Introduction to Multilevel Models for Longitudinal Data

- This hour:
  - Concepts and terminology
  - Modeling person dependency
  - Fixed and random intercepts
  - Fixed and random time slopes
  - Time-invariant predictors
  - Slides available at: https://www.lesahoffman.com/Workshops/index.html

- Next hour: time-varying predictors!

Sources of Longitudinal Relations

- **Between**-Person* (**BP**) Variation:
  - Macro – **Level-2** – "**INTER**-individual Differences" – Time-Invariant
  - All longitudinal studies that begin as cross-sectional studies have this

- **Within**-Person* (**WP**) Variation:
  - Micro – **Level-1** – "**INTRA**-individual Differences" – Time-Varying
  - Only longitudinal studies can provide this extra type of 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 usual

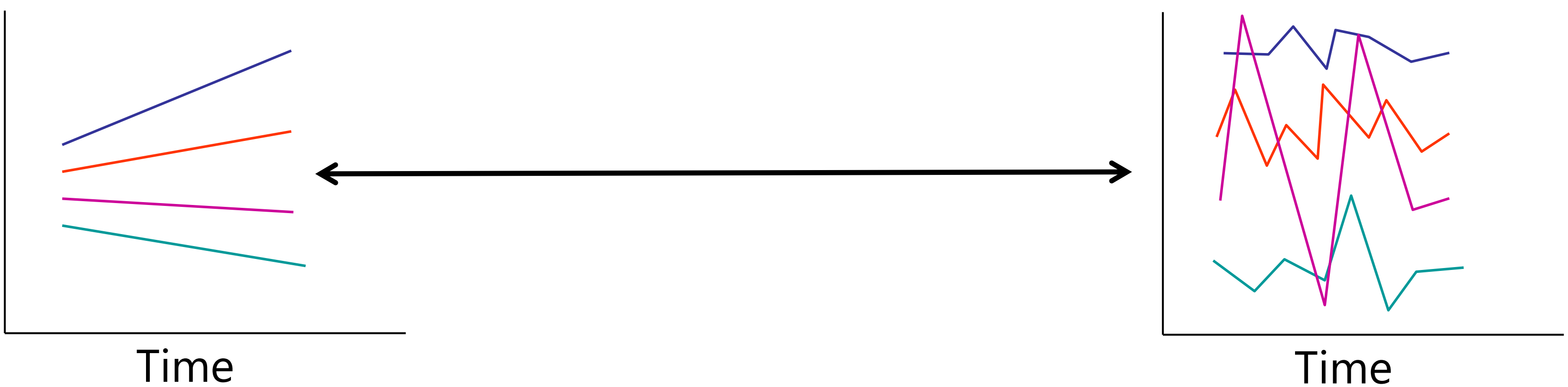
- **I will use* ***person***, *but "between" units can be anything that is measured repeatedly (like animals, schools, countries…)*

A Longitudinal Data Continuum

- **Within-Person (WP) Change**: Expect systematic effect(s) of time
  - e.g., "(Latent) Growth Curve Models" → **Time is meaningfully sampled**
  - If magnitude or direction of change differs across individuals, then the outcome's variance and covariance will change over time, too!

- **Within-Person (WP) Fluctuation**: Few *expected* effects of time
  - Outcome just varies/fluctuates over time (e.g., emotion, mood, stress)
  - **Time is just a way to get lots of data per person** (e.g., EMA studies)
  - Lends itself to questions about effects of relative changes and inconsistency

Pure WP Change | Pure WP Fluctuation



Sources of "Time" in Longitudinal Data

- What aspects of "**time**" are relevant?
  - **WP change**: e.g., time in study, age, grade, time to/from event
  - **WP fluctuation**: e.g., time of day, day of week, day in study

- Does time vary **within persons (WP)** AND **between persons (BP)**?
  - If people differ in time at the study beginning (e.g., accelerated designs), we will need to **differentiate BP time effects from WP time effects**
  - If there is more than one kind of WP "time" (e.g., occasions within days), we will need to **differentiate distinct sources of WP time effects**

- Is time *balanced* or *unbalanced*?
  - **Balanced** = **shared** measurement schedule (not necessarily equal interval)
    - Although some people may miss some occasions, making their data "incomplete"
  - **Unbalanced** = people have **different** possible time values
    - By definition, the possible outcomes are at least partially "incomplete" across persons
    - This may be a consequence of using a time metric that also varies between persons

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 (how many "piles"):**
  - *Aka* **Random Effects and Residuals**, Stochastic Part of Model
  - What you *were* used to **making assumptions about** instead
  - How **residuals are distributed and related** across sampling dimensions (persons, occasions) → these relationships are known as "**dependency**" and ***this is the primary way that longitudinal models differ from "regular" (GLM) regression models***

A Statistician's World View

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

- **<u>General Linear Models:</u>** conditionally normal outcome distribution, **fixed effects** (identity link; only one dimension of sampling)

Note: OLS is only for GLM

- **<u>General*ized* Linear Models:</u>** **any conditional outcome distribution**, **fixed effects** through **link functions**, no random effects (one dimension)
- **<u>General Linear Mixed Models:</u>** conditionally normal outcome distribution, **fixed and random effects** (identity link, but multiple sampling dimensions)
- **<u>General*ized* Linear Mixed Models:</u>** **any conditional outcome distribution**, **fixed and random effects** through **link functions** (multiple dimensions)
  - Not this week—Many of the same concepts, but with more complexity in estimation

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

Multilevel Model (MLM) Word Salad

- MLM is the same as other terms you have heard of:
  - **Linear Mixed-Effects Model** (fixed + random effects, of which intercepts and slopes are specific kinds of effects)
  - **Random Coefficients Model** (because coefficients also = effects)
  - **Hierarchical Linear Model** (not same as hierarchical regression)

- Special cases of MLM:
  - Random Effects ANOVA or Repeated Measures ANOVA
  - (Latent) Growth Curve Model (where "Latent" implies SEM software)
    - Btw, most MLMs can be equivalently estimated as single-level SEMS
  - Within-Person Fluctuation Model (e.g., for EMA or daily diary data)
    - See also "dynamic" SEM or multilevel SEM (even without measurement models!)
  - Clustered/Nested Observations Model (e.g., for kids in schools)
    - If followed over time in same group, is "clustered longitudinal model"
  - Cross-Classified Models (e.g., teacher "value-added" models)
  - Psychometric Models (e.g., factor analysis, item response theory, SEM)

The Two Sides of a General Linear Model

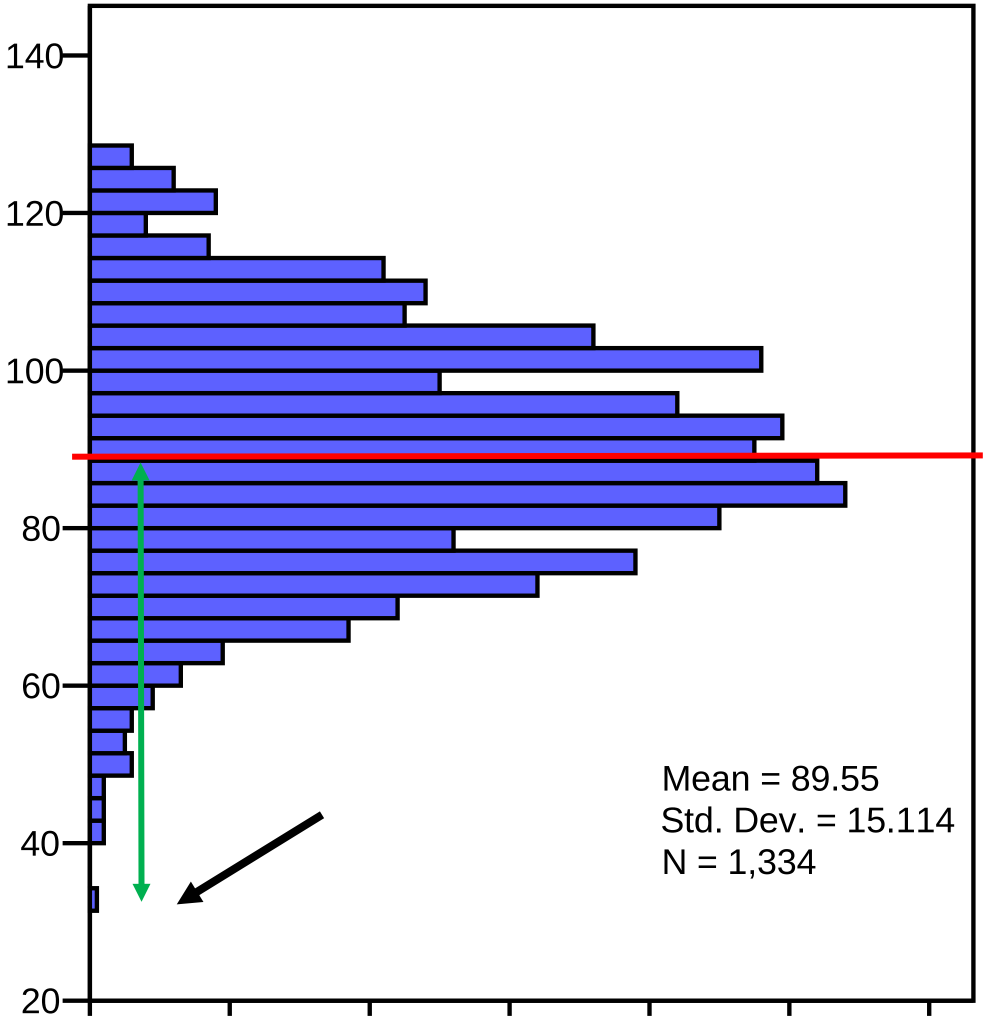
$$y_i = \beta_0 + \beta_1(x1_i) + \beta_2(x2_i) + \cdots + e_i$$

Our focus now

- **Model for the Means (→ Predicted Values):**
  - Each person's expected (predicted) outcome is a weighted linear function of his/her values on $x1_i$ and $x2_i$ (and any other predictors); each variable is measured once per person
  - **Estimated constants are called fixed effects** (here, β_0, β_1, and β_2)
  - Number of fixed effects will show up in formulas as $\boldsymbol{k}$ (so $\boldsymbol{k} = \mathbf{3}$ here)

- **Model for the Variance (→ "Piles" of Variance):**
  - $e_i \sim \mathrm{N}(0, \sigma_e^2)$ → ONE (BP) source of residual (unexplained) error
  - In GLMs, e_i has a mean of 0 with some estimated constant variance σ_e^2, is normally distributed, is unrelated to $x1_i$ and $x2_i$, and is **independent** across all observations (which is just one outcome per person here)
  - **There is only ONE source of residual variance in the above GLM because it was designed for only ONE (BP) dimension of sampling!**

An "Empty Means" General Linear Model → Single-Level (BP) Model for the Variance



$$y_i = \beta_0 + e_i$$

Filling in values:

$$32 = \underbrace{90}_{\hat{y}_i} + -58$$

$\hat{y}_i$ = "y-hat" model-predicted outcome

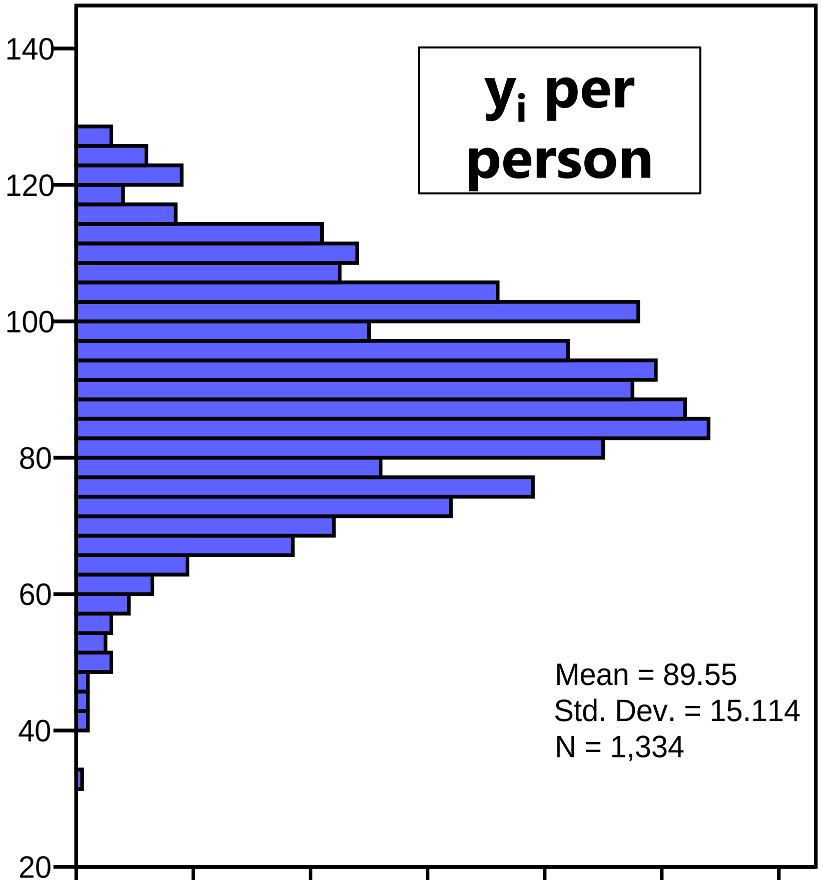
Model for the Means

y_i **residual ("error") variance:**

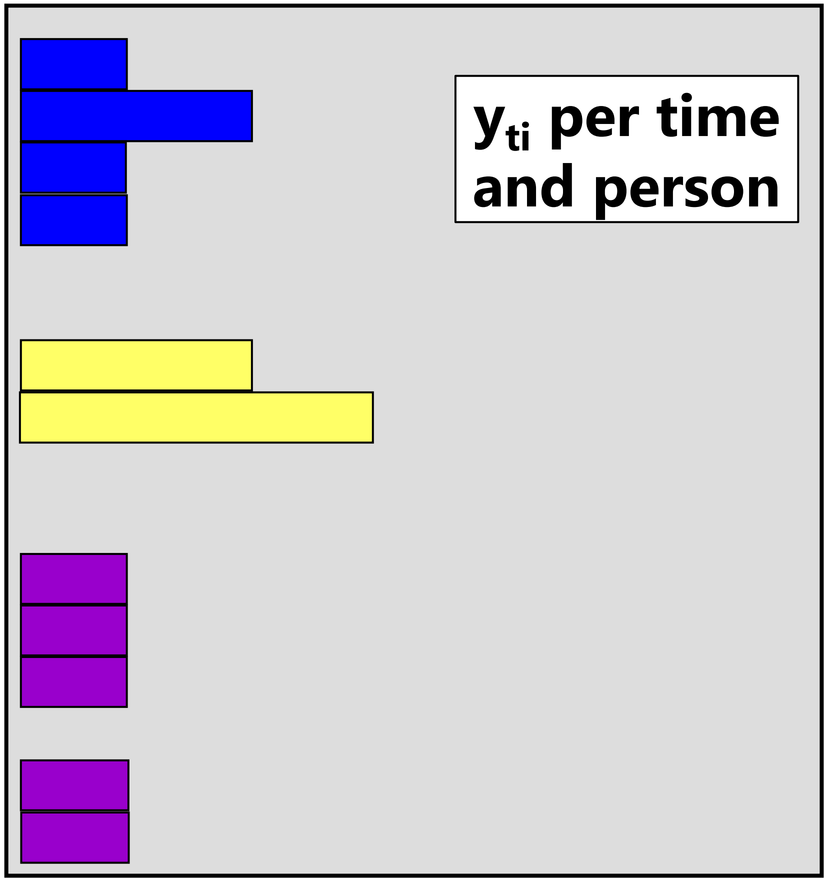
$$\frac{\sum(y_i - \hat{y}_i)^2}{N-1}$$

Adding Repeated Occasions → Two-Level (+WP) Model for the Variance

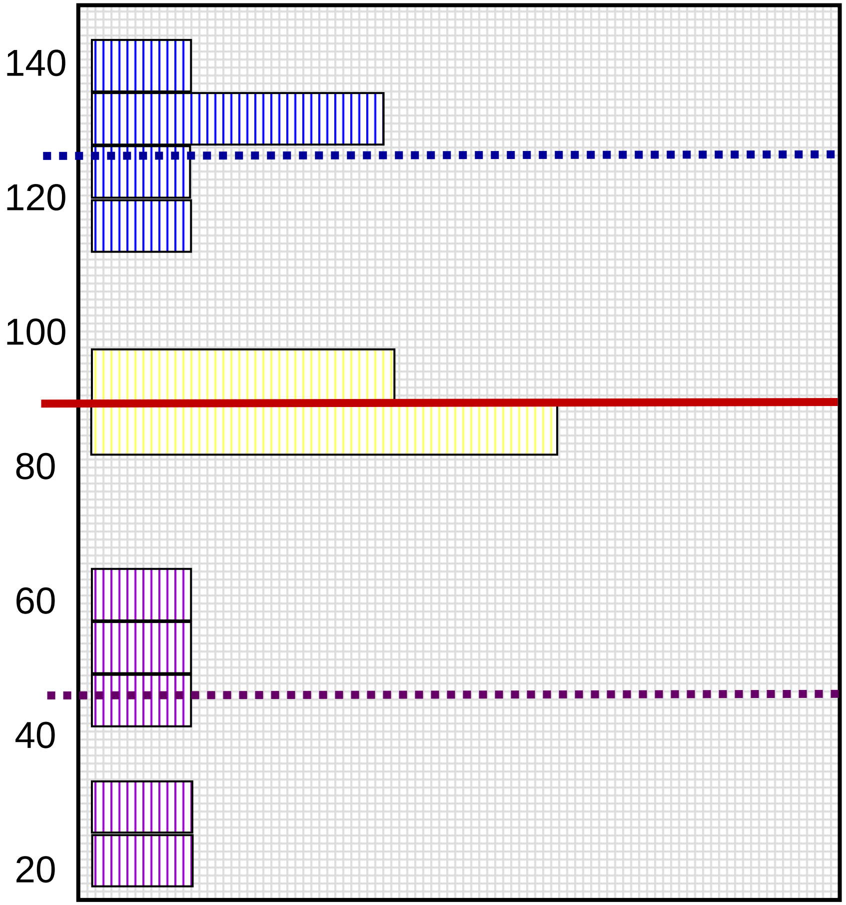
Full Sample Distribution



5 Occasions (t); 3 People (i)



Empty Means (+WP) Two-Level Model



Start off with Mean of y_{ti} as "best guess" for any value:

= Grand Mean

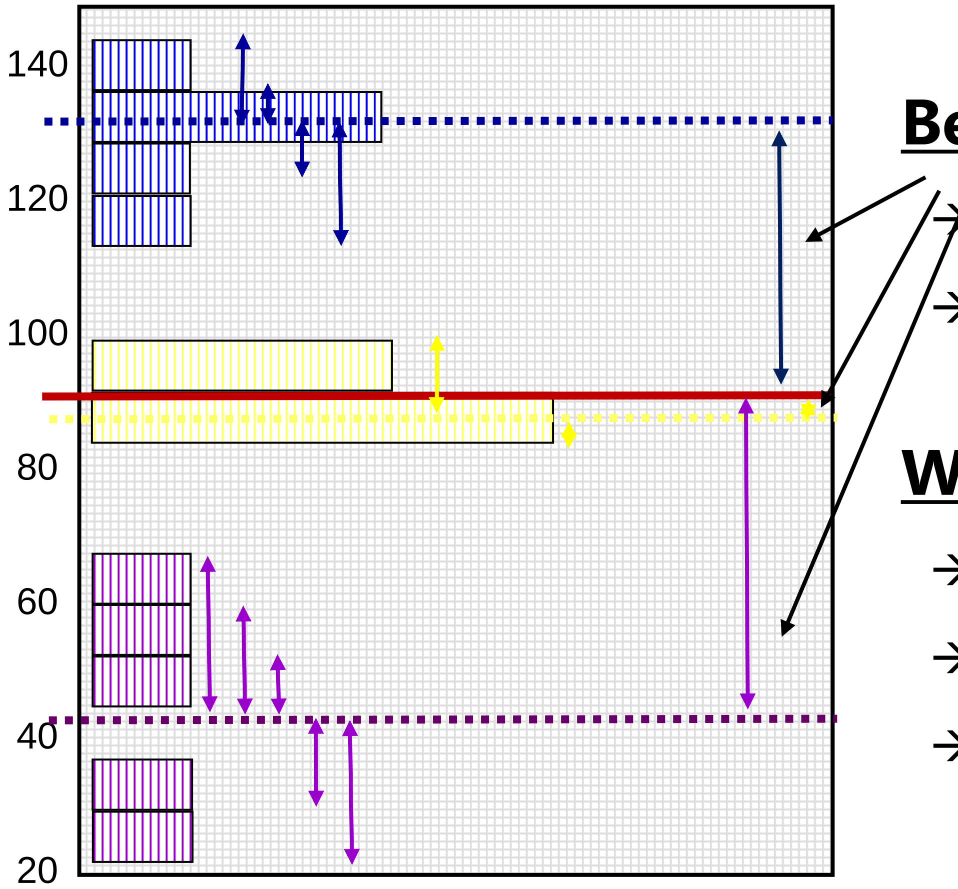
= **Fixed Intercept**

Can make better guess by taking advantage of repeated observations:

= Person Mean

→ Random Intercept

Empty Means (+WP) Two-Level Model



Variance of y_{ti} → 2 sources:

Between-Person (BP) Variance:

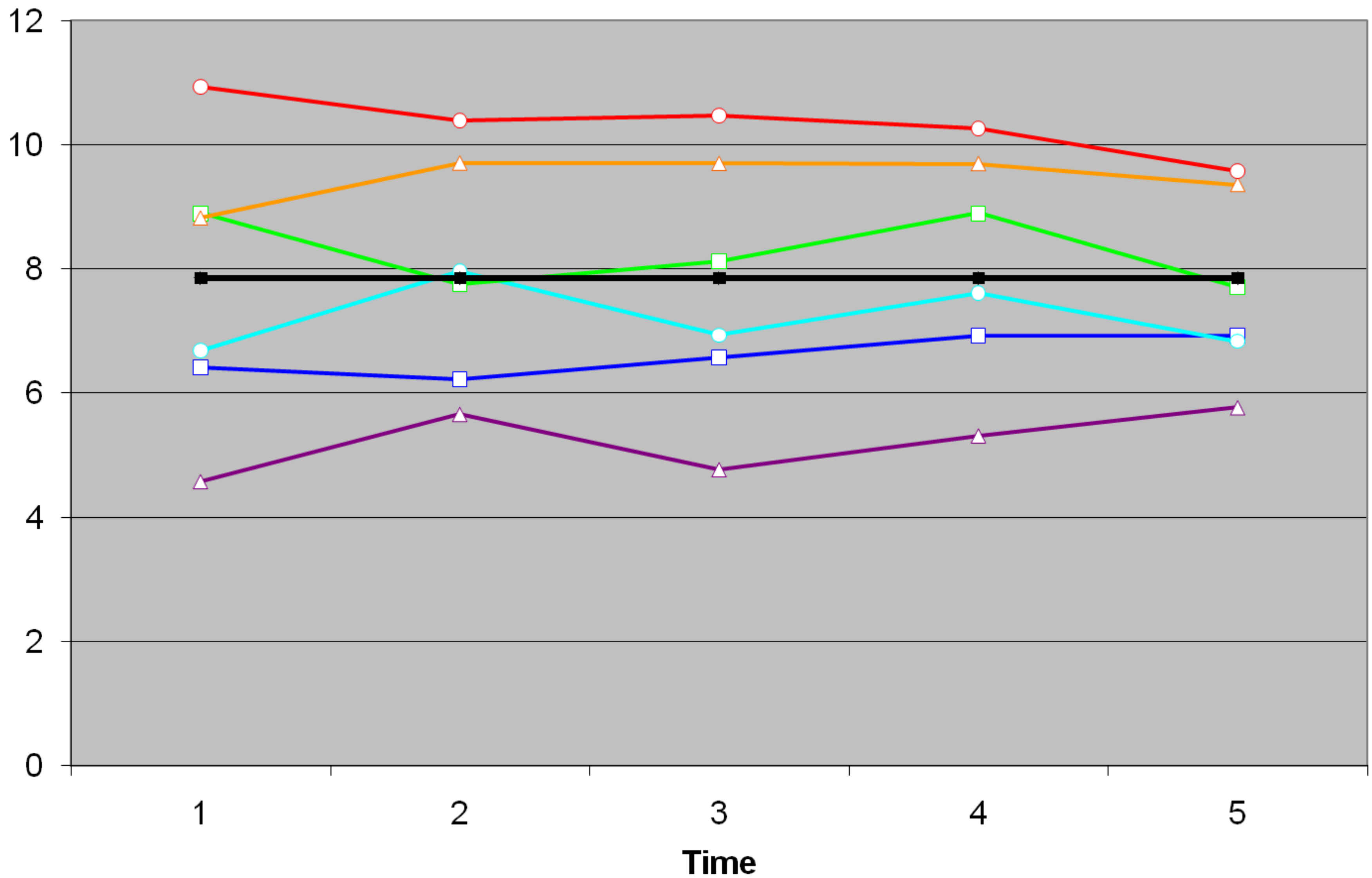
- → Differences from **GRAND** mean
- → **INTER**-Individual Differences

Within-Person (WP) Variance:

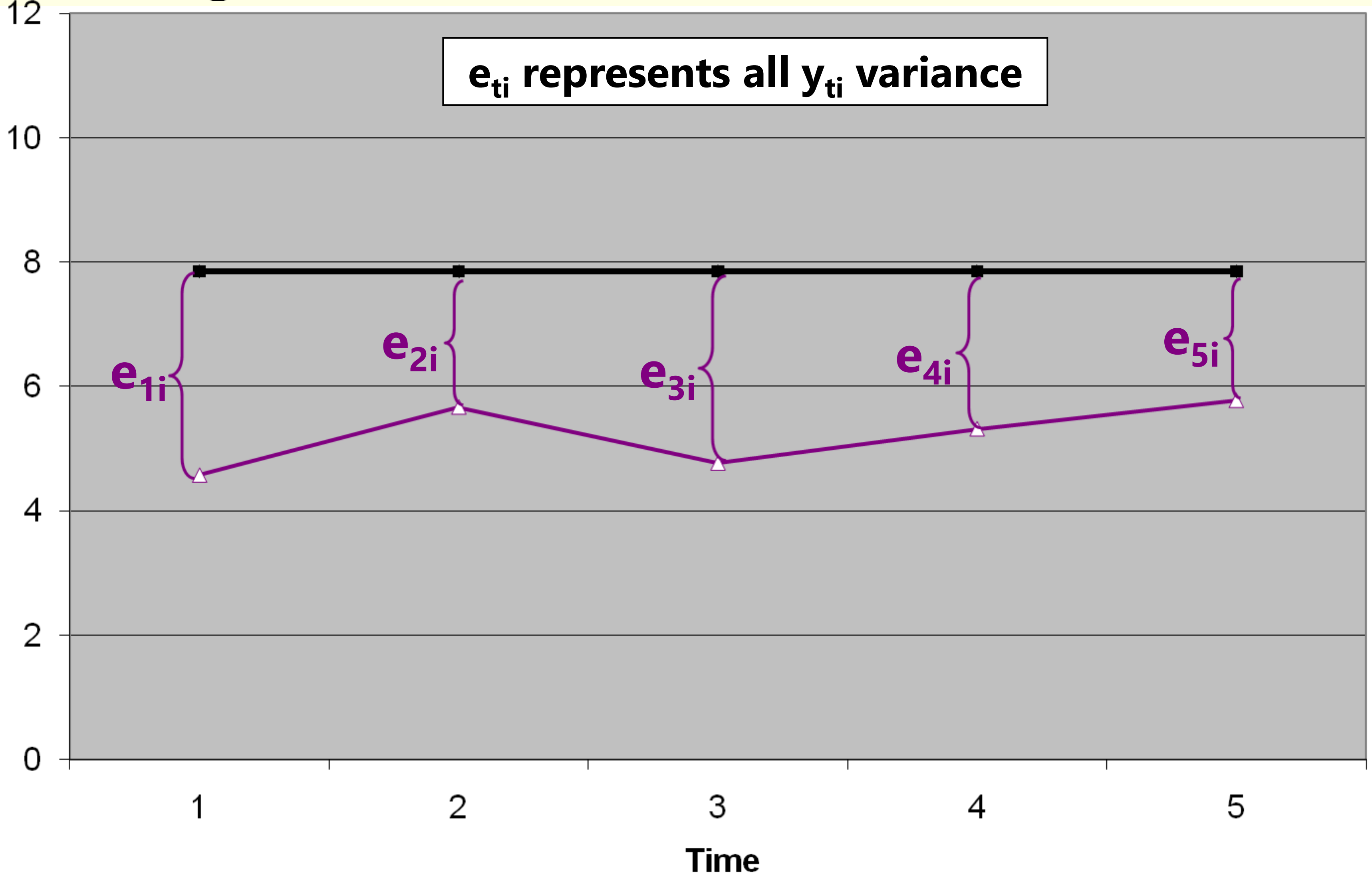
- → Differences from **OWN** mean
- → **INTRA**-Individual Differences
- → This part is only observable through longitudinal data.

Now we have 2 piles of variance in y_{ti} to predict.

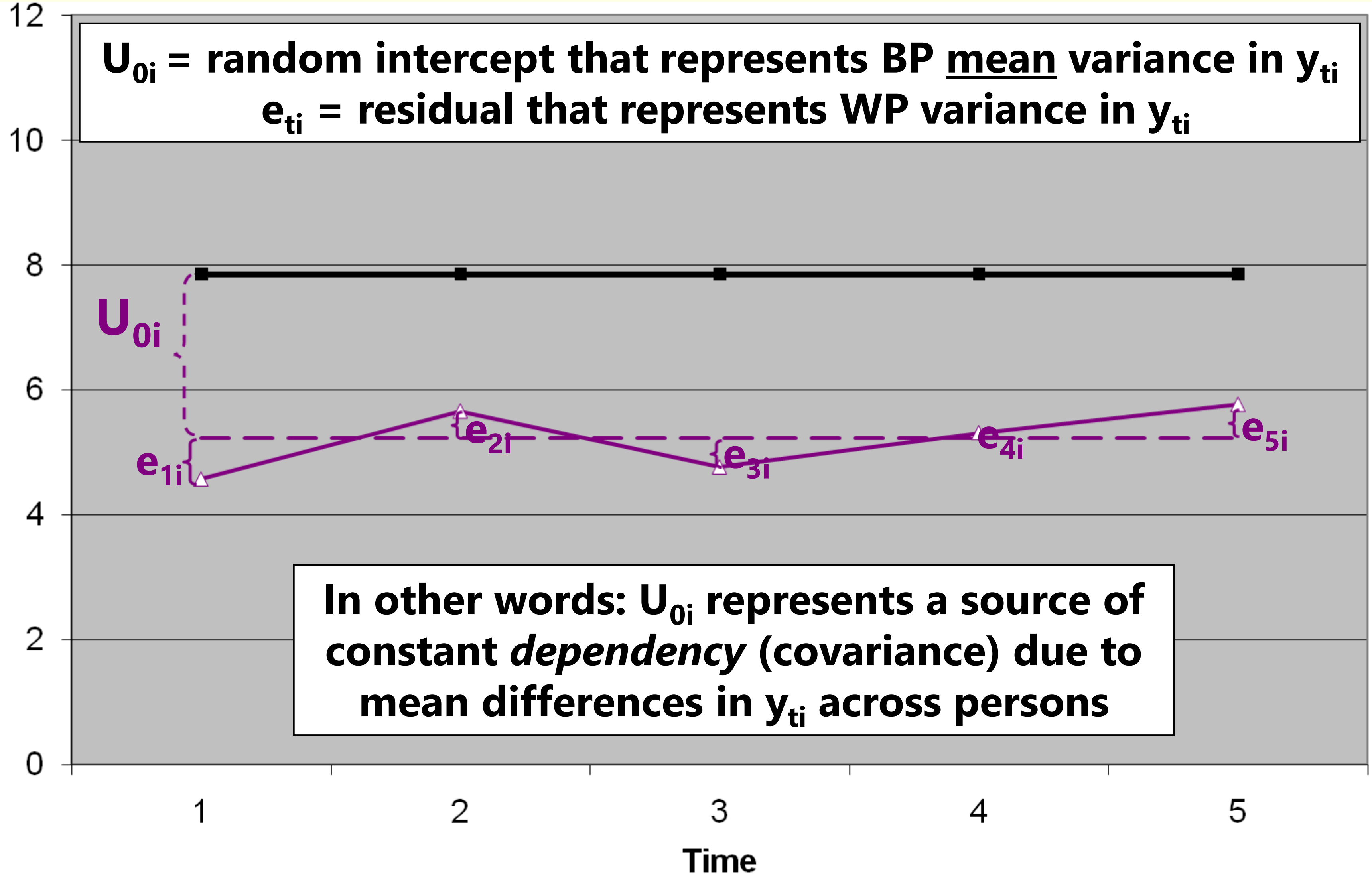
Hypothetical Longitudinal Data



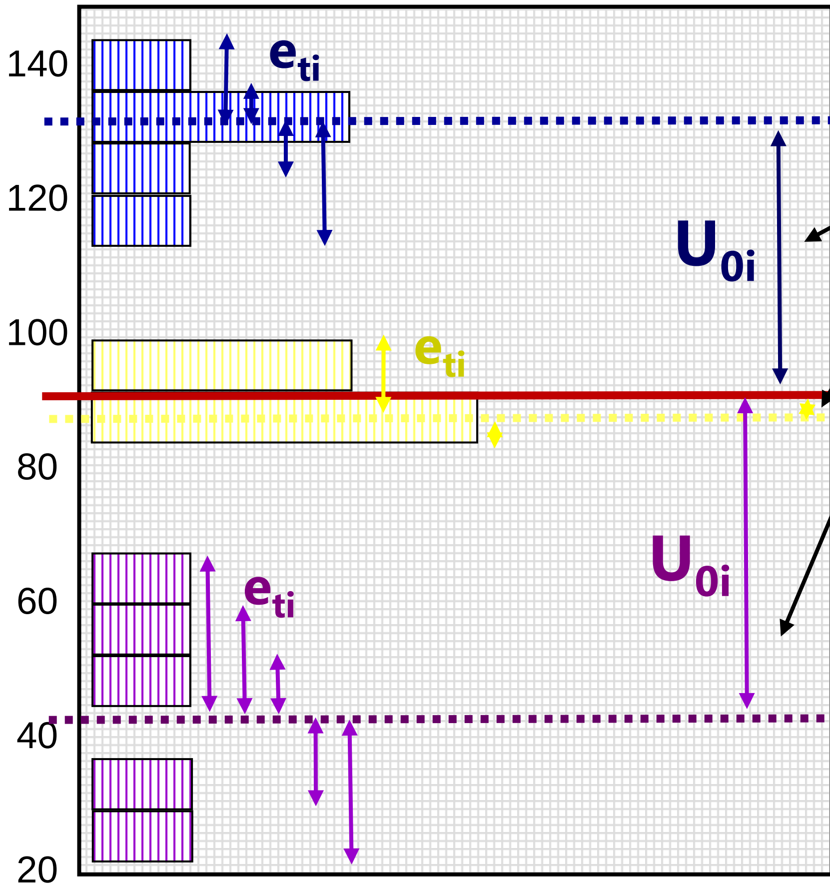
Only One Kind of "Error" in a Single-Level Model for the Variance



Two Distinct Kinds of "Error" in a Two-Level Model for the Variance



Empty Means, Two-Level Model



y_{ti} variance → 2 sources:

Level-2 Random Intercept Variance (of U_{0i}, as $\tau^2_{U_0}$):

→ **Between**-Person variance in means

→ **INTER**-Individual differences from **GRAND** mean to be explained by time-invariant predictors

Level-1 Residual Variance (of e_{ti}, as σ^2_e):

→ **Within**-Person variance

→ **INTRA**-Individual differences from **OWN** mean to be explained by time-varying predictors

Two-Level Model Using Multilevel Notation: Empty Means, Random Intercept Model

GLM Empty Model:

- $y_i = \beta_0 + e_i$

MLM Empty Model:

- Level 1:

$$y_{ti} = \beta_{0i} + e_{ti}$$

- Level 2:

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

3 Parameters:

Model for the Means (1):

- Fixed Intercept γ_{00}

Model for the Variance (2):

- Level-1 **WP** Variance of e_{ti} → σ_e^2
- Level-2 **BP** Variance of U_{0i} → $\tau_{U_0}^2$

Residual = time-specific deviation from individual's predicted outcome

Fixed Intercept = mean of person means (because no predictors yet)

Random Intercept = individual-specific deviation from predicted intercept

Composite equation:

$$y_{ti} = (\gamma_{00} + U_{0i}) + e_{ti}$$

Intraclass Correlation (ICC)

Intraclass Correlation (ICC; also known as "ICC1"):

$$\text{ICC} = \frac{\text{BP}}{\text{BP} + \text{WP}} = \frac{\text{Intercept Var.}}{\text{Intercept Var.} + \text{Residual Var.}} = \frac{\boldsymbol{\tau^2_{U_0}}}{\boldsymbol{\tau^2_{U_0}} + \boldsymbol{\sigma^2_e}}$$

ICC =

$$r(y_{1i}, y_{2i}) = \frac{Cov(y_{1i}, y_{2i})}{\sqrt{Var(y_{1i})}\sqrt{Var(y_{2i})}}$$

$$\begin{bmatrix} \sigma^2_e + \tau^2_{u_0} & \tau^2_{u_0} & \tau^2_{u_0} \\ \tau^2_{u_0} & \sigma^2_e + \tau^2_{u_0} & \tau^2_{u_0} \\ \tau^2_{u_0} & \tau^2_{u_0} & \sigma^2_e + \tau^2_{u_0} \end{bmatrix} \begin{bmatrix} 1 & \text{ICC} & \text{ICC} \\ \text{ICC} & 1 & \text{ICC} \\ \text{ICC} & \text{ICC} & 1 \end{bmatrix}$$

- ICC = Proportion of total variance that is between persons
- ICC = Correlation of occasions from same person (in RCORR)
- ICC is a standardized way of expressing how much we need to worry about *dependency due to person mean differences* **(i.e., ICC is an effect size for <u>constant</u> person dependency)**

Augmenting the Empty Means, Random Intercept Model with **Time**

- 2 questions about the possible effects of **"time"** (e.g., time in study in WP change; time of day or day of week in WP fluctuation):

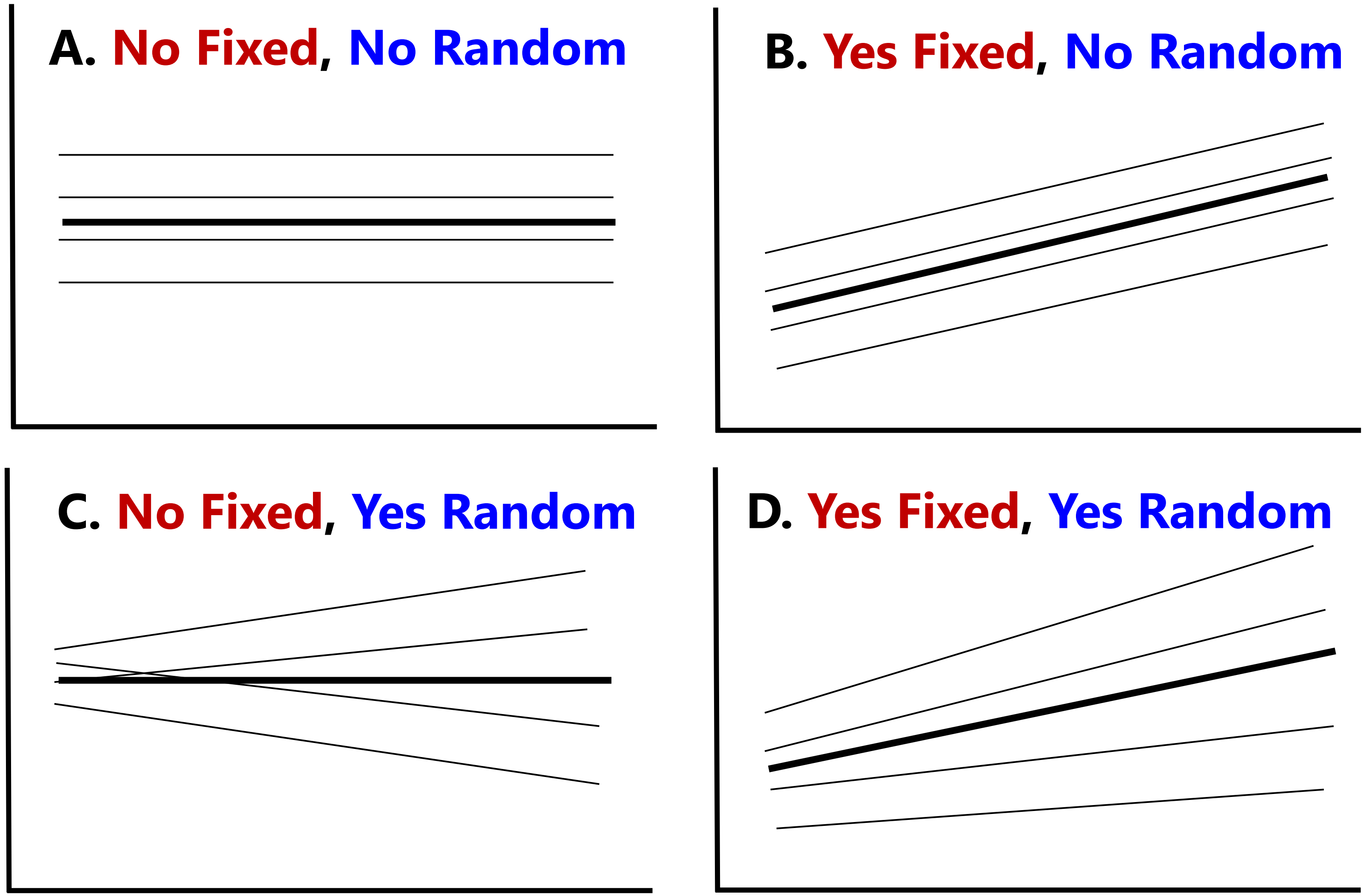
1. **Is there an effect of time on average?**
   - Is the line connecting the sample means not flat?
   - If so, you need **FIXED** effect(s) of time

2. **Does the average effect of time vary across individuals?**
   - Does each individual need their *own* version of that line?
   - If so, you need **RANDOM** effect(s) of time

- Let's look at examples using **linear time** effects to start…

Fixed and Random Effects of **Time**

(Note: The intercept is random in every figure)



B. Fixed Linear Time, Random Intercept Model
(4 parameters: effect of time is FIXED only)

Multilevel Model

Residual = time-specific deviation from individual's predicted outcome → estimated variance of σ_e^2

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

Fixed Intercept = predicted mean outcome at time 0

Fixed Linear Time Slope = predicted mean rate of change per unit time

Level 2: $$\beta_{0i} = \gamma_{00} + U_{0i} \qquad \beta_{1i} = \gamma_{10}$$

Random Intercept = individual-specific deviation from fixed intercept → estimated variance of $\tau_{U_0}^2$

Composite Model

$$y_{ti} = \underbrace{(\gamma_{00} + U_{0i})}_{\beta_{0i}} + \underbrace{(\gamma_{10})}_{\beta_{1i}}(Time_{ti}) + e_{ti}$$

Because the effect of time is **fixed**, everyone is predicted to change at the same rate

C or D: Random Linear Time Model (6 parms)

Multilevel Model

Residual = time-specific deviation from individual's predicted outcome → estimated variance of σ_e^2

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

Fixed Intercept = predicted mean outcome at time 0

Fixed Linear Time Slope = predicted mean rate of change per unit time

Level 2: $\beta_{0i} = \gamma_{00} + U_{0i}$ $\quad \beta_{1i} = \gamma_{10} + U_{1i}$

Random Intercept = individual-specific deviation from fixed intercept at time 0 → estimated variance of $\tau_{U_0}^2$

Random Linear Time Slope = individual-specific deviation from fixed linear time slope → estimated variance of $\tau_{U_1}^2$

Also has an estimated **covariance** of random intercepts and slopes of $\tau_{U_{01}}$

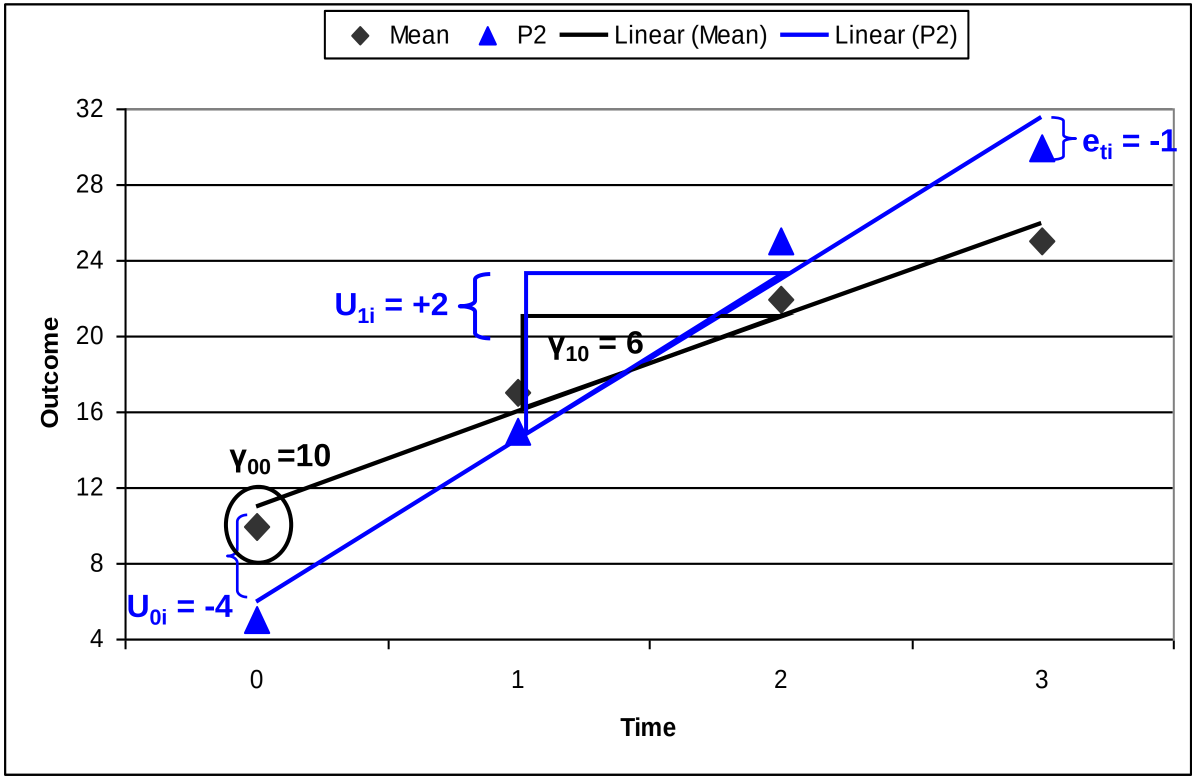
Composite Model

$$y_{ti} = \underbrace{(\gamma_{00} + U_{0i})}_{\beta_{0i}} + \underbrace{(\gamma_{10} + U_{1i})}_{\beta_{1i}}(Time_{ti}) + e_{ti}$$

Random Linear Time Model

$$y_{ti} = (\gamma_{00} + U_{0i}) + (\gamma_{10} + U_{1i})(Time_{ti}) + e_{ti}$$

- γ_{00}: Fixed Intercept
- U_{0i}: Random Intercept Deviation
- γ_{10}: Fixed Slope
- U_{1i}: Random Slope Deviation
- e_{ti}: error for person *i* at time *t*



6 Parameters:

2 Fixed Effects:

γ_{00} Intercept, γ_{10} Slope

U_{0i} Random Intercept Variance = $\tau^2_{U_0}$

U_{1i} Random Slope Variance = $\tau^2_{U_1}$

Random Int-Slope Covariance = $\tau_{U_{01}}$

e_{ti} Residual Variance = σ^2_e

Random Linear Time Models Imply:

- **People differ from each other systematically in TWO ways**—in intercept (U_{0i}) and time slope (U_{1i}), which implies **TWO kinds of BP variance**, which translates to **TWO sources of person dependency** (covariance or correlation in the outcomes from the same person)
- If so, after controlling for both BP intercept and slope differences (by estimating the $\tau^2_{U_0}$ and $\tau^2_{U_1}$ variances in the **G** matrix), the **e_{ti} residuals** (whose variance and covariance are estimated in the **R** matrix) should be **uncorrelated with homogeneous variance across time**, as shown (or else a different **R** matrix is needed):

Level-2 **G** matrix: RANDOM TYPE=UN

$$\begin{bmatrix} \tau^2_{U_0} & \tau_{U_{10}} \\ \tau_{U_{01}} & \tau^2_{U_1} \end{bmatrix}$$

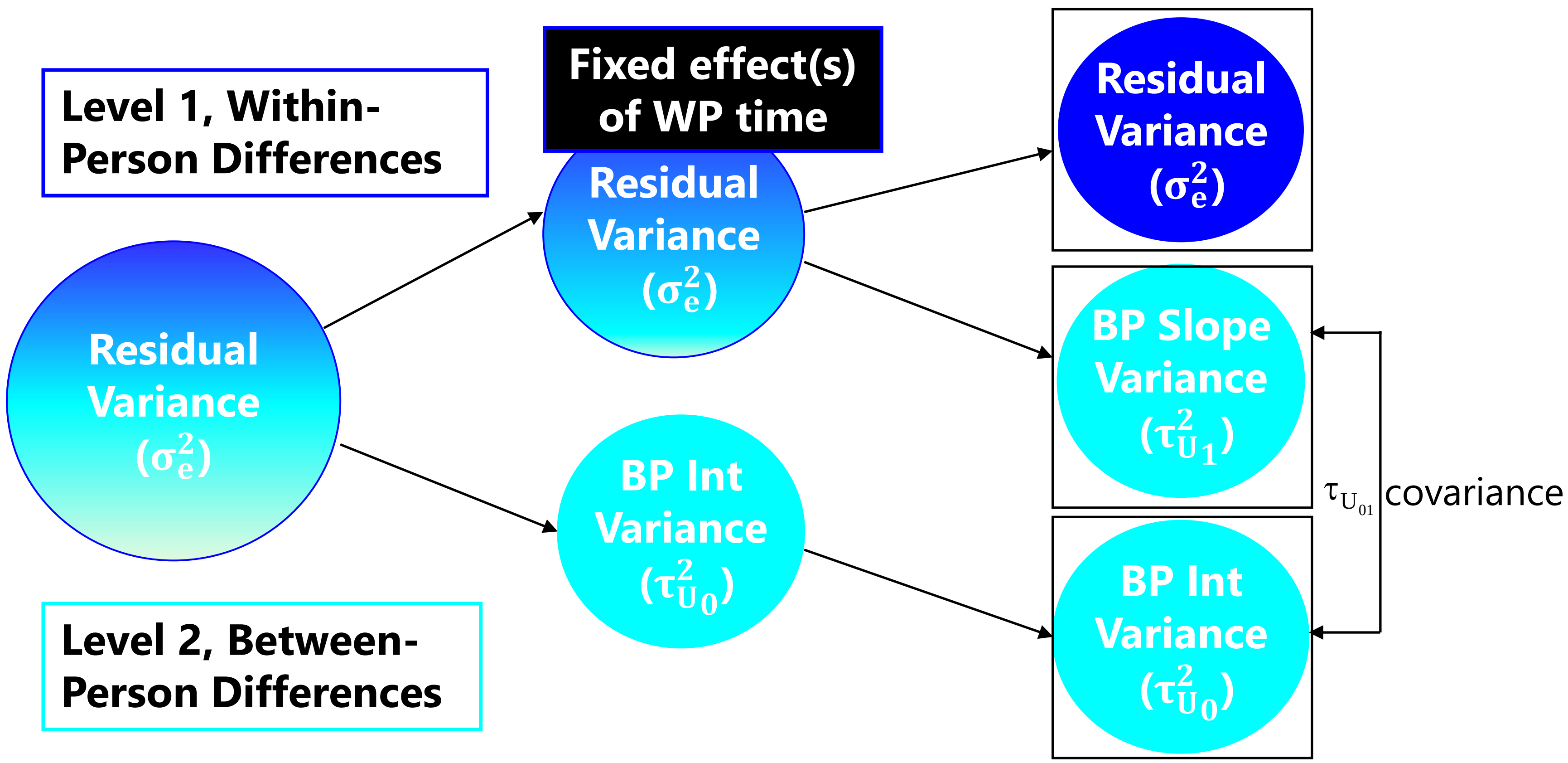
Level-1 **R** matrix: REPEATED TYPE=VC

$$\begin{bmatrix} \sigma^2_e & 0 & 0 & 0 \\ 0 & \sigma^2_e & 0 & 0 \\ 0 & 0 & \sigma^2_e & 0 \\ 0 & 0 & 0 & \sigma^2_e \end{bmatrix}$$

G and **R** combine to create a total **V** matrix whose per-person structure depends on the **specific time occasions** for each person in **Z** (flexible for unbalanced time)

Summary: "Handling" Person Dependency

- The process of fitting "unconditional models for time" (fixed and random effects) can be depicted as follows:



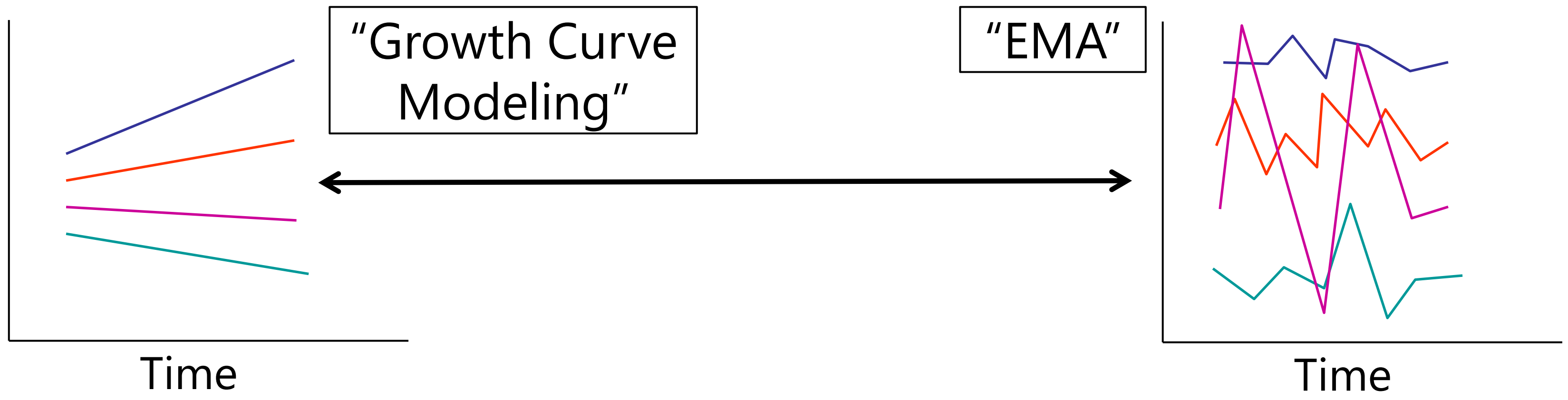
Families of Nonlinear Change

- Polynomial functions (e.g., $time^2$, $time^3$)
  - Best suited for time slopes that should change directions (in which time is treated as continuous)
- Piecewise (linear spline) functions
  - Best suited for distinct phases of time (known "knot" points)
  - Otherwise, location of "latent" knots can be model parameters
- Linear effect of log(time) → exponential-ish
  - Good for time slopes that should level off (hit upper or lower asymptote)
  - Adding quadratic log(time) adjusts how fast the time slope levels off
- Latent basis → single slope with estimated nonlinearity
  - In SEM software, for random time slope factor: fix first loading to 0, last loading to 1, and estimate the other loadings to capture proportion of change by each occasion
- Truly nonlinear models (e.g., logistic, exponential)
  - Harder to estimate, particularly for random effects variances

Summary: Unconditional Models for Time

Pure WP Change

Pure WP Fluctuation



Role of "Time" in the Model for the Means:

- WP Change → describe pattern of ***average*** change (e.g., growth curves)
- WP Fluctuation → describe ***average*** time-specific trends that may not have been expected (e.g., reactivity, day of the week, circadian/schedule effects)

Role of "Time" in the Model for the Variance:

- WP Change → describe ***individual differences*** in change (random effects)
  → this allows variances and covariances to differ over time
- WP Fluctuation → mostly describe pattern(s) of covariance over time (may need random effects of time for differing variances)

Summary: Unconditional Models for Time

- 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 two-level longitudinal model:**

Level 1 (one source of) <u>Within-Person Variation</u>: 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) <u>Between-Person Variation</u>: 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**

Multiple BP time slope variances are possible…

We can add predictors to reduce each pile of variance!

Modeling Time-Invariant Predictors

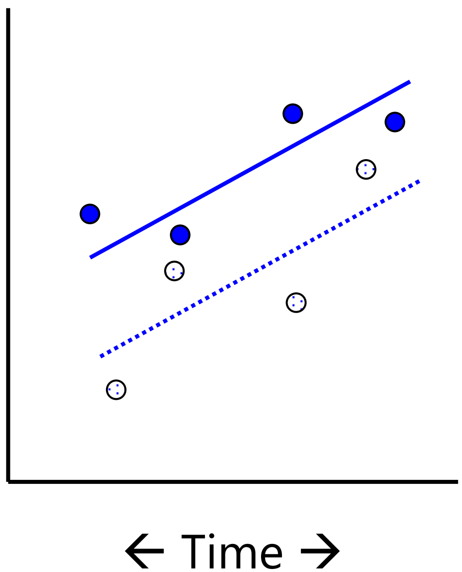
- Which independent variables can be time-invariant predictors?
  - Aka, "**person-level**" or "**level-2**" or predictors ($\boldsymbol{x_i}$) in two-level models
  - Includes substantive predictors, controls, and predictors of missingness
  - Includes anything that either **does not change across time**, or that might change across time but that **you've only measured once** (you may need to argue why this is conceptually ok or limit conclusions accordingly)
  - Also includes **BP variance in time or time-varying predictors** (stay tuned)

- All predictors should be **centered** so that 0 values are meaningful:
  - This is needed to create a meaningful fixed/random intercept, and/or meaningful fixed main effects of predictors also included in interactions
    - e.g., if fixed effects of X, Z, and X*Z, the main effect of X is specifically for Z=0
  - **Quantitative** predictors can be **centered at any constant**, such as the sample mean (common, and useful if it has an unfamiliar scale) or any meaningful reference (better for translating across studies)
  - **Categorical** predictors can have their **dummy-code contrasts** created for you as "factor" variables (e.g., SAS CLASS, SPSS BY, STATA i.), but not in Mplus; I do not like ± 1 coding for group differences (because then 0 = ???)
    - I find indicator or sequential dummy-coding variants most useful

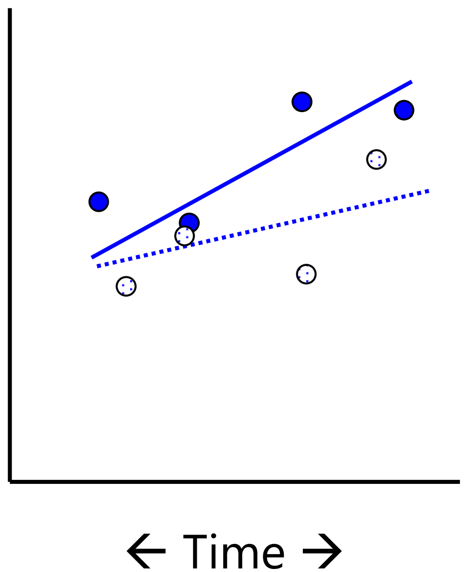
The Role of Time-Invariant Predictors in the **Model for the Means**

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

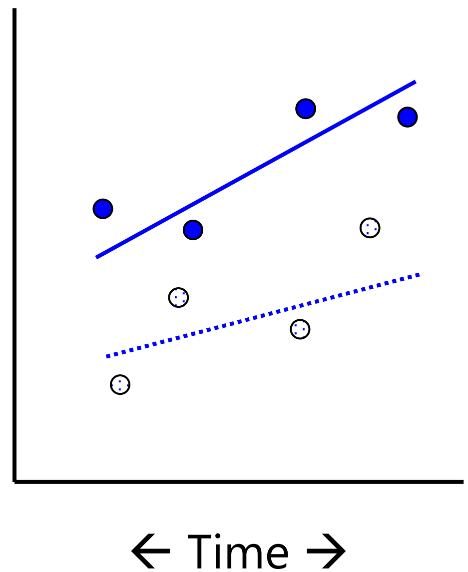
Main effect of x_i, no interaction with time



Interaction with time, main effect of x_i?



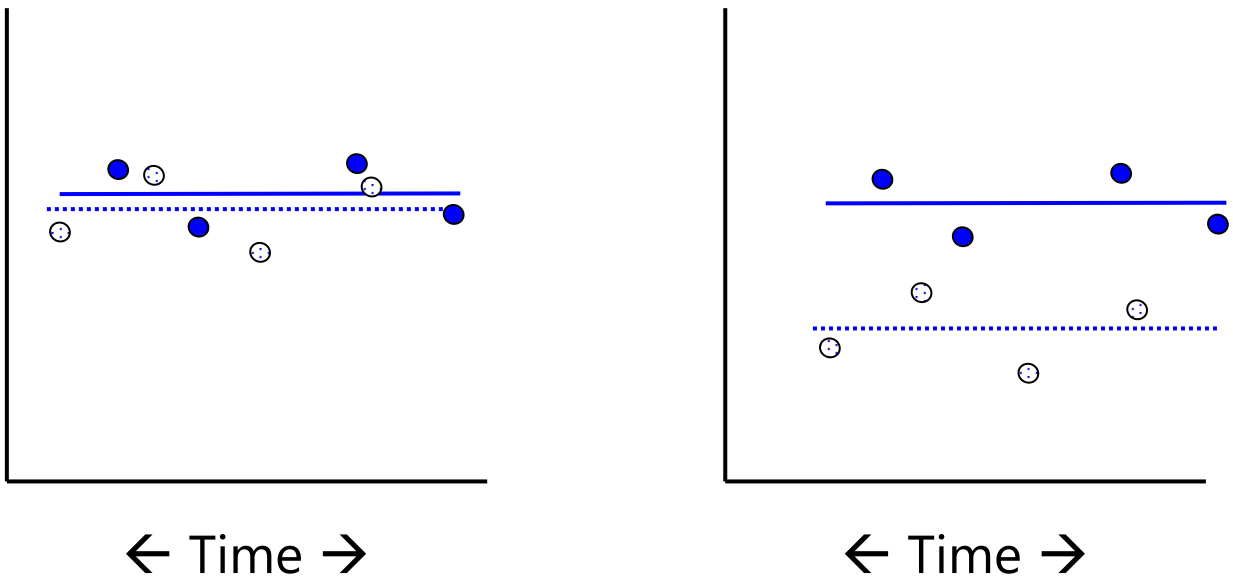
Main effect of x_i, and interaction with time



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_i | Main effect of x_i

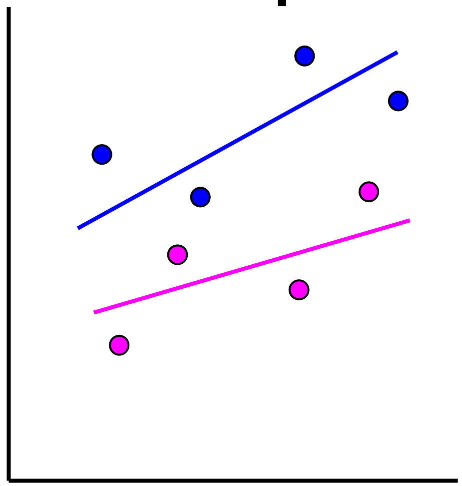


The Role of Time-Invariant Predictors in the **Model for the Variance**

- Beyond fixed effects in the model for the means, time-invariant predictors can be used to allow **heterogeneity of variance** at their level or below in "**location–scale models**"
- 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, in Mplus v 8+ using "logV")
  - Also described with examples in Hoffman & Walters (2022)

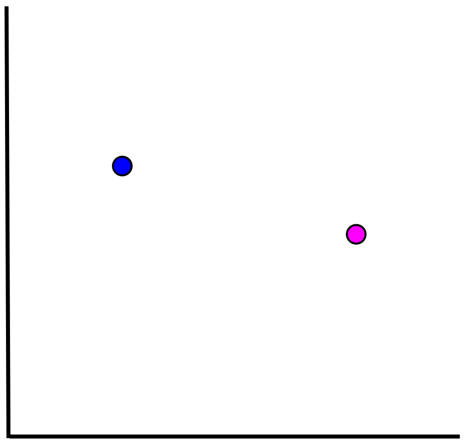
Why Level-2 Predictors Cannot* Have Random Effects in Two-Level Models

Random Slopes for Time



Time
(or Any Level-1 Predictor)

Random Slopes for Group?

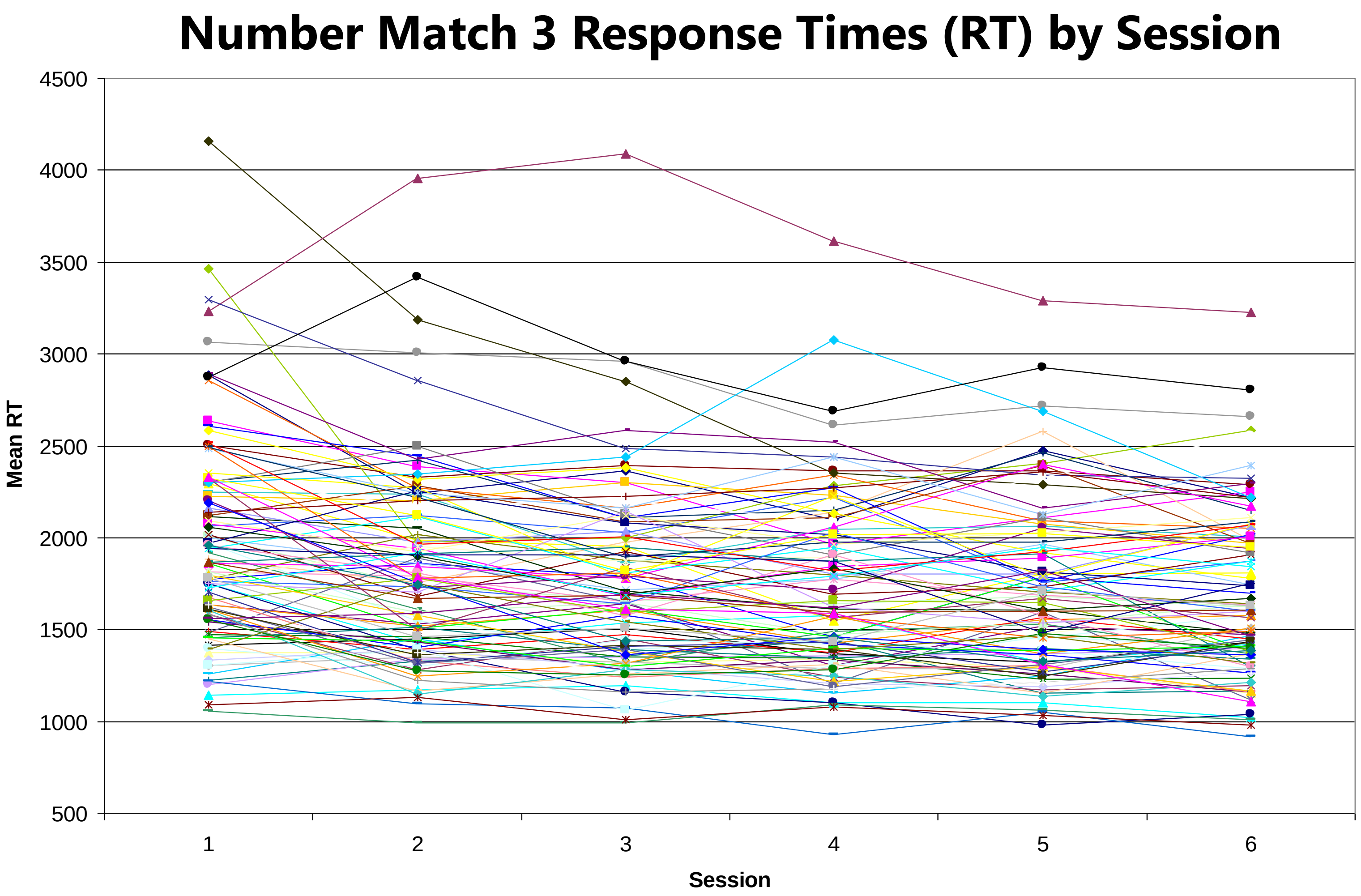


Group
(or any Level-2 Predictor)

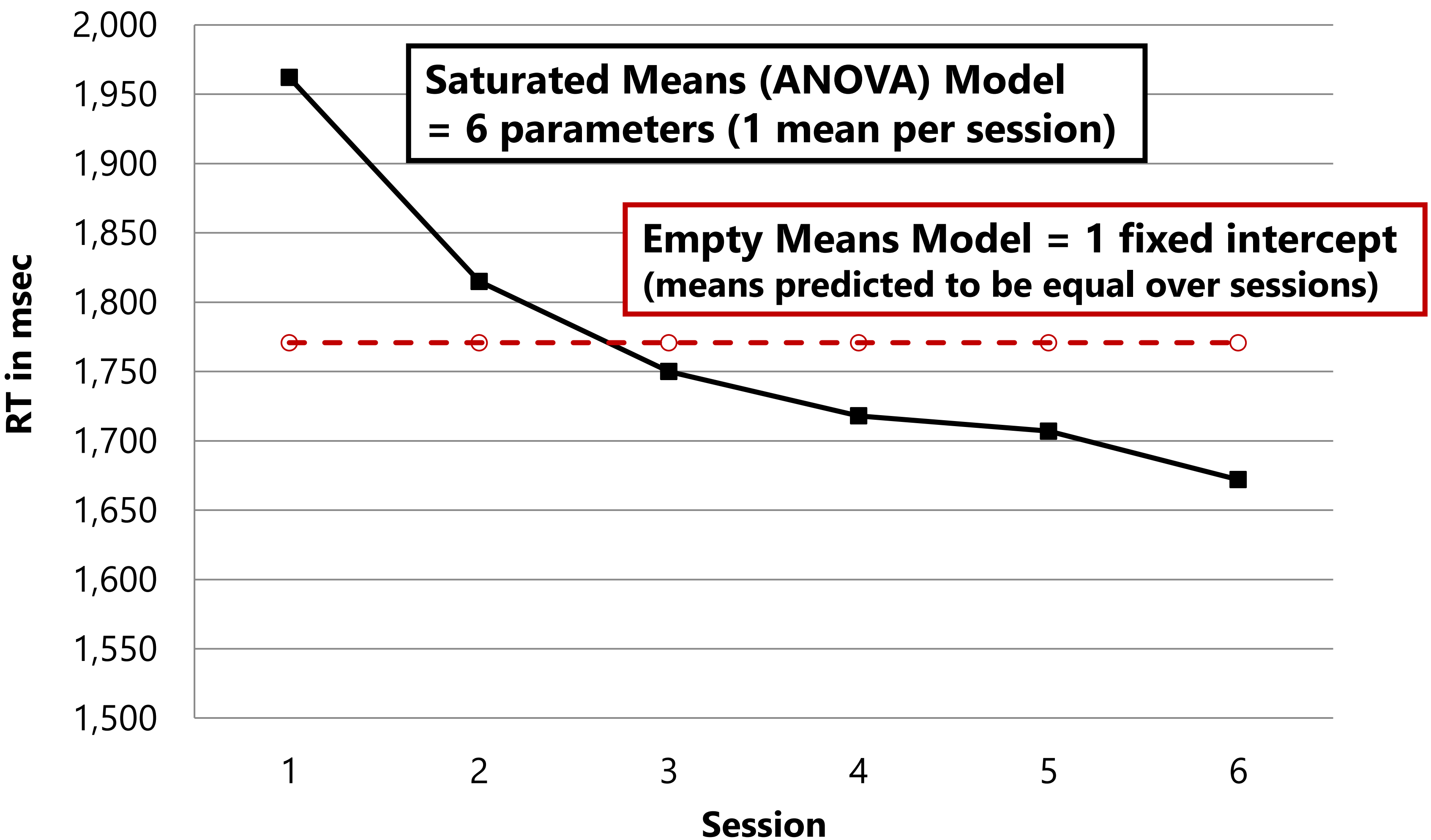
You cannot make a line out of a dot, so level-2 effects cannot vary randomly over persons.

* *Level-2 predictors can be included as predictors of heterogeneity of variance, which technically is a random slope of sorts (but interpretation is different)*

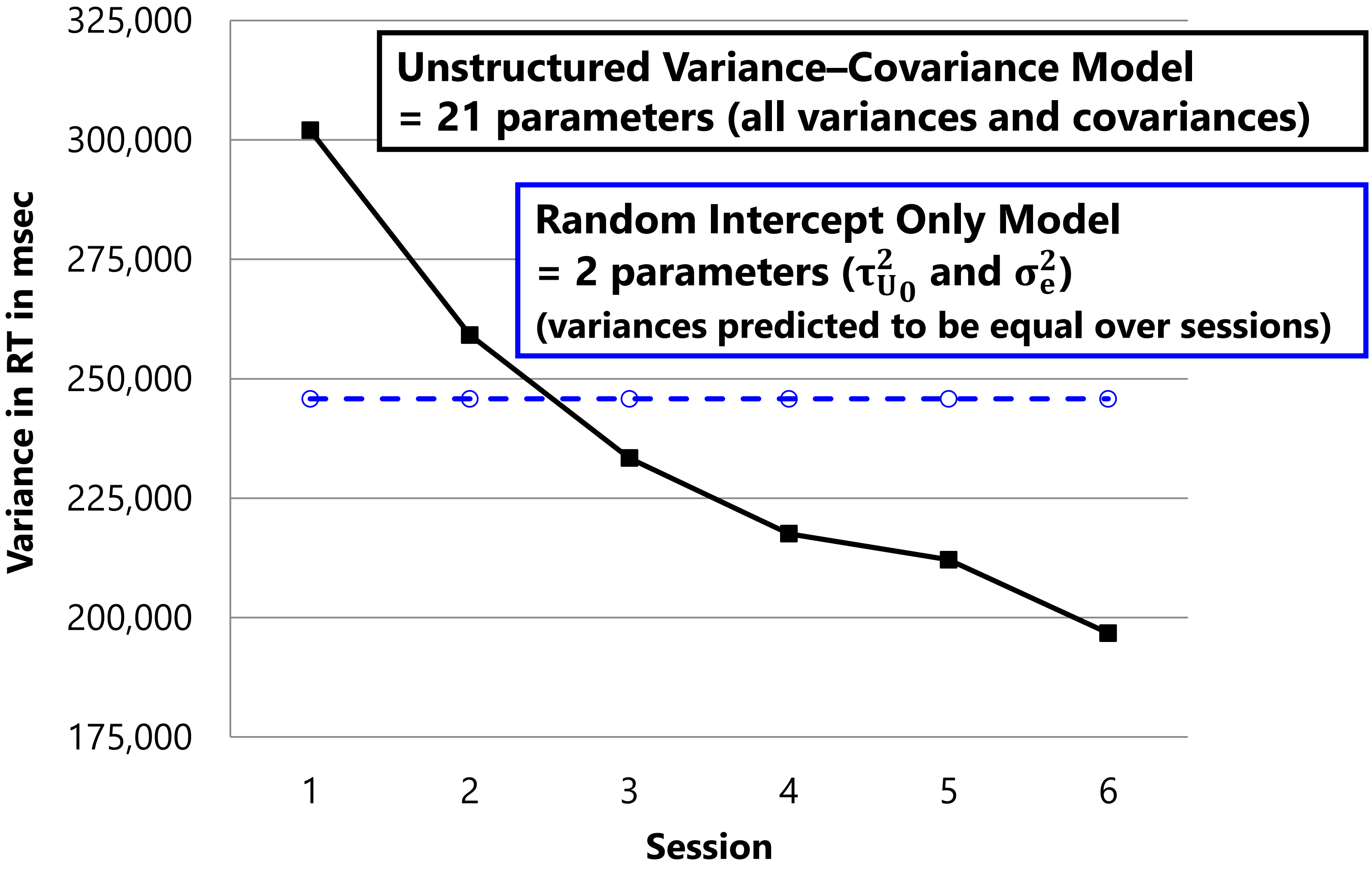
Example: Individual Trajectories 101 older adults, 6 occasions within 2 weeks



Example Mean RT by Session: Baseline Models for the Means



Example Variance in RT by Session: Baseline Models for the Variance



Random Quadratic Time Unconditional Model

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

Level 2 Equations (one per β):

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

- β_{0i}: **Intercept** for person *i*
- γ_{00}: Fixed (mean) **Intercept**
- U_{0i}: Random (Deviation) **Intercept**

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

- β_{1i}: **Linear Time Slope** for person *i*
- γ_{10}: Fixed (mean) **Linear Slope**
- U_{1i}: Random (Deviation) **Linear Slope**

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

- β_{2i}: **Quadratic Time Slope** for person *i*
- γ_{20}: Fixed (mean) **Quad Slope**
- U_{2i}: Random (Deviation) **Quad Slope**

Time = session − 1
REML estimation using stacked data (univ MLM)
U_i covariances also estimated

Fixed Effect Subscripts:
1st = which level-1 term
2nd = which level-2 term

# of Possible Time-Related Slopes by # of Occasions (*n*):

Fixed time slopes = $n - 1$
Random time slopes = $n - 2$

Need $n = 4$ occasions to fit random quadratic time model

Adding Reasoning (0=22) as a Time-Invariant Predictor: Is RT Improvement Predicted by Fluid Intelligence?

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

Level 2 Equations (one per β):

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

- β_{0i}: **Intercept** for person *i*
- γ_{00}: **Fixed Intercept when Time=0 and Reas=22**
- γ_{01}: **Δ in Intercept per unit Δ in Reas**
- U_{0i}: **Random (Deviation) Intercept after controlling for Reas**

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

- β_{1i}: **Linear Slope** for person *i*
- γ_{10}: **Fixed Linear Time Slope when Time=0 and Reas=22**
- γ_{11}: **Δ in Linear Time Slope per unit Δ in Reas (=Reas*time)**
- U_{1i}: **Random (Deviation) Linear Time Slope after controlling for Reas**

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

- β_{2i}: **Quad Slope** for person *i*
- γ_{20}: **Fixed Quad Time Slope when Reas=22**
- γ_{21}: **Δ in Quad Time Slope per unit Δ in Reas (=Reas*time^2)**
- U_{2i}: **Random (Deviation) Quad Time Slope after controlling for Reas**

Reasoning (0=22) as a Time-Invariant Predictor: Is RT Improvement Predicted by Fluid Intelligence?

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

Level 2 Equations (one per β):

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

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

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

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

Each fixed slope of reasoning will predict the random U_i variance in its level-2 equation if present, or e_{ti} residual variance otherwise. That's why random slopes should be tested **before** adding cross-level interactions!

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

Reasoning (0=22) as a Time-Invariant Predictor: Is RT Improvement Predicted by Fluid Intelligence?

$$RT_{ti} = (1966 + -27{*}Reas_i + U_{0i}) + (-120 + -3.6{*}Reas_i + U_{1i})Time_{ti} + (13 + 1.2{*}Reas_i + U_{2i})Time_{ti}^2 + e_{ti}$$

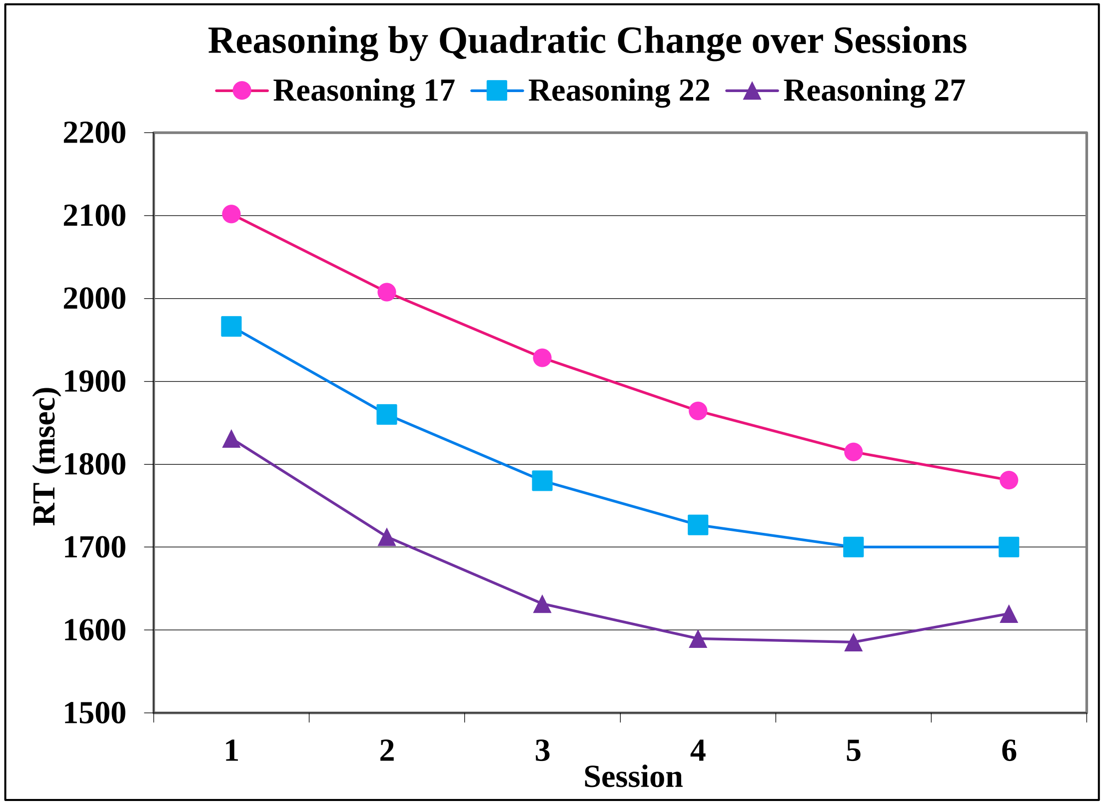
BP Pseudo-R^2 Values:

Intercept U_{0i} = .049

Linear Time U_{1i} = −.006

Quadratic Time U_{2i} = .024

WP Residual e_{ti} = **0**



People with better reasoning:

- started out faster/lower *(intercept at session 1)*,
- improved more initially *(linear slope at session 1)*,
- and had a greater rate of deceleration with practice *(quadratic slope*2!)*

Example: Syntax by Univariate MLM Program (Stacked Data)

SAS:

```
PROC MIXED DATA=work.Example2 COVTEST METHOD=REML;
     CLASS ID;
     MODEL RT = time timesq reas time*reas timesq*reas / SOLUTION DDFM=Satterthwaite;
     RANDOM INTERCEPT time timesq / GCORR TYPE=UN SUBJECT=ID;
RUN;
```

R (lmer from lme4 package)—using lmertest package, which does provide correct denominator DF:

```
model2 = lmer(data=Example2, REML=TRUE,
         formula=RT~1+time+timesq+reas+reas
                    +time:reas+timesq:reas+(1+time+timesq|ID))
summary(model2, ddf="Satterthwaite")
```

STATA:

```
mixed RT time timesq reas time#reas timesq#reas, || ID: time timesq, ///
      variance reml covariance(un) dfmethod(satterthwaite) dftable(pvalue)
```

SPSS:

```
MIXED RT BY ID WITH time timesq reas
      /METHOD = REML
      /PRINT = SOLUTION TESTCOV
      /FIXED = time timesq reas time*reas timesq*reas
      /RANDOM = INTERCEPT time timesq | COVTYPE(UN) SUBJECT(ID).
```

Example: Mplus M-SEM Syntax

Just showing **MODEL** part, which would be preceded by **DATA**, **VARIABLE**, and **ANALYSIS** as usual (estimated using **long** data)

```
%WITHIN%
  RT;                     ! Level-1 residual variance
  Lin | RT ON time;       ! Create beta1i placeholder
  Qua | RT ON timesq;     ! Create beta2i placeholder

%BETWEEN%
  [RT Lin Qua];                      ! Intercepts
   RT Lin Qua;                       ! Level-2 random effect variances
   RT Lin Qua WITH RT Lin Qua;       ! Level-2 random effect covariances
   RT Lin Qua ON reas;               ! Fixed effects of reasoning
```

- Note: R's lavaan package does have M-SEM capability, but it is much more limited than M-SEM in Mplus:
  - Listwise deletion for any rows (occasions) with missing values
  - No random slopes!

Sources of Explained Variance by Person-Level-2 Time-Invariant Predictors

- **Fixed effects of level-2 predictors *by themselves*:**
  - Level-2 (BP) main effects reduce level-2 random intercept variance
  - Level-2 (BP) interactions also reduce level-2 random intercept variance
- **Fixed effects of *cross-level interactions* (level-1* level-2):**
  - If a 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 pred1*time*, pred2*time*, and pred1*pred2*time* can each reduce the level-2 random linear time slope variance
  - If the 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 $time^2$ does not have a level-2 random slope, then pred1*$time^2$, pred2*$time^2$, and pred1*pred2*$time^2$ will reduce the level-1 residual variance → Different quadratic slopes by pred1 and pred2 create better level-1 trajectories, thus reducing level-1 residual variance around the trajectories
    - But always **test the random slope first** before fitting cross-level interactions!

Variance Explained… Continued

- **Pseudo-R^2** is named that way for a reason… piles of variance can shift around, such that **it can actually become negative**
  - Sometimes is a sign of model mis-specification (but not always)
  - See Rights & Sterba (2019, 2020) for alternative marginal versions of R^2
    - Ensure positive R^2 values, but they don't quantify R^2 for slope variances (boo)

- **A simple alternative: Total R^2** (Singer & Willett, 2003)
  - Generate model-predicted $\hat{y}_{ti}$ from fixed effects only (NOT including random effects, so no cheating) and correlate it with observed y_{ti}
  - Then square that correlation → total R^2 (same as in GLM regression)
  - Total R^2 = total reduction in overall outcome variance across levels
  - Can be "unfair" in models with large unexplained sources of variance (i.e., for sampling dimensions you didn't have predictors for)

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

Wrapping Up

- Multilevel models are used to quantify and predict the sources of variance within different dimensions ("levels") of sampling
  - Longitudinal data → Level-1 occasions in Level-2 persons
  - Clustered data → Level-1 persons in Level-2 clusters
- MLMs differ from GLMs (regression, ANOVA) by including both fixed effects AND random effects
  - Fixed effects = everyone gets the same term in predicting the outcome
  - Random effects = everyone gets their own (intercept and time slope)
- Person characteristics (time-invariant level-2 predictors) can explain random intercept and slope variances across persons
  - Why do people start out or change differently?
  - Time-varying level-1 predictors are next!