additive to* must be included

- The role of a two-way interaction is to adjust its main effects…
- However, the idea of a "main effect" no longer applies… each main effect is ***conditional*** on the interacting predictor = 0

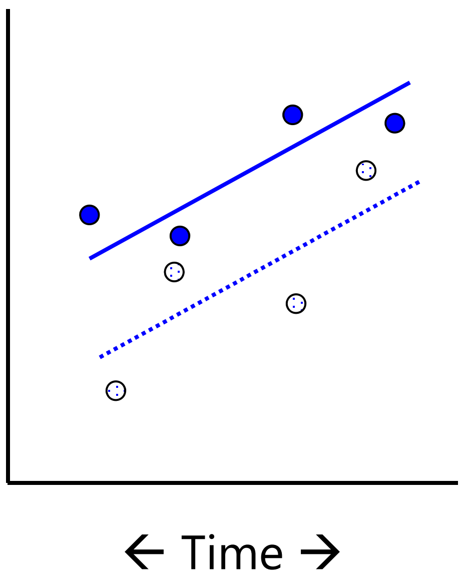
- e.g., Model of Y = W, X, Z, X*Z:
  - The effect of W is still a "main effect" because it is not part of an interaction
  - The effect of X is now the conditional main effect of X *specifically when Z=0*
  - The effect of Z is now the conditional main effect of Z *specifically when X=0*

- The trick is keeping track of what 0 means for every interacting predictor, which depends on the way each predictor is being represented, as determined by you, or by the software without you!

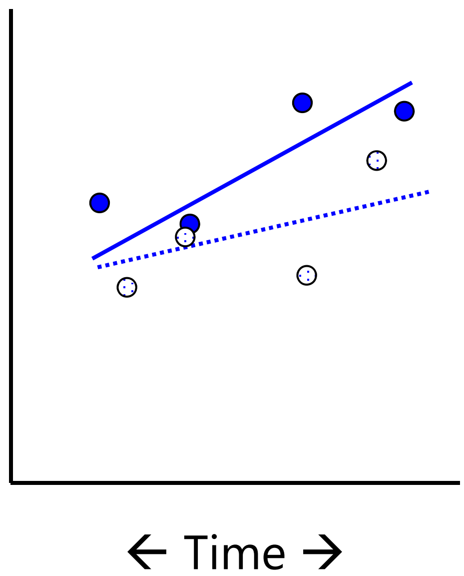
The Role of Time-Invariant Predictors in the **Model for the Means**

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

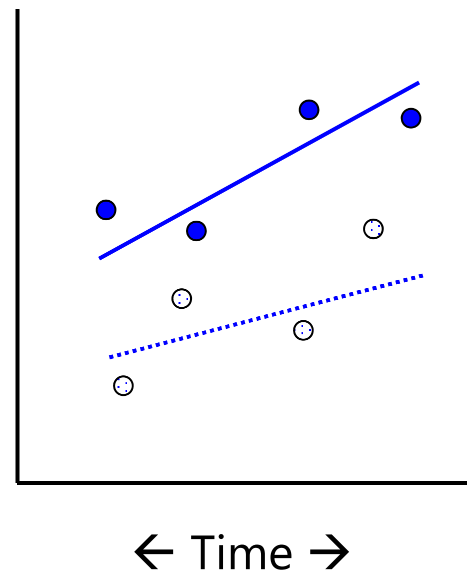
Main effect of X, No interaction with time



Interaction with time, Main effect of X?

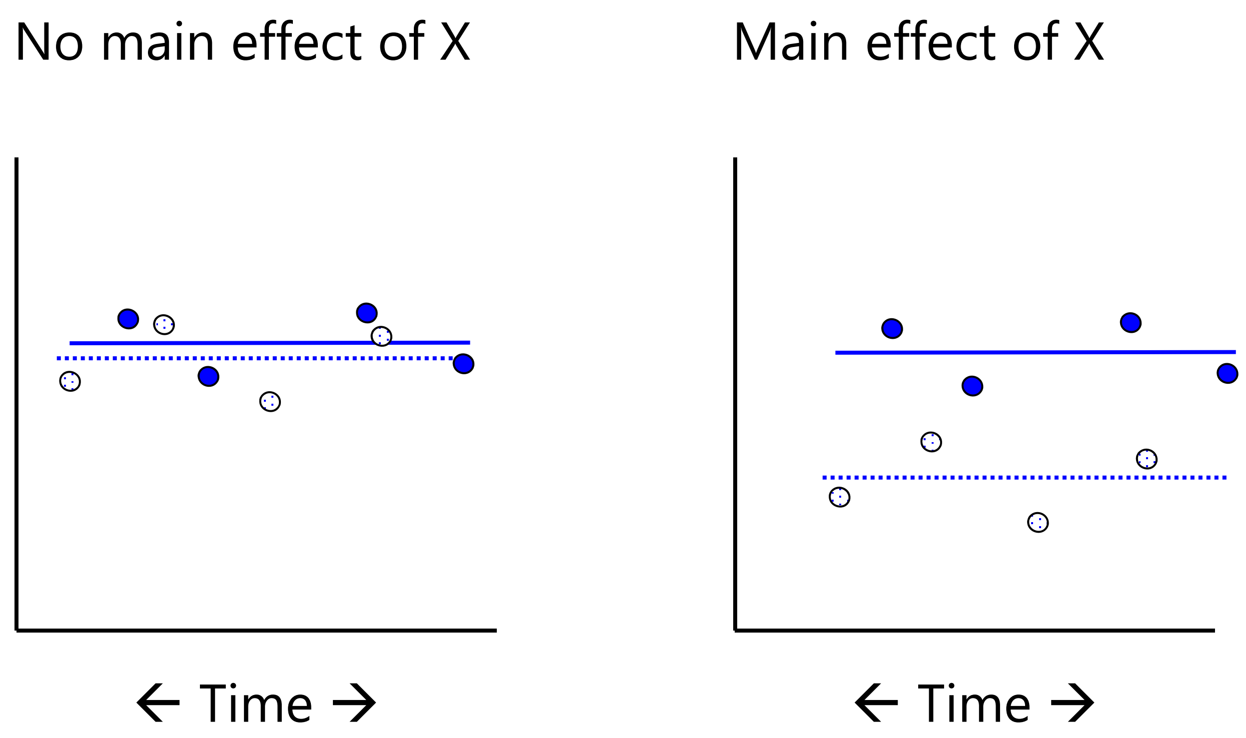


Main effect of X, and Interaction with time



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

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

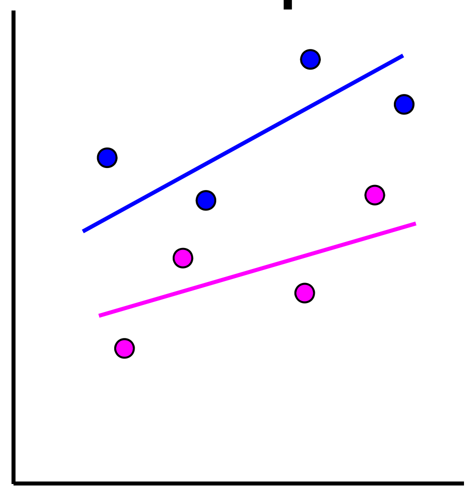


The Role of Time-Invariant Predictors in the **Model for the Variance**

- 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 software (e.g., NLMIXED in SAS)

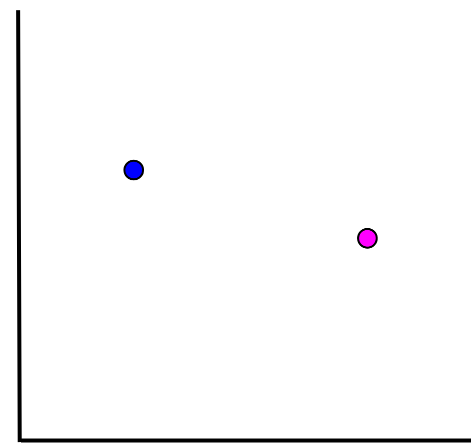
Why Level-2 Predictors Cannot Have Random Effects in 2-Level Models

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*time²)**
- 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: $\mathbf{y_{ti} = \beta_{0i} + \beta_{1i}Time_{ti} + \beta_{2i}Time_{ti}^2 + e_{ti}}$

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

Time-Invariant Predictors in Longitudinal Models

- Topics:
  - Missing predictors in MLM
  - Effects of time-invariant predictors
  - **Fixed, systematically varying, and random level-1 effects**
  - Model building strategies and assessing significance

Fixed Effects of Time-Invariant Predictors

- Question of interest: <u>Why do people change differently?</u>
  - 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*

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

  - Can calculate pseudo-R^2 for each level-2 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?

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

  - YES, surprisingly enough….
  - **In theory**, 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)
  - 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
  - Just make sure you test for random effects BEFORE testing any cross-level interactions with that level-1 predictor!

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 effect?	Significant Sex*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 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 <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 sex**time*, ed**time*, and sex*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 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

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!!

Time-Invariant Predictors in Longitudinal Models

- Topics:
  - Missing predictors in MLM
  - Effects of time-invariant predictors
  - Fixed, systematically varying, and random level-1 effects
  - **Model building strategies and assessing significance**

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

- 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 df=3: −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

Evaluating Statistical Significance of New Individual Fixed Effects

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 $\boldsymbol{\chi^2}$ 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=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 (listwise deletion)

- 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