Introduction to Concepts and Terminology in Longitudinal Modeling

- Topics:
  - **Concepts and terminology in longitudinal models**
  - Modeling person dependency
  - Fixed and random intercepts
  - Fixed and random time slopes
  - From multilevel models (MLMs) to single-level structural equation models (SEMs) to multilevel SEMs (M-SEMs)
  - Details

Sources of Longitudinal Relations

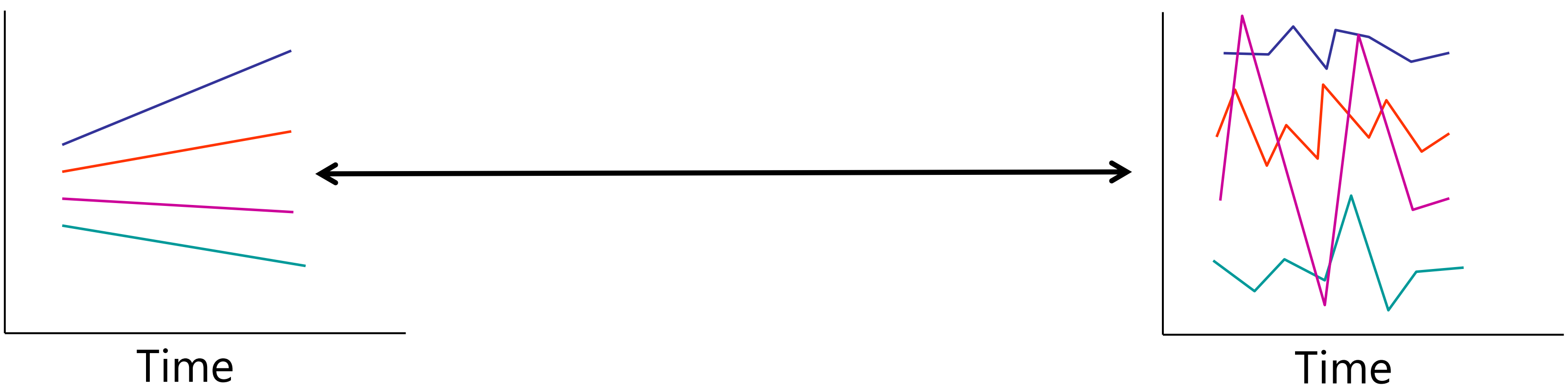
- **Between-Person (BP) Variation:**
  - "**INTER**-individual differences" from "**time-invariant**" **measures**
  - All longitudinal studies that begin as cross-sectional studies have this
- **Within-Person (WP) Variation:**
  - "**INTRA**-individual differences" from "**time-varying**" **measures**
  - 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 use "person" here, but "between" units can be anything that is measured repeatedly (e.g., schools, countries, companies…)

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



Why Do Longitudinal Research?

- To explore **within-person change** over time and its relations
  - On average (→ fixed effects): e.g., Does my new treatment result in greater (or faster) improvement than the standard approach?
  - BP differences (→ random effects): e.g., Do some people improve more (or more rapidly) over time than others? And if so, why?
  - Because cross-sectional age differences ≠ longitudinal age changes!
  - Btw, this is the purpose of "(latent) growth curve models"

- To explore **within-person fluctuation, "dynamics",** and their relations
  - On average (→ fixed effects): e.g., When you sleep less than usual, are you more impatient than usual the next day, too (or vice-versa, as "reciprocal" relations)?
  - BP differences (→ random effects): e.g., Are some people more affected by (relative) sleep deficits than others? And if so, why?
  - Btw, this is (often) the purpose of "multilevel models" or "multilevel SEM", as well as "cross-lag panel models" (or "auto-regressive cross-lag models)"

- To explore **within-person (in)stability** and its relations
  - e.g., Why are some people *moodier* than others?
  - e.g., Does inconsistency precede long-term age-related decline?
  - Btw, this is the purpose of "location–scale mixed effects models"

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 function of values on known predictor variables
    - Fixed effects are **estimated constants** that multiply predictors

- **Model for the Variance:**
  - *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" regression models***

Introduction to Concepts and Terminology in Longitudinal Modeling

- Topics:
  - Concepts and terminology in longitudinal models
  - **Modeling person dependency**
  - Fixed and random intercepts
  - Fixed and random time slopes
  - From multilevel models (MLMs) to single-level structural equation models (SEMs) to multilevel SEMs (M-SEMs)
  - Details

Modeling Longitudinal Dependency

- Outcomes from the same sampling unit (i.e., person) will have one or more sources of **dependency → correlated residuals**
  - If ignored, dependency in a longitudinal outcome will result in incorrect fixed effect standard errors and *p*-values (well-known problem)
  - If ignored, dependency in a longitudinal predictor variable will result in incorrect fixed effect estimates, too (relatively less well-known problem)
    - Because time-varying predictors have both BP and WP variation—stay tuned!

- The sources of residual correlation of occasions from same person can be captured by a model in three main ways:
  1. **Fixed effects:** Add Person ID as a predictor (via *N*-1 dummy codes)
  2. **(Multivariate) alternative covariance structures (ACS):** Just allow correlation over occasions to exist (for unknown reasons)
  3. **Add a "level" (or more):** Use random effect (latent factor) variances, as possible within multilevel or structural equation modeling

1. Modeling Longitudinal Dependency

- **Fixed effects:** Add Person ID as a categorical predictor
- Estimate fixed effects of $N - 1$ dummy codes for person ID
  - Person ID **main effects** capture dependency due to mean differences
  - Interactions of Person ID with time-varying predictors (like time) capture other predictor-specific sources of person dependency
- Pro: Does adequately control for person dependency
  - Very common in econometrics, political science, sociology…
  - Does a better job in studies with "few" persons (< 15ish)
  - Useful to make individual-specific conclusions (i.e., as in aggregated N-of-1 randomized control trials)
- Con: Does not allow prediction of WHY any of those individual differences occurred ☹
  - Model would be saturated with respect to between-person differences

2. Modeling Longitudinal Dependency

- **Alternative multivariate variance–covariance structures:** Change model to allow correlation over occasions (and any residual heterogeneity) to exist
- Is only possible given **balanced data** (all people on same schedule) and conditionally normal outcomes (i.e., not when using generalized models)

- Is the basis of **repeated measures ANOVA**, of which there are **2 kinds**
  - "**Univariate approach**": residuals have equal variance and equal correlations across all repeated measures outcomes—but this "compound symmetry" pattern can only possibly hold if all people change the same!
  - "**Multivariate approach**": all residual variances and correlations are separately estimated—but this "unstructured" (MANOVA) model becomes difficult-to-impossible given many outcomes (especially with few people)
  - Estimation using ordinary least squares → listwise deletion of missing data ☹

- Switching to maximum likelihood estimation uses all complete occasions AND offers more choices for patterns of residual variance and correlation
  - Btw, residual maximum likelihood = ordinary least squares given complete outcomes
  - e.g., Compound Symmetry Heterogeneous (diff variances, equal correlation)
  - Options that use time-lagged covariances also require equal-interval occasions: e.g., First-order auto-regressive, moving average, or antedependence; Toeplitz

3. Modeling Longitudinal Dependency

- **Add a "level"** → Add random effect (latent variable) variances
- Random effect = model term that each person gets their own version of (in theory); directly incorporated by estimating the variance of each random effect across persons → BP differences
  - Capture patterns of non-constant variance and covariance for testable reasons
  - Works for general or generalized models (i.e., for any kind of outcome)
  - Works for balanced or unbalanced longitudinal data
- More generally, a "level" is a dimension of sampling that has unexplained outcome variability represented by 1+ random effects
  - "time" is not a level once sufficient fixed effects for its mean diffs are included
  - e.g., Randomized Control Trial (RCT) of 5 monthly occasions → 2 levels (1. within-person, 2. between-person)
  - e.g., Ecological Momentary Assessment (EMA) design of 4 observations per day for 3 weeks → 3 levels (1. within-day, 2. between-day, 3. between-person)

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)

Introduction to Concepts and Terminology in Longitudinal Modeling

- Topics:
  - Concepts and terminology in longitudinal models
  - Modeling person dependency
  - **Fixed and random intercepts**
  - Fixed and random time slopes
  - From multilevel models (MLMs) to single-level structural equation models (SEMs) to multilevel SEMs (M-SEMs)
  - Details

The Two Sides of a General Linear Model

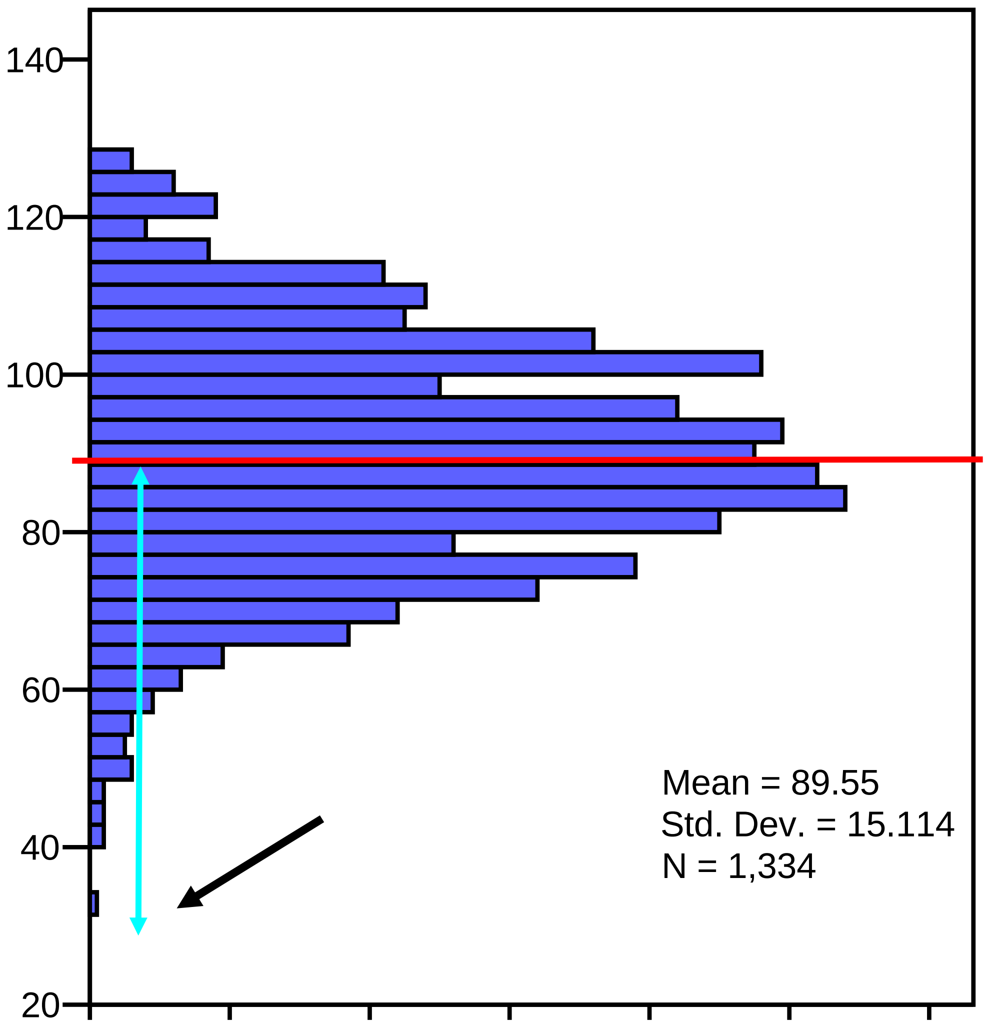
$$y_i = \boxed{\beta_0 + \beta_1(x1_i) + \beta_2(x2_i) + \cdots} \boxed{+\, e_i}$$

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 Model for the Variance



$$y_i = \beta_0 + e_i$$

Filling in values:

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

Model for the Means

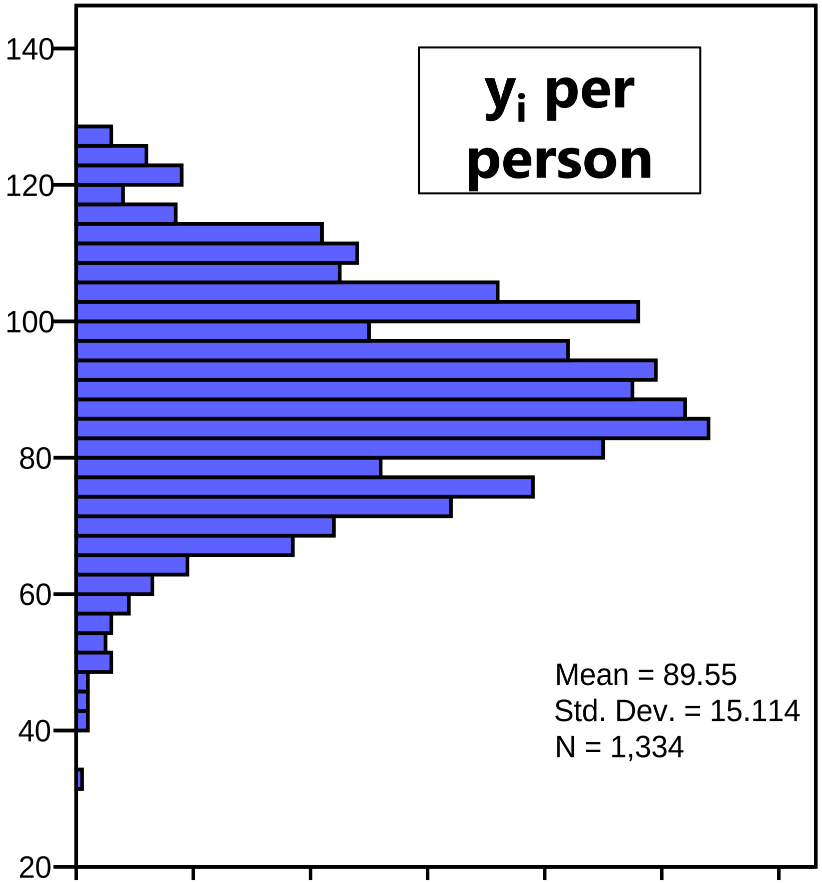
$\hat{y}_i$ = **"y-hat" model-predicted outcome**

y_i residual ("error") variance:

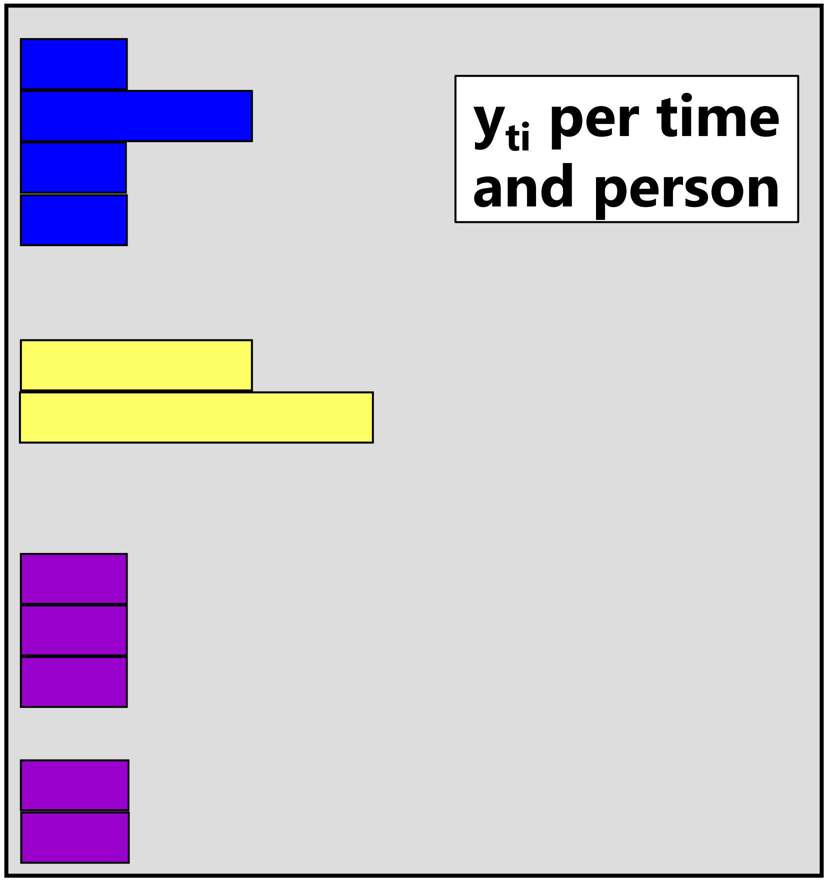
$$\frac{\sum(y_i - \hat{y}_i)^2}{N-1}$$

Adding Repeated Occasions → Two-Level Model for the Variance

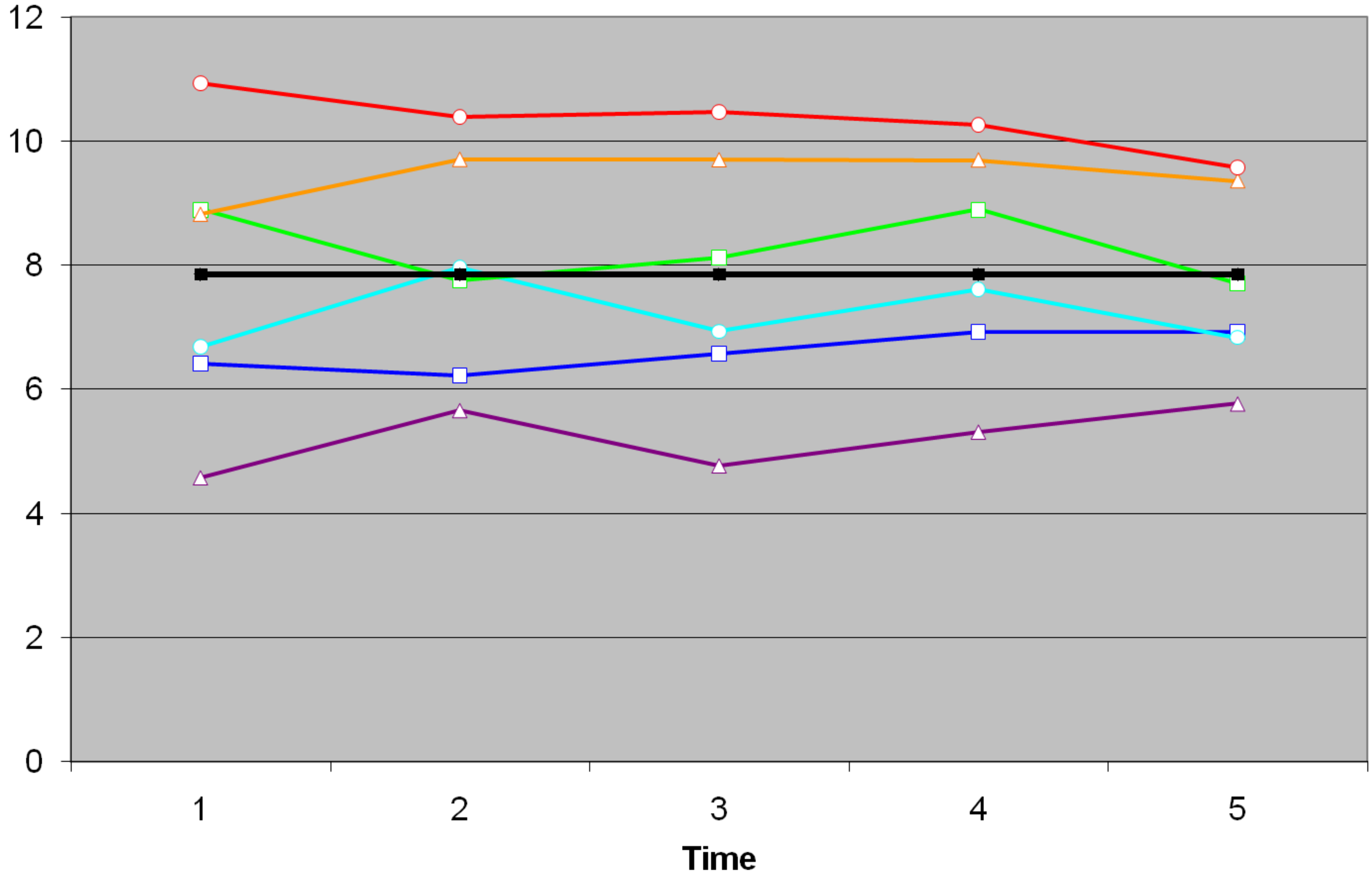
Full Sample Distribution



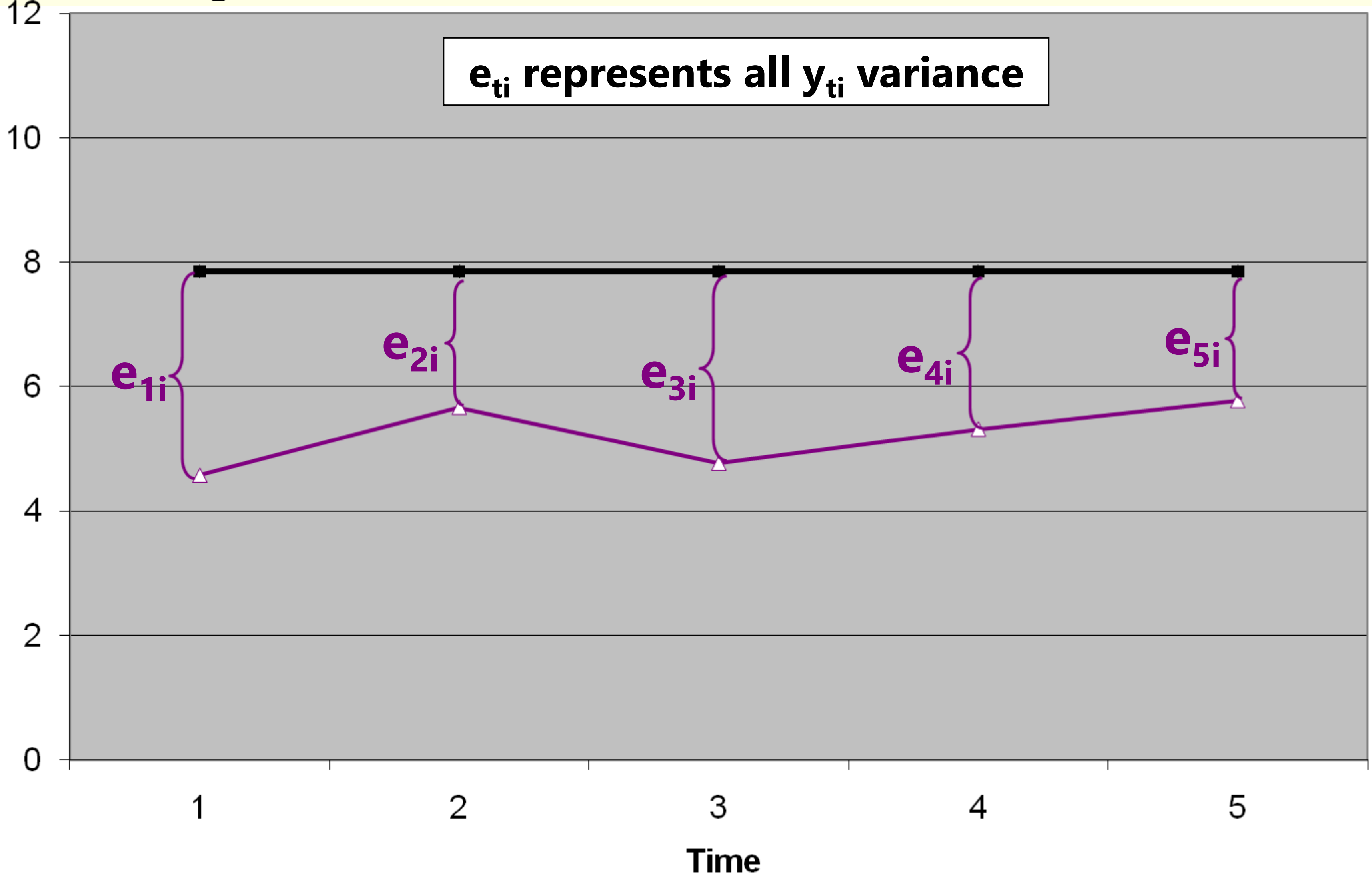
5 Occasions (t); 3 People (i)



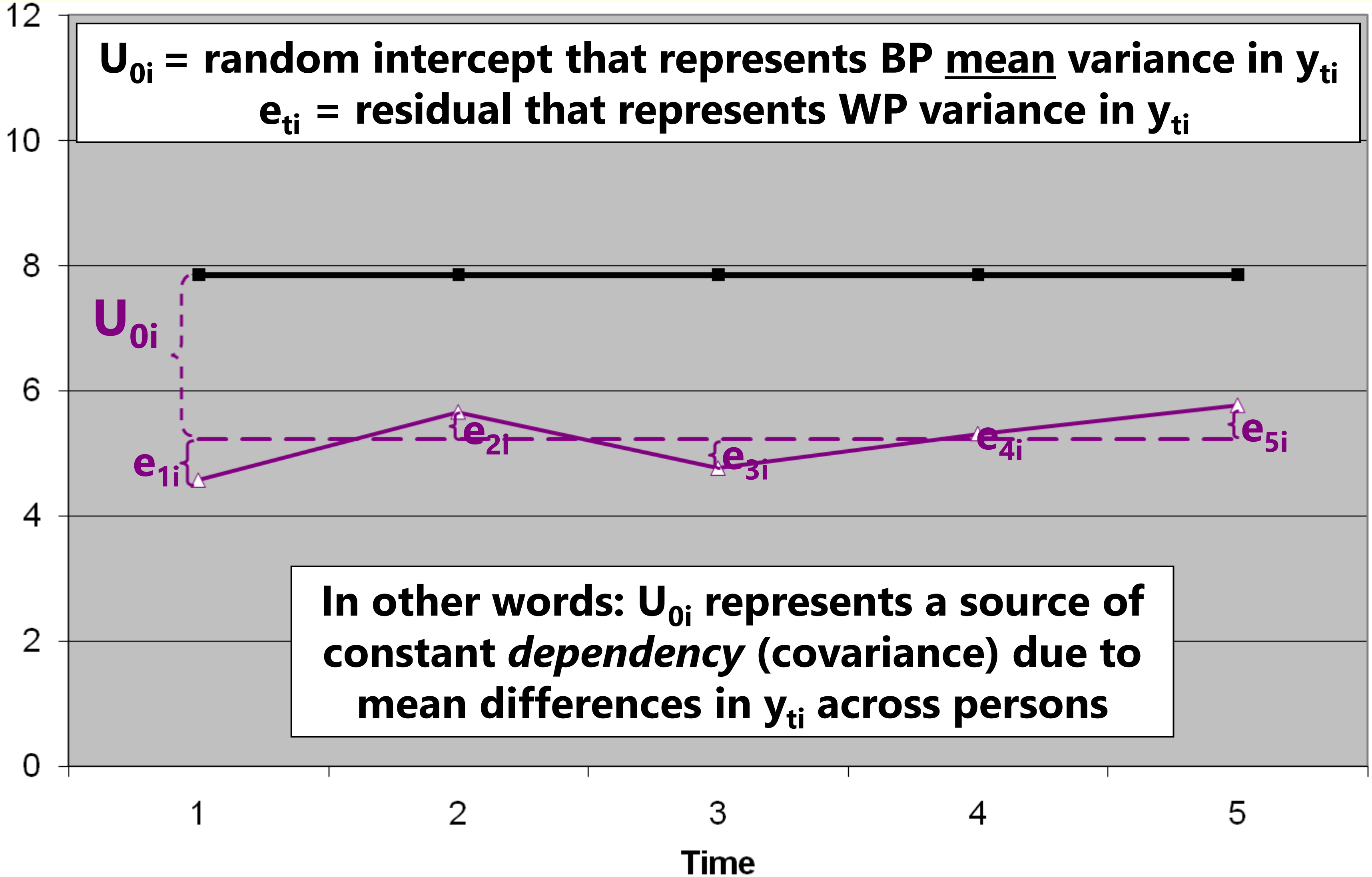
Hypothetical Longitudinal Data



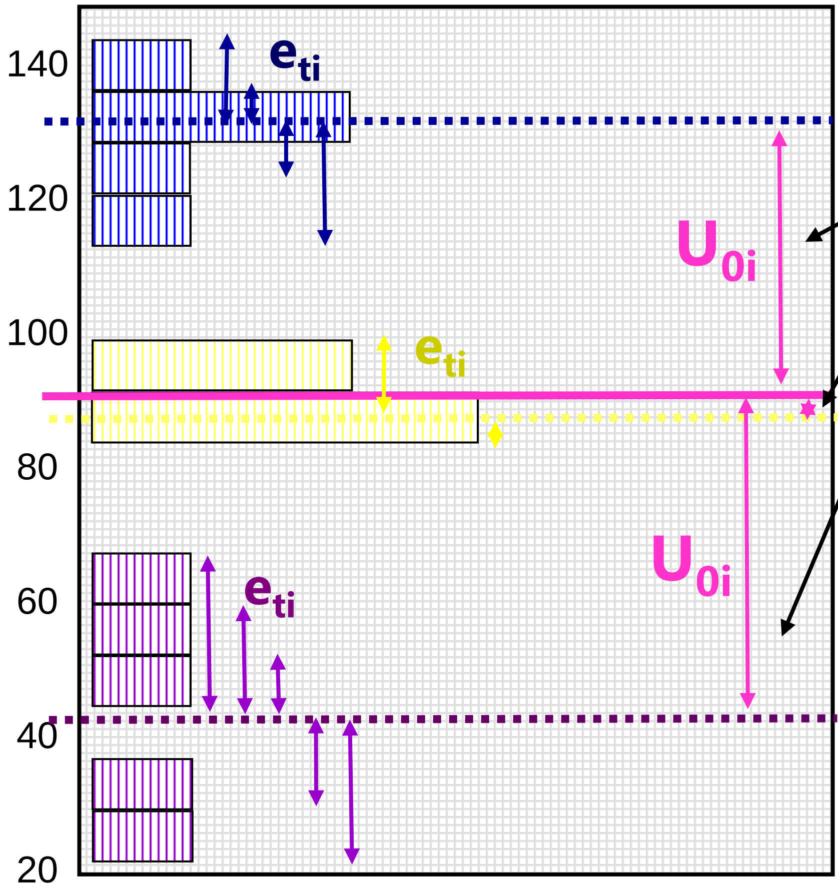
Only One Kind of "Error" in a Single-Level Model for the Variance



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

Empty Means Models: Single-Level vs. Two-Level

- Empty Means, **Single-Level Model** (used for 1 occasion):

$$y_i = \beta_0 + e_i$$

  - β_0 = fixed intercept = grand mean
  - e_i = residual deviation from GRAND mean

- Empty Means, **Two-Level** Model (for 2+ occasions):

$$y_{ti} = \beta_0 + U_{0i} + e_{ti}$$

  - β_0 = fixed intercept = grand mean
  - U_{0i} = random intercept = individual deviation from GRAND mean
  - e_{ti} = time-specific residual deviation from OWN mean

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 $e_{ti} \rightarrow \sigma_e^2$
- Level-2 **BP** Variance of $U_{0i} \rightarrow \tau_{U_0}^2$

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:

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

A "Random Intercept" Model for the Variance

Total Predicted Data Matrix is called V Matrix, and each person gets their own!

$$\begin{bmatrix} \tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} \\ \tau^2_{U_0} & \tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} \\ \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0} & \tau^2_{U_0} \\ \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0} \\ \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0}+\sigma^2_e \end{bmatrix}$$

***N* = total obs**
***n* = # occasions (5 here)**

Level 2, BP Variance

Unstructured **G Matrix** (**RANDOM** statement)

Each person has same **1 x 1 G** matrix (no covariance across persons in two-level model)

1 Random Intercept Variance only $\begin{bmatrix} \tau^2_{U_0} \end{bmatrix}$

To be added to **R** in order to form **V**, **G** is pre- and post-multiplied by an ***N* x 1 Z** matrix that holds the values of the predictors with random effects (just the intercept here): $\mathbf{V_i} = \mathbf{Z_i G_i Z_i^T} + \mathbf{R_i}$

Level 1, WP Variance

Diagonal (VC) **R Matrix** (**REPEATED** statement)

Each person has same ***n* x *n* R** matrix → **equal variances and 0 covariances** across time (and no covariance across persons)

1 Residual Variance only

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

Intraclass Correlation (ICC)

ICCs for two-level longitudinal data:

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

$$\text{Corr}(y_1,y_2) = \frac{\text{Cov}(y_1,y_2)}{\sqrt{\text{Var}(y_1)} * \sqrt{\text{Var}(y_2)}}$$

V matrix

$$\begin{bmatrix} \tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} \\ \tau^2_{U_0} & \tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} \\ \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0} & \tau^2_{U_0} \\ \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0} \\ \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0}+\sigma^2_e \end{bmatrix}$$

VCORR Matrix

$$\begin{bmatrix} 1 & \text{ICC} & \text{ICC} & \text{ICC} & \text{ICC} \\ \text{ICC} & 1 & \text{ICC} & \text{ICC} & \text{ICC} \\ \text{ICC} & \text{ICC} & 1 & \text{ICC} & \text{ICC} \\ \text{ICC} & \text{ICC} & \text{ICC} & 1 & \text{ICC} \\ \text{ICC} & \text{ICC} & \text{ICC} & \text{ICC} & 1 \end{bmatrix}$$

- ICC = Proportion of total variance that is **between persons**
- ICC = **Correlation of occasions** from same person (in VCORR)
- ICC is a standardized way to express *dependency due to person mean differences* → **effect size for <u>constant</u> person dependency**

Introduction to Concepts and Terminology in Longitudinal Modeling

- Topics:
  - Concepts and terminology in longitudinal models
  - Modeling person dependency
  - Fixed and random intercepts
  - **Fixed and random time slopes**
  - From multilevel models (MLMs) to single-level structural equation models (SEMs) to multilevel SEMs (M-SEMs)
  - Details

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):

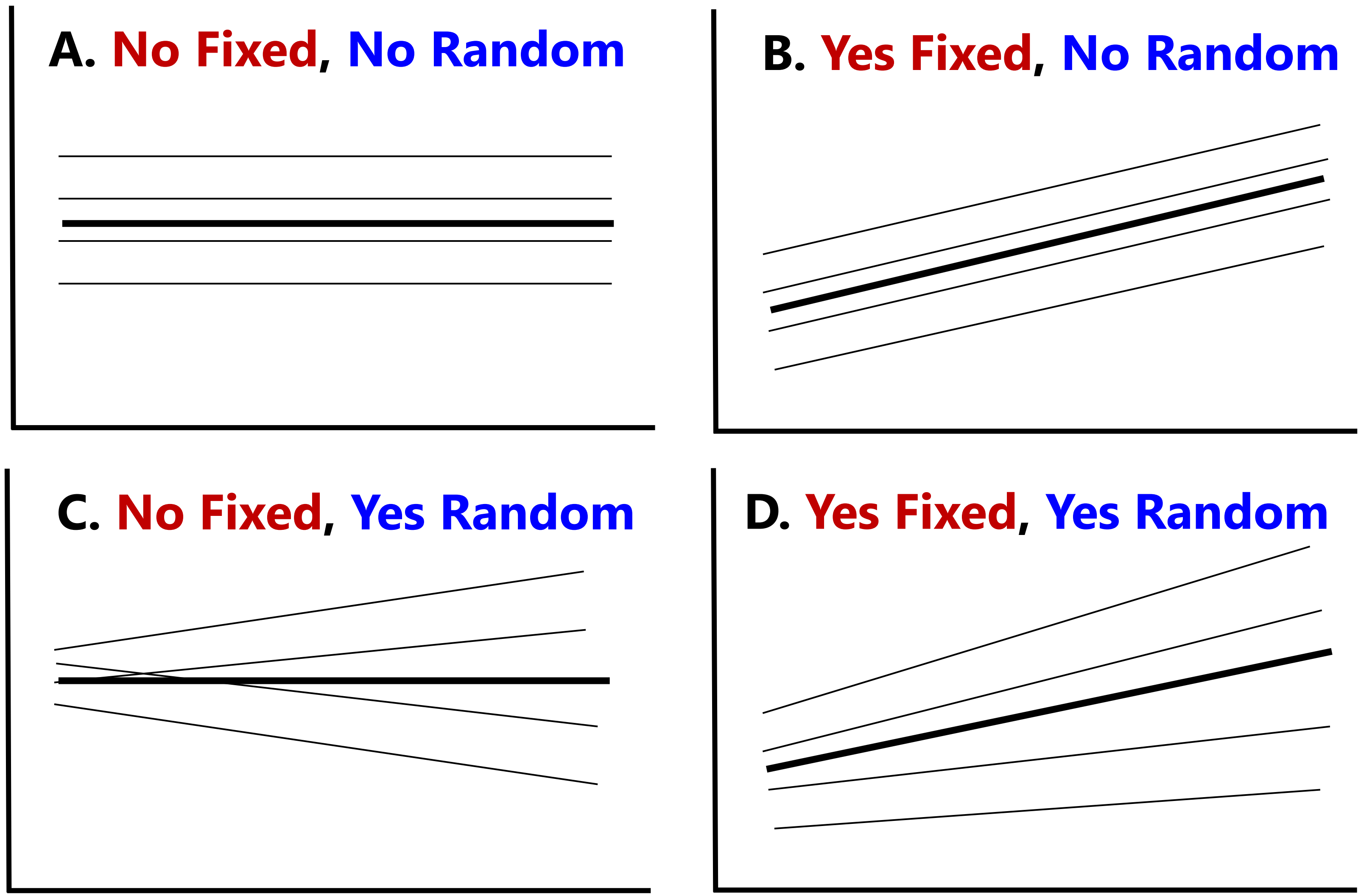
1. **Is there an effect of time on average?**
   - Is the line connecting the sample means not flat?
   - If so, you need **FIXED** effect(s) of time

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)

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

Multilevel Model

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 exactly the same rate

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

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

Multilevel Model

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

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

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

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

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

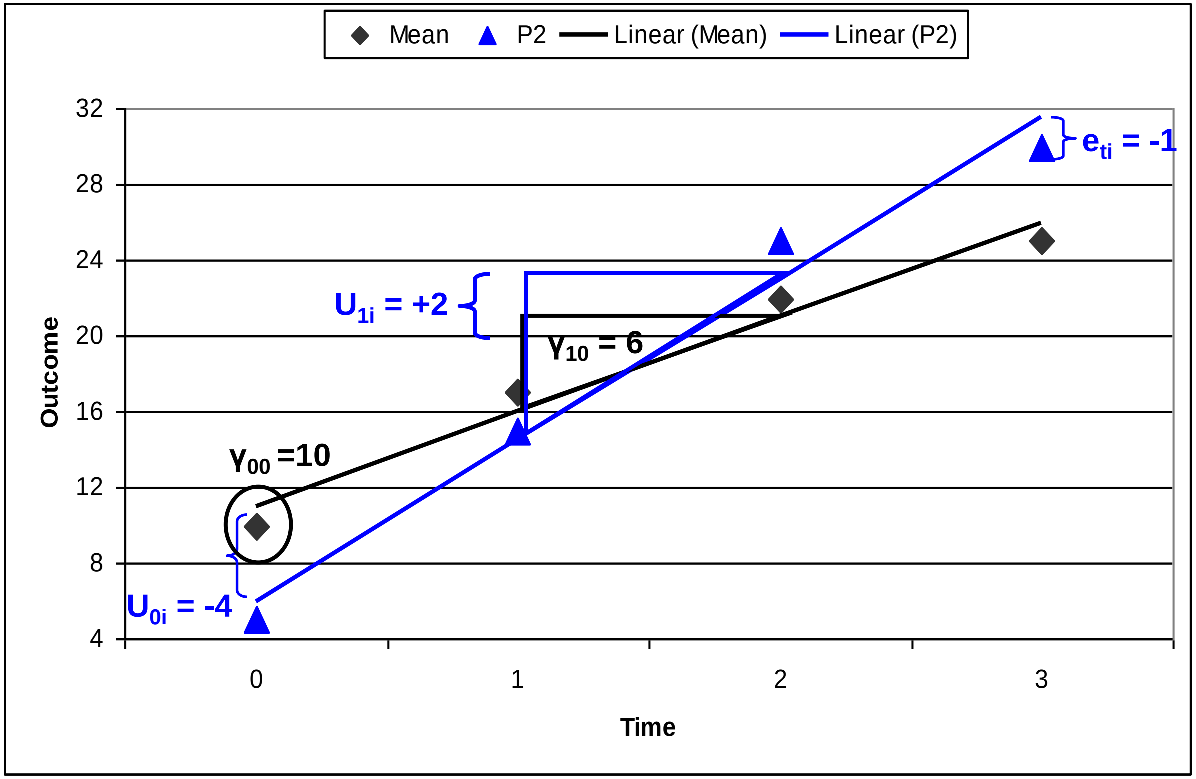
γ_{00}: **Fixed Intercept**

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

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

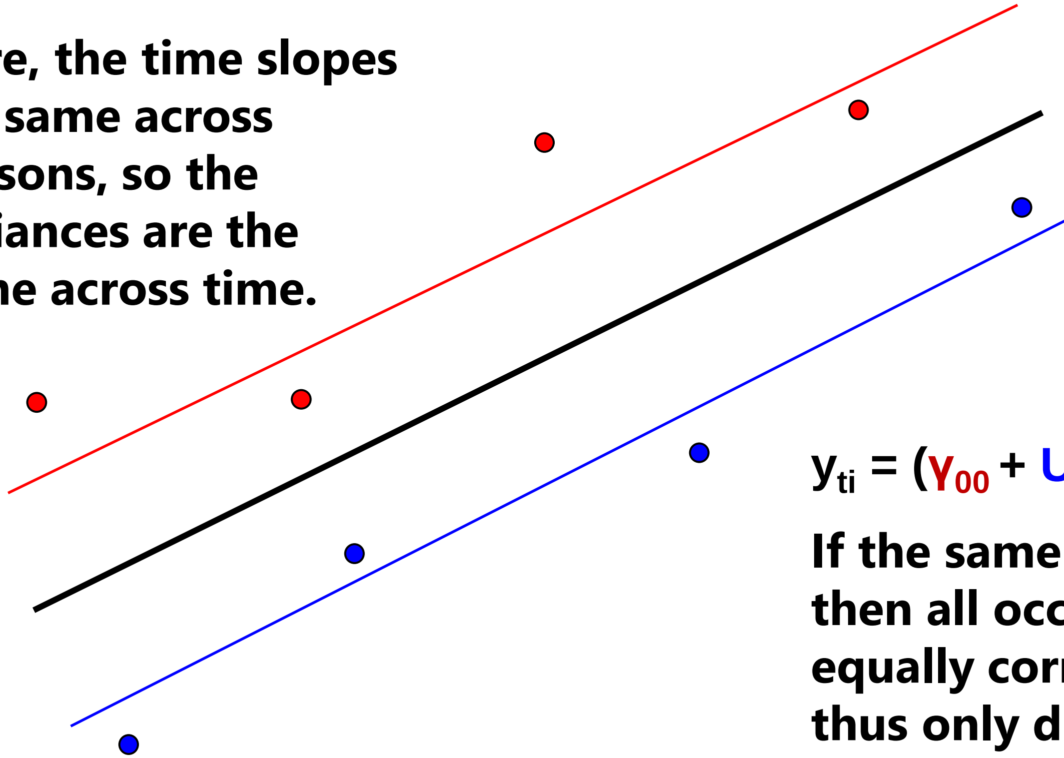
Level-1 **R** matrix: REPEATED TYPE=VC

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

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)

Choices in Modeling Variances: Random Intercept Only (Compound Symmetry)

Here, the time slopes are same across persons, so the variances are the same across time.



$$\begin{bmatrix} \tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} \\ \tau^2_{U_0} & \tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} \\ \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0} & \tau^2_{U_0} \\ \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0} \\ \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0}+\sigma^2_e \end{bmatrix}$$

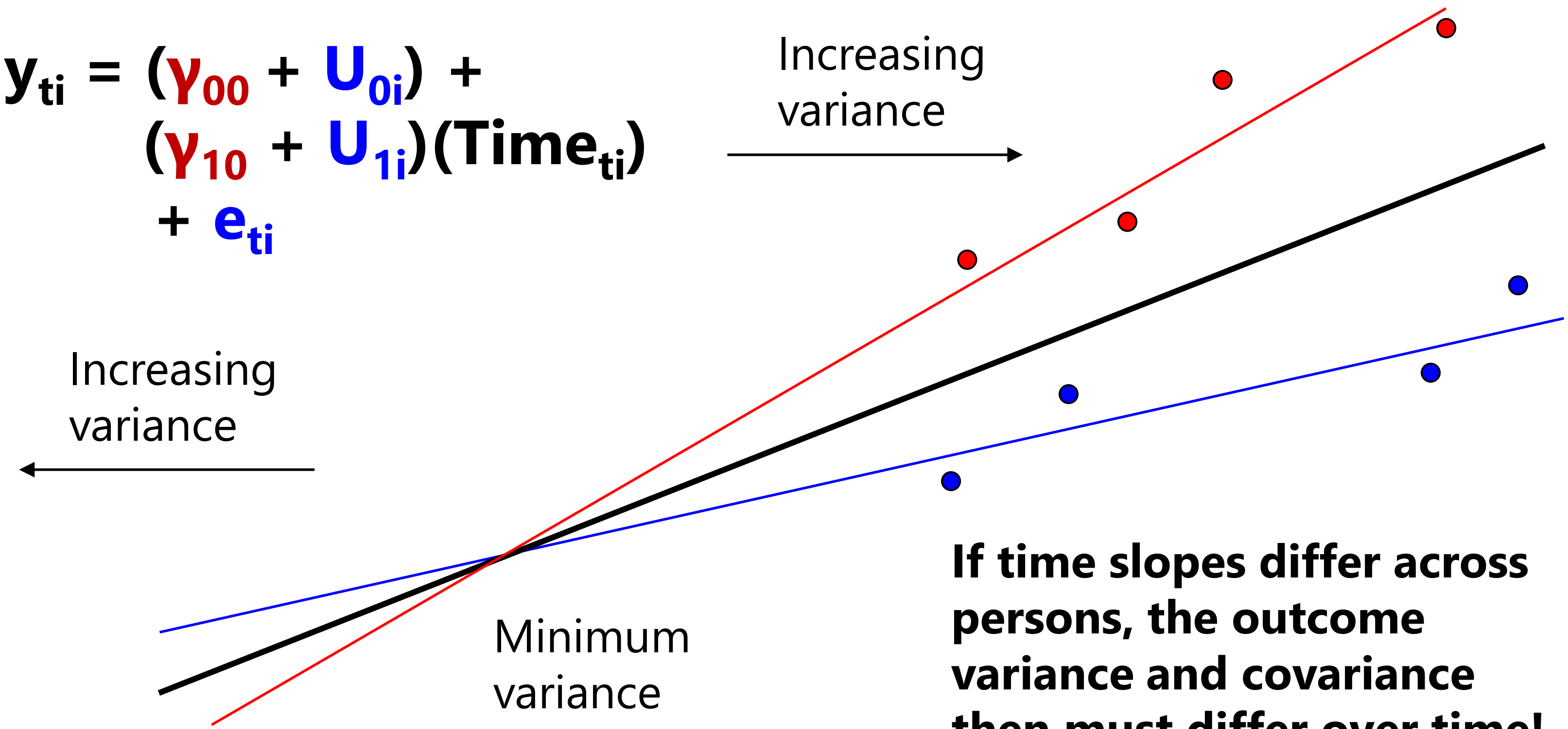
$$y_{ti} = (\gamma_{00} + U_{0i}) + (\gamma_{10})(Time_{ti}) + e_{ti}$$

If the same slope fits all persons, then all occasions should be equally correlated over time (and thus only due to U_{0i} variance).

If the time slopes are the same across people, then people differ from each other systematically in only 1 way (i.e., their **U_{0i}** level) → THIS IS COMPOUND SYMMETRY.

Choices in Modeling Variances: Random Intercepts and Time Slopes

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



If time slopes differ across persons, the outcome variance and covariance then must differ over time!

If slopes are different across people, then people differ from each other systematically in 2 ways (U_{0i} and U_{1i})
→ this implies compound symmetry will NOT hold.

Random Linear Time Model
(6 parameters: effect of time is RANDOM)

- Scalar "mixed" model equation per person:

$$\mathbf{Y}_i = \boxed{\mathbf{X}_i * \boldsymbol{\gamma}} + \boxed{\mathbf{Z}_i * \mathbf{U}_i + \mathbf{E}_i}$$

$$\begin{bmatrix} y_{0i} \\ y_{1i} \\ y_{2i} \\ y_{3i} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} \gamma_{00} \\ \gamma_{10} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} U_{0i} \\ U_{1i} \end{bmatrix} + \begin{bmatrix} e_{0i} \\ e_{1i} \\ e_{2i} \\ e_{3i} \end{bmatrix}$$

$$\begin{bmatrix} y_{0i} \\ y_{1i} \\ y_{2i} \\ y_{3i} \end{bmatrix} = \begin{bmatrix} \gamma_{00} + \gamma_{10}(0) \\ \gamma_{00} + \gamma_{10}(1) \\ \gamma_{00} + \gamma_{10}(2) \\ \gamma_{00} + \gamma_{10}(3) \end{bmatrix} + \begin{bmatrix} U_{0i} + U_{1i}(0) \\ U_{0i} + U_{1i}(1) \\ U_{0i} + U_{1i}(2) \\ U_{0i} + U_{1i}(3) \end{bmatrix} + \begin{bmatrix} e_{0i} \\ e_{1i} \\ e_{2i} \\ e_{3i} \end{bmatrix}$$

$$\begin{bmatrix} y_{0i} \\ y_{1i} \\ y_{2i} \\ y_{3i} \end{bmatrix} = \begin{bmatrix} \gamma_{00} + \gamma_{10}(0) + U_{0i} + U_{1i}(0) + e_{0i} \\ \gamma_{00} + \gamma_{10}(1) + U_{0i} + U_{1i}(1) + e_{1i} \\ \gamma_{00} + \gamma_{10}(2) + U_{0i} + U_{1i}(2) + e_{2i} \\ \gamma_{00} + \gamma_{10}(3) + U_{0i} + U_{1i}(3) + e_{3i} \end{bmatrix}$$

$\mathbf{X}_i$ = n x k values of **predictors with fixed effects**, so can differ per person (k = 2: intercept, linear time)

$\boldsymbol{\gamma}$ = k x 1 estimated **fixed effects**, so will be the same for all persons (γ_{00} = intercept, γ_{10} = linear time)

$\mathbf{Z}_i$ = n x u values of **predictors with random effects**, so can differ per person (u = 2: intercept, linear time)

$\mathbf{U}_i$ = u x 2 estimated individual **random effects**, so can differ per person

$\mathbf{E}_i$ = n x n time-specific residuals, so can differ per person

Random Linear Time Model
(6 parameters: effect of time is RANDOM)

- Predicted total variances and covariances per person:

$$\mathbf{V_i} = \mathbf{Z_i} * \mathbf{G_i} * \mathbf{Z_i^T} + \mathbf{R_i}$$

$$\mathbf{V_i} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} \tau^2_{U_0} & \tau_{U_{01}} \\ \tau_{U_{01}} & \tau^2_{U_1} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{bmatrix} + \begin{bmatrix} \sigma^2_e & 0 & 0 & 0 \\ 0 & \sigma^2_e & 0 & 0 \\ 0 & 0 & \sigma^2_e & 0 \\ 0 & 0 & 0 & \sigma^2_e \end{bmatrix}$$

$\mathbf{V_i}$ matrix: $\text{Variance}\left[y_{time}\right]$

$$= \tau^2_{U_0} + \left[(\text{time})^2 \tau^2_{U_1}\right] + \left[2(\text{time})\tau_{U_{01}}\right] + \sigma^2_e$$

$\mathbf{V_i}$ matrix: $\text{Covariance}\left[y_A, y_B\right]$

$$= \tau^2_{U_0} + \left[(A+B)\tau_{U_{01}}\right] + \left[(AB)\tau^2_{U_1}\right]$$

$\mathbf{V_i}$ matrix = complicated ☺

$\mathbf{Z_i}$ = *n* x *u* values of **predictors with random effects**, so can differ per person (*u* = 2: int., time slope)

$\mathbf{Z_i}^T$ = *u* x *n* values of predictors with random effects (just $\mathbf{Z_i}$ transposed)

$\mathbf{G_i}$ = *u* x *u* estimated **random effects variances and covariances**, so will be the same for all persons ($\tau^2_{U_0}$ = int. var., $\tau^2_{U_1}$ = slope var.)

$\mathbf{R_i}$ = *n* x *n* **time-specific residual variances and covariances**, so will be same for all persons (here, just diagonal σ^2_e)

Building **V** across persons: Random Linear Time Model

- **V** for two persons also with ***different n*** per person:

$$\mathbf{V} = \qquad \mathbf{Z} \qquad * \qquad \mathbf{G} \qquad * \qquad \mathbf{Z}^T \qquad + \qquad \mathbf{R}$$

$$\mathbf{V} = \begin{bmatrix} 1 & 0.0 & 0 & 0 \\ 1 & 1.0 & 0 & 0 \\ 1 & 2.0 & 0 & 0 \\ 1 & 3.0 & 0 & 0 \\ 0 & 0 & 1 & 0.2 \\ 0 & 0 & 1 & 1.4 \\ 0 & 0 & 1 & 3.5 \end{bmatrix} \begin{bmatrix} \tau^2_{U_0} & \tau_{U_{01}} & 0 & 0 \\ \tau_{U_{01}} & \tau^2_{U_1} & 0 & 0 \\ 0 & 0 & \tau^2_{U_0} & \tau_{U_{01}} \\ 0 & 0 & \tau_{U_{01}} & \tau^2_{U_1} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0.0 & 1.0 & 2.0 & 3.0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0.2 & 1.4 & 3.5 \end{bmatrix} + \begin{bmatrix} \sigma^2_e & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma^2_e & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2_e & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2_e & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2_e & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma^2_e & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2_e \end{bmatrix}$$

- The "block diagonal" does not need to be the same size or contain the same time observations per person…
- **R** matrix can also include non-0 covariance or differential residual variance across time (as in ACS models), although many models based on the idea of a "lag" won't work for unequal-interval time (but AR1 can be modified to work)

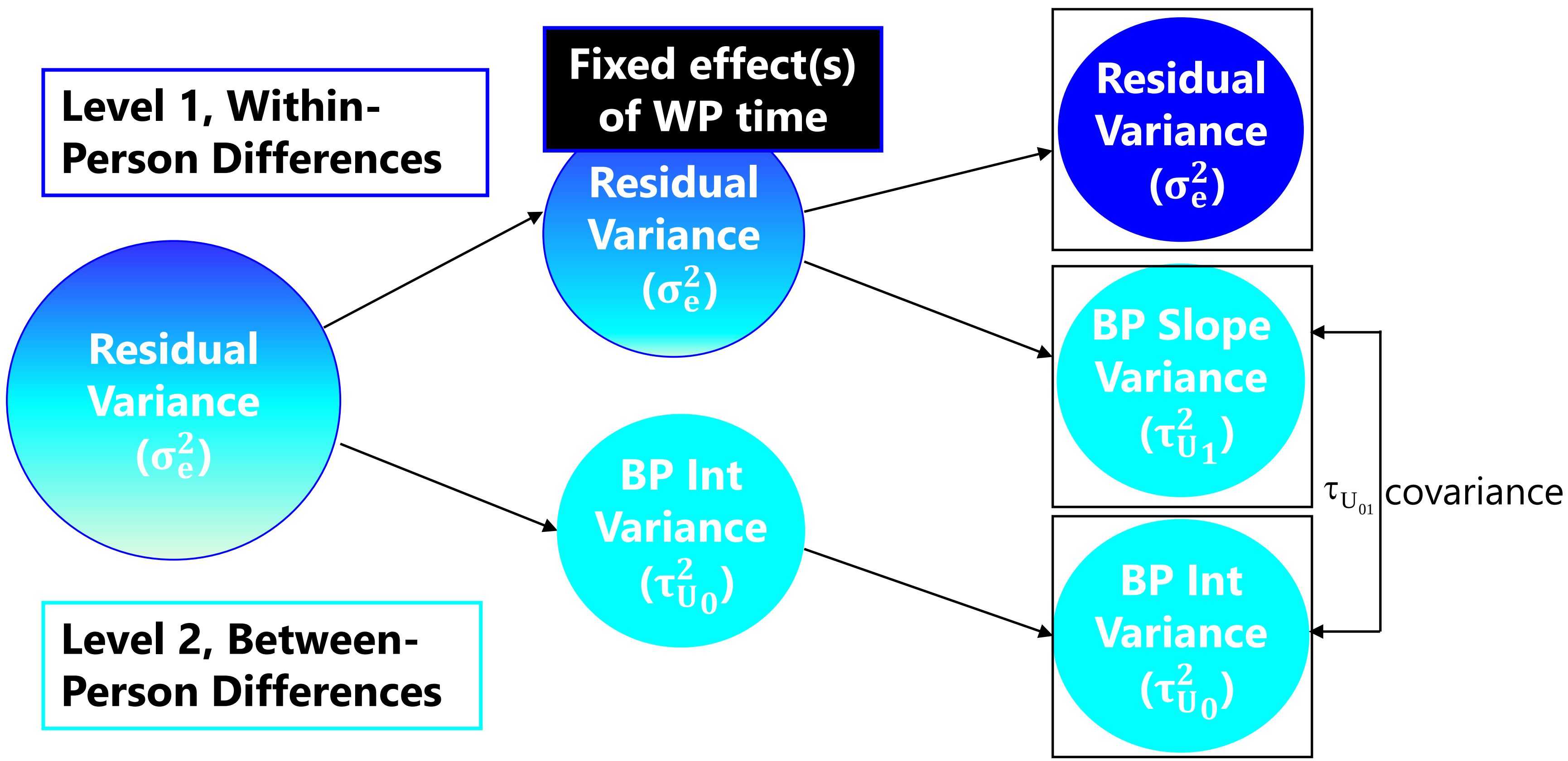
The Bigger Picture

- **Random effects** (new "piles" of variance, partitioned out of what used to be a single residual variance) are used to capture sources of **person dependency**
  - Random **intercept → constant** correlation over time due to person mean differences → univariate RM ANOVA
  - Random **time slope(s) → non-constant** correlation over time and non-constant variance over time due to between-person differences in rate(s) of change over time
  - Foreshadowing: random time-varying x_{ti} slope → heterogeneity over x_{ti}

- After accounting for BP level-2 random effects (intercepts, and any slopes for change over time), **WP level-1 residuals** are usually assumed **uncorrelated** with **constant variance**
  - But these are both testable assumptions! (fewer alternatives in unbalanced data, largely due to software inflexibility)
  - All sources of person dependency related to time should be addressed before considering other predictors!
  - Any longitudinal model not accounting for person dependency due to intercepts (at a minimum) is most likely to be WAY wrong (AR-CLPM!)

Summary: "Handling" Person Dependency

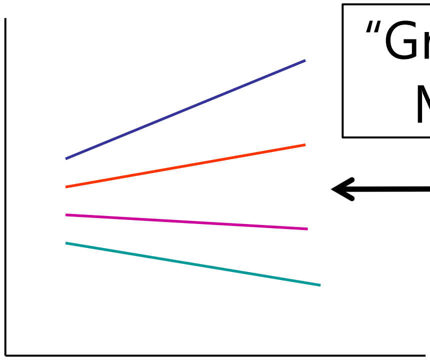
- The process of fitting "unconditional models for time" (fixed and random effects) can be depicted as follows:



Summary: Unconditional Models for Time

Pure WP Change

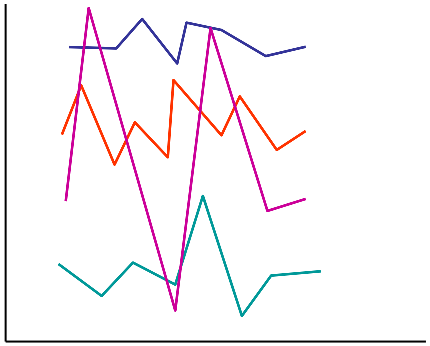
“Growth Curve Modeling”



Time

Pure WP Fluctuation

“EMA”



Time

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)

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

- 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...

Soon we will add predictors to account for each pile!

Introduction to Concepts and Terminology in Longitudinal Modeling

- Topics:
  - Concepts and terminology in longitudinal models
  - Modeling person dependency
  - Fixed and random intercepts
  - Fixed and random time slopes
  - **From multilevel models (MLMs) to single-level structural equation models (SEMs) to multilevel SEMs (M-SEMs)**
  - Details

Translating MLMs into SEMs…

- "**Random effects**" = "pile of variance" = "variance components"
  - Random effects represent "person*predictor" interaction terms
  - Random intercept → person*intercept (person "main effect")
  - Random linear slope → person*time interaction
  - Capture **specific patterns of covariation** of unknown origin…
    - *Why do people need their own random intercepts and slopes?* (We can add person-level predictors to answer these questions)

- Random effects can also be seen as **latent variables**
  - Latent variable = unobservable construct (ability or trait)
    - Latent variables are created from the common variance across indicators
    - In longitudinal data, the latent variables can be thought of as "general tendency" and "propensity to change" as created by measuring the same outcome over time (occasions → indicators)
  - Let's see how MLMs can be estimated as single-level SEMs…

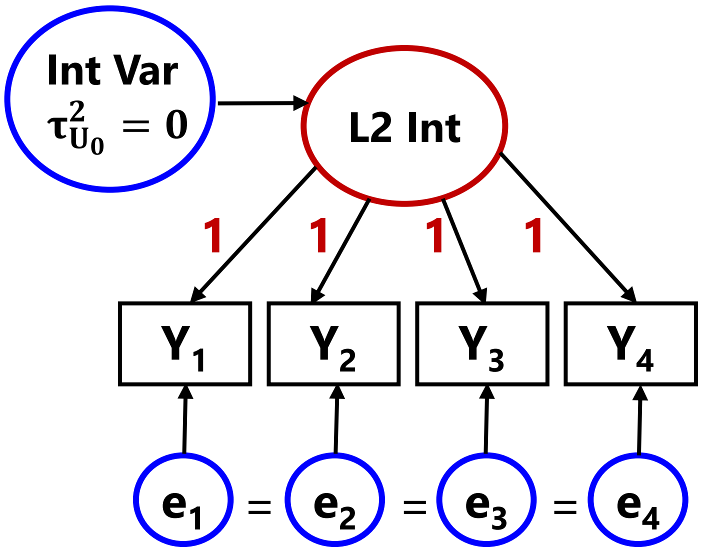
Structural Equation Models (SEMs)

- **Measurement model:** $\mathbf{y_{is} = \mu_i + \lambda_i F_s + e_{is}}$
  - Observed response for item *i* and subject *s*
    = intercept of item *i* (μ_i)
    + subject *s*'s latent trait/factor (F_s), item-weighted by λ_i
    + residual error (e_{is}) of item *i* and subject *s*

- Two big differences when using two factors for longitudinal data:
  - Usually two factors for "level" and "change" (intercept and slope):
    $\mathbf{y_{ti} = (\gamma_{00} + U_{0i}) + (\gamma_{10} + U_{1i})time_{ti} + e_{ti}}$ → **so each U → F**
  - Fixed effects → factor means; random effects → factor variances
  - The **occasion-specific intercepts** μ_i cannot be separately identified from the "intercept" latent factor and therefore must be fixed to 0
  - Factor loadings λ_i for how each outcome relates to the latent factor are (usually) pre-determined by how much time has passed → fixed to the difference in time across longitudinal outcomes
  - Unbalanced time requires "definition variables" → use variables for person-specific time loadings rather than fixing loadings to same values for all
    - In Mplus, is TSCORES option; could not find an equivalent option in R lavaan

Random Effects as Latent Variables

- **Single-level model for the variance → σ^2_e only**
  - **$y_{ti} = \gamma_{00} + e_{ti}$**



Mean of the intercept factor = fixed intercept γ_{00}

Loadings of intercept factor = 1 (all occasions contribute equally)

Indicator intercepts = 0 (always)

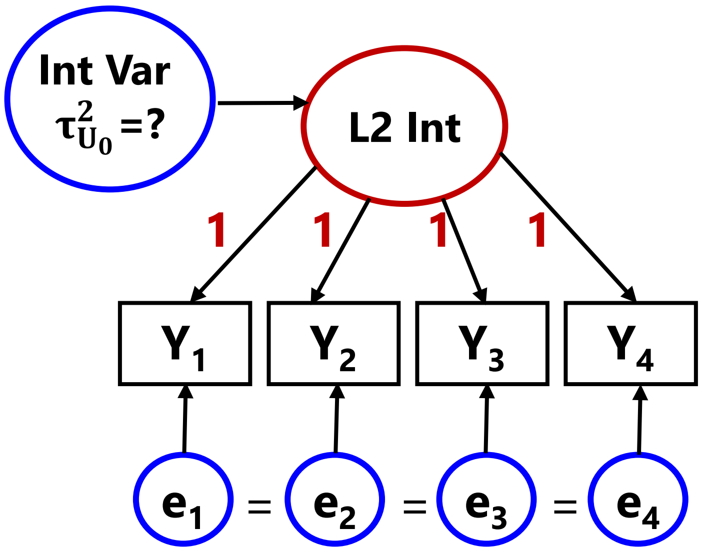
L2 variance of intercept factor $\tau^2_{U_0}$ = 0 so far

L1 residual variance (σ^2_e) is predicted to be equal across occasions

  - After controlling for the *fixed* intercept (factor mean), level-1 residuals are predicted to be uncorrelated

Random Effects as Latent Variables

- **Two-level model for the variance → add $\tau^2_{U_0}$**
  - $y_{ti} = \gamma_{00} + U_{0i} + e_{ti}$



Mean of the intercept factor = fixed intercept γ_{00}

Loadings of intercept factor= 1 (all occasions contribute equally)

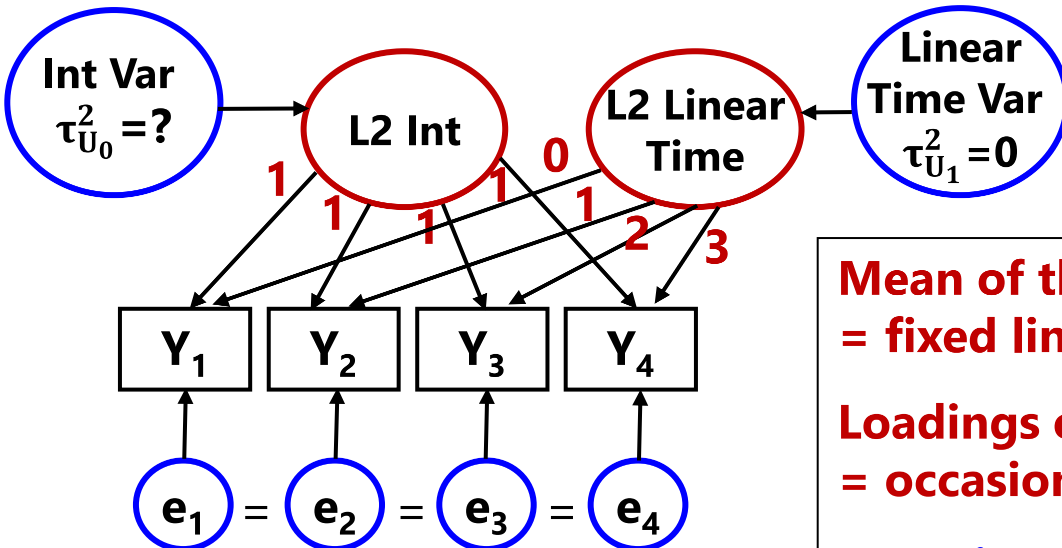
L2 variance of intercept factor $\tau^2_{U_0}$= random intercept variance

L1 residual variance (σ^2_e) is predicted to be equal across occasions

  - After controlling for the *random* intercept (factor mean and variance), level-1 residuals are predicted to be uncorrelated

Random Effects as Latent Variables

- **Fixed linear time, random intercept model:**
  - $y_{ti} = \gamma_{00} + (\gamma_{10}Time_{ti}) + U_{0i} + e_{ti}$



Mean of the linear time factor = fixed linear slope γ_{10}

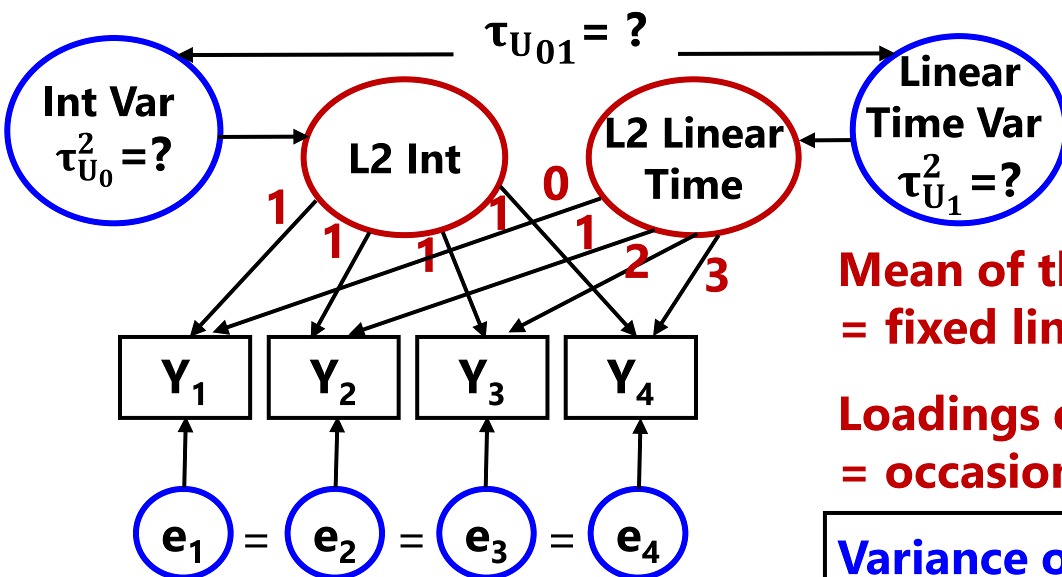
Loadings of linear time factor = occasions (keep real time)

L2 variance of linear time factor $\tau^2_{U_1}$ = 0

- After controlling for the *fixed linear time slope* (factor mean) and *random* intercept (factor mean and variance), level-1 residuals are predicted to be uncorrelated

Random Effects as Latent Variables

- **Random linear time model:**
  - **$y_{ti} = \gamma_{00} + (\gamma_{10}Time_{ti}) + U_{0i} + (U_{1i}Time_{ti}) + e_{ti}$**



Mean of the linear slope factor = fixed linear slope γ_{10}

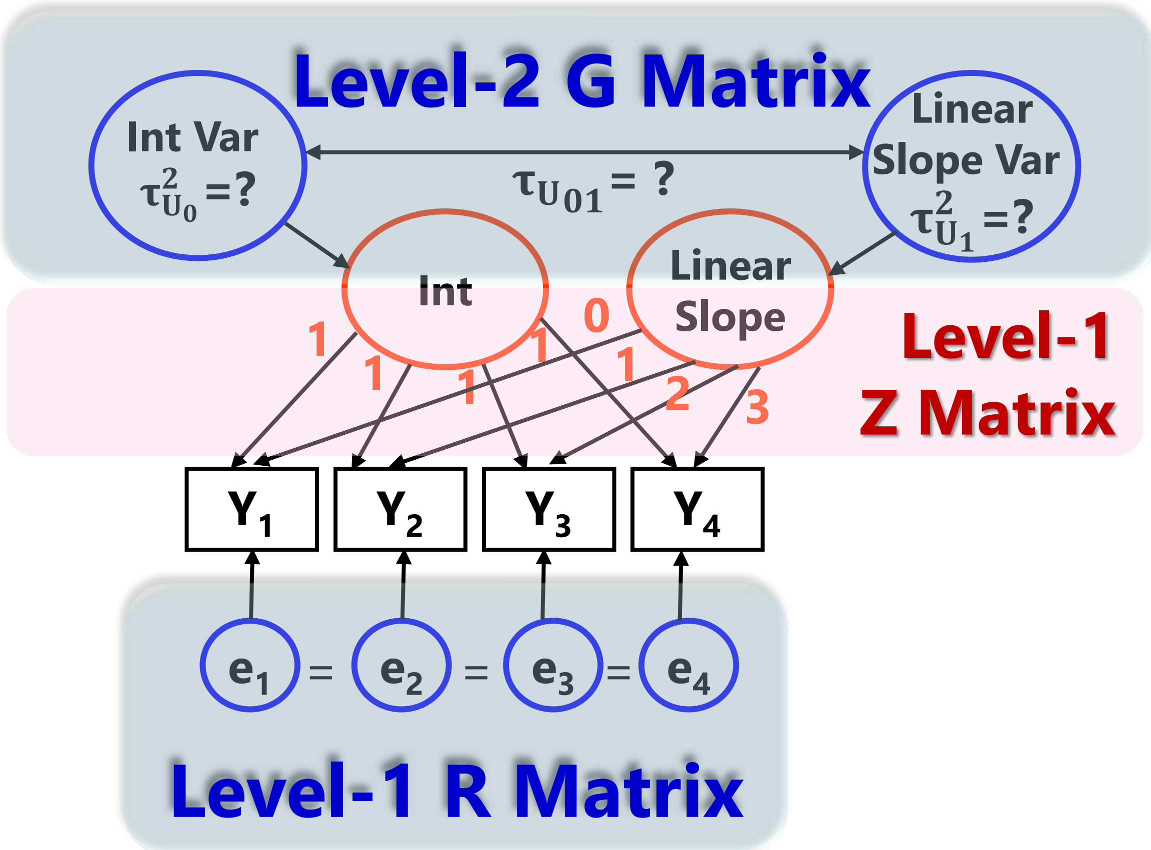
Loadings of linear slope factor = occasions (keep real time)

Variance of linear time factor $\tau^2_{U_1}$ = random slope variance

  - After controlling for the *random* linear time slope and *random* intercept (both factor means and variances), level-1 residuals are predicted to be uncorrelated

Random Linear Time Model: From MLM to Single-Level SEM

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

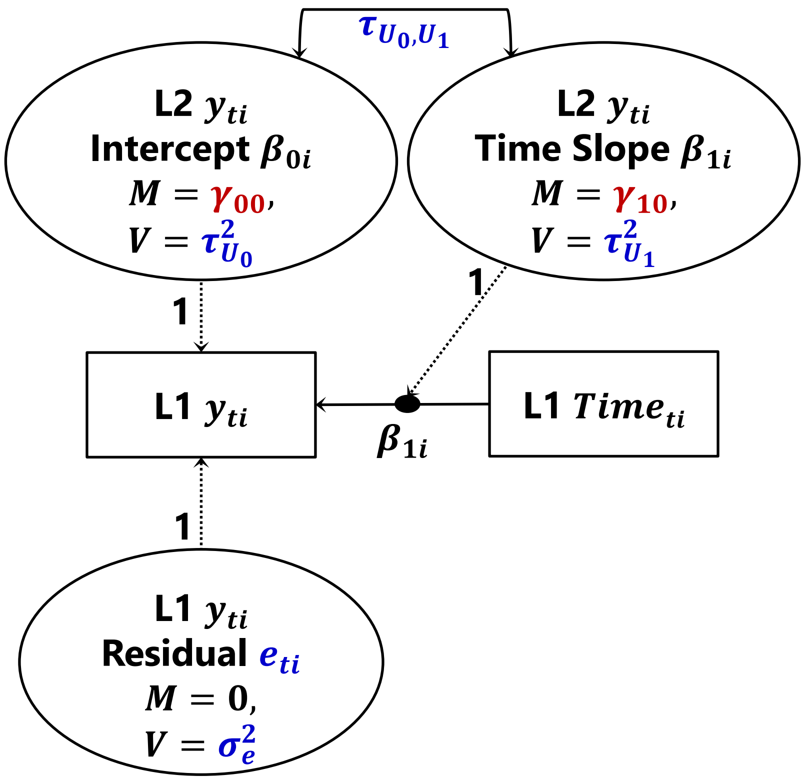


For unbalanced time, you need "definition variables" (like Mplus TSCORES) that allow different loadings (→ occasions) per person

Btw, allowing different residual variances for every occasion is going to be redundant with the random effects—they already predict variance to change over time!

Random Linear Time Model: From MLM to Multilevel SEM

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



Multilevel SEM (or what I prefer to call "multivariate MLM" in absence of a true measurement model) uses a long (stacked) data structure with one row per level-1 unit (so per occasion per person), just like univariate MLM.

The difference is that in M-SEM, multiple variables (predictors or outcomes) can have their variance partitioned into BP intercepts, BP slopes, and WP residuals at the same time (with additional features for autoregressive relations possible in "dynamic SEM", which is still M-SEM).

Summary: Three Frameworks for the Estimation of Longitudinal Models

- **Multilevel/Mixed/Hierarchical Linear Models: MLM → I start here**
  - Person dependency is captured primarily by **random effects** (through "**levels**" in **stacked/long data**, so occasions can be unbalanced and have multiple types of WP time)
  - **Univariate MLMs** (single outcome over time) are common (SAS, SPSS, or STATA MIXED; R lme4 and nlme), have REML and denominator DF for **small samples**, but can't do it all!
- **Single-Level Structural Equation Models: SEM**
  - Person dependency is captured by **latent variables** (through **multivariate outcomes** in **wide, single-level data**, so univariate occasions are treated as observed boxes)
  - Single-level SEM is common (Mplus, R lavaan), but **may not work for unbalanced data** or designs with **more than one level of time** (e.g., occasions within days within persons)
  - SEM software does not have **REML** or denominator DF (DDF) → bad for small samples
- **Multilevel Structural Equation Models: M-SEMs**
  - Estimated on stacked/long data, are more flexible for unbalanced time, less available (Mplus mainly), but **may break down in small *N*** (b/c no REML, no DDF, more parameters)
  - What is "**multilevel SEM**" (M-SEM) to others, I call "multivariate MLM" when they do not include true latent variable measurement models (i.e., as used in CFA, IFA, or IRT)

Introduction to Concepts and Terminology in Longitudinal Modeling

- Topics:
  - Concepts and terminology in longitudinal models
  - Modeling person dependency
  - Fixed and random intercepts
  - Fixed and random time slopes
  - From multilevel models (MLMs) to single-level structural equation models (SEMs) to multilevel SEMs (M-SEMs)
  - **Details**

Details: ML vs. REML Estimation

- What are REML and ML? Two flavors of likelihood estimation:
- **REML = "Restricted (or residual) maximum likelihood"**
  - Only available for general linear models or general linear mixed models (that assume normally distributed residuals); not in any SEM software
  - Is same as OLS given complete outcomes, but it doesn't require them
  - Estimates variances the same way as in OLS (accurate) → $\frac{\sum(y_{ti} - \boldsymbol{\hat{y}_{ti}})^2}{N-k}$
- **ML = "Maximum likelihood" (also called FIML*)**
  - Is more general, is available for the above plus for non-normal outcomes and latent variable models (CFA/SEM/IRT; multilevel SEM)
  - Is NOT the same as LS: it under-estimates variances by not accounting for the # of estimated fixed effects → $\frac{\sum(y_{ti} - \boldsymbol{\hat{y}_{ti}})^2}{N}$

**FI = Full information → it uses all the original data (they both do)*

Details: ML vs. REML Estimation

Remember "population" vs. "sample" formulas for calculating variance?

"Population"

$$\frac{\sum(y_i - \hat{\boldsymbol{y}}_{\boldsymbol{ti}})^2}{N}$$

"Sample"

$$\frac{\sum(y_i - \hat{\boldsymbol{y}}_{\boldsymbol{ti}})^2}{N-k}$$

All comparisons must have same *N*!!!	**ML**	**REML**
In software:	Only choice in SEM or M-SEM; available in MLM	Default in univariate general MLM programs
In estimating variances, it treats fixed effects as…	**Known** (df for having to also estimate fixed effects is not factored in)	**Unknown** (df for having to estimate fixed effects is factored in)
So, in small samples, L2 variances will be…	**Too small** (less difference after *N*=30-50 or so)	**Unbiased** (correct)
But because it indexes the fit of the…	**Entire model** (means + variances)	**Variances model only**
You can compare models differing in…	**Fixed and/or random effects** (either/both)	**Random effects only** (same fixed effects)

Details: Assessing Significance

- **Model for the Means** → which **fixed effects** of predictors should be included in the model (e.g., main effects, interactions)
  - **Significance tests** do not require assessment of relative model fit using −2ΔLL (can always use univariate or multivariate Wald tests)
  - **Effect sizes** can come from the significance tests (e.g., *t* → Cohen's *d* or partial *r*), or from reductions in variance (pseudo-R^2 or total-R^2)

- **Model for the Variance** → what pattern(s) of variance and covariance the residuals from the same unit have; what **random effects** are needed to describe these pattern(s)
  - **Significance tests** DO require assessing relative model fit via −2ΔLL
    - Cannot use the Wald test *p*-values for variances because those *p*-values use a two-sided sampling distribution, but variances cannot be negative
  - **Effect sizes** (less commonly provided) can come from random effects confidence intervals (CI) or random effects reliability measures
    - $\text{Random Effect 95\% CI} = \text{fixed effect} \pm \left(1.96 * \sqrt{\text{Random Variance}}\right)$

Details: Significance of Fixed Effects

	Denominator DF is infinite (Proper Wald test)	**Denominator DF is estimated instead** ("Modified" Wald test)
Numerator DF = 1 *(test one fixed effect)* is **Univariate Wald Test**	use z distribution (all of SEM; Mplus MLM, STATA MIXED default)	use t distribution (R nlme or lme4; SAS; SPSS; STATA MIXED with *dfmethod* and *small*)
Numerator DF > 1 *(test 2+ fixed effects)* is **Multivariate Wald Test**	use χ^2 distribution (Mplus, STATA default)	use F distribution (R glht; SAS, SPSS; STATA MIXED with *dfmethod* and *small*)
Options for estimating Denominator DF (DDF)	not applicable	R, SAS, STATA: Kenward-Roger R, SAS, STATA, SPSS: **Satterthwaite**

Details: Comparing Models for the Variance

- **Two strategies for choosing a model for the variance:**
  - Does the more complex model fit better (than a simpler model)?
  - Does the simpler model fit worse (than a more complex model)?

- Nested models are compared using a **"likelihood ratio test": −2ΔLL test** (aka, "χ^2 test" in SEM; "deviance difference test" in MLM)

"fewer" = from model with fewer parameters "more" = from model with more parameters	Results of 1. & 2. must be positive values!

1. Calculate **−2ΔLL:** if given −2LL, do $-2\Delta LL = (-2LL_{fewer}) - (-2LL_{more})$
   if given LL, do $-2\Delta LL = -2 * (LL_{fewer} - LL_{more})$
2. Calculate **Δdf** = $(\#\ Parms_{more}) - (\#\ Parms_{fewer})$
3. **Compare −2ΔLL to χ^2 distribution with df = Δdf**
4. Get *p*-value (e.g., from Excel CHIDIST, R anova, or STATA LRTEST)

Details: Comparing Models for the Variance

- What your *p*-value for the −2ΔLL test means:
  - If you **ADD** parameters, then your model can get **better** (if −2ΔLL test is significant) or **not better** (not significant)
  - If you **REMOVE** parameters, then your model can get **worse** (if −2ΔLL test is significant) or **not worse** (not significant)

- Nested or non-nested models can also be compared by **Information Criteria** that also reflect model parsimony
  - No significance tests or critical values, just "smaller is better"
  - **AIC** = Akaike IC = −2LL + 2 *(#parameters)
  - **BIC** = Bayesian IC = −2LL + log(N)*(#parameters)
  - What "parameters" means depends on flavor (except in STATA):
    - ML = ALL parameters; REML = variance model parameters only