Review of CLDP 944: Multilevel Models for Longitudinal Data

- Topics:
  - Review of general MLM concepts and terminology
  - Model comparisons and significance testing
  - Fixed and random effects of time
  - Significance testing and effect size in MLM

What is a Multilevel Model (MLM)?

- Same as other terms you have heard of:
  - **General(*ized*) Linear Mixed Model** (if you are from statistics)
    - *Mixed* = Fixed and Random effects
  - **Random Coefficients Model** (also if you are from statistics)
    - Random coefficients = Random effects = **latent variables/factors**
  - **Hierarchical Linear Model** (if you are from education)

- MLM is for modeling **dependency**. Special cases include:
  - Random Effects ANOVA or Repeated Measures ANOVA
  - (Latent) Growth Curve Model (where "Latent" implies use of SEM)
  - Within-Person Fluctuation Model (e.g., for daily diary data)
  - Clustered/Nested Observations Model (e.g., for kids in schools)
  - Cross-Classified Models (e.g., "value-added" models)
  - Psychometric Models (e.g., factor analysis, item response theory)

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 predictor variables

- **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 observations (persons, groups, time, etc.) → these relationships are called "dependency" and ***this is the primary way that multilevel models differ from general linear models*** *(e.g., regression)*

For Example: A Single-Level (BP) Model

$$y_i = \beta_0 + \beta_1 X_i + \beta_2 Z_i + \beta_3 X_i Z_i + e_i$$

= Single-Level

- **Model for the Means (→Predicted Values):**
  - Each person's expected (predicted) outcome is a weighted linear function of his/her values on X and Z (and here, their interaction), each measured once per person (i.e., this is a between-person model)
  - Estimated parameters are called **fixed effects** (here, β_0, β_1, β_2, and β_3)

- **Model for the Variance (→"Piles" of Variance):**
  - $e_i \sim N(0, \sigma_e^2)$→ ONE residual (unexplained) deviation, so **estimated parameter is residual variance in single-level (BP) model**
  - e_i residuals have a mean of 0 with some estimated **constant variance** σ_e^2, are **normally distributed**, are unrelated to X and Z, and are **independent** across all observations
  - *We should change models when any of these assumptions do not hold…*

Models We Will Learn in CLDP 945

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

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

Note: Least Squares is only for GLM

- **General*ized* Linear Models:** **any conditional outcome distribution**, **fixed** effects through **link functions**, no random effects (one dimension)

- **General Linear Mixed Models:** conditionally normal outcome distribution, **fixed and random effects** (identity link, but multiple sampling dimensions)
- **General*ized* Linear Mixed Models:** **any conditional outcome distribution**, **fixed and random effects** through **link functions** (multiple dimensions)
  - Many of the same concepts, but with more complexity in estimation

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

What kinds of designs will we analyze?

- "Longitudinal" data (still, but with more complexity)
  - Same individual units of analysis measured at different *occasions* (which could range from milliseconds to days to years)
  - Accelerated longitudinal designs; multiple levels of "time"
  - Multivariate models (e.g., for families, dyads, and mediation)
- "Repeated measures" (RM) data (not involving "time")
  - Same individual units of analysis measured via different *items*, using different *stimuli*, or under different *conditions*
- "Clustered" and "cross-classified" data
  - Same individual units of analysis (one or more kinds of groups) measured via different *people* (cross-sectionally or longitudinally)

Options for Modeling Dependency

- Many sampling designs have one or more sources/types of **dependency**, or **correlation of observations from same unit**
- Three main ways of building dependency into a model:
  - **Fixed effects in the model for the means**: add ID variable as a categorical predictor to represent differences across upper-level units
    - Main effects of ID represent intercept dependency; interactions of ID with lower-level predictors represent predictor-specific dependency
    - Does not allow prediction of *why* those differences occurred
  - **Multivariate variance–covariance structures:** for balanced longitudinal or repeated measures data; those using lags also require equal intervals
    - e.g., VC(H), CS(H), AR1(H), TOEP(H), or Unstructured (UN) as "answer key"
    - Can create a pattern of non-constant variance and covariance over time/RM
  - **Add one level (or more):** add random intercept (and slope) variances
    - Can create multiple patterns of non-constant variance and covariance even with unbalanced data (longitudinal or clustered) → **LET'S REVIEW THIS…**

Two-Level Longitudinal Data

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

- Within-Person (WP) Variation **over Time**:
  - **Level 1** – "**INTRA**-individual Differences" – Time-Varying
  - Only longitudinal studies can provide this extra 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 one's average

- I use "person" here, but level 2 can be any entity that is measured repeatedly (like animals, schools, houses, countries…)

Characterizing Longitudinal Data

- What should "**time**" be?
  - e.g., time in study, age, grade, time from event/diagnosis

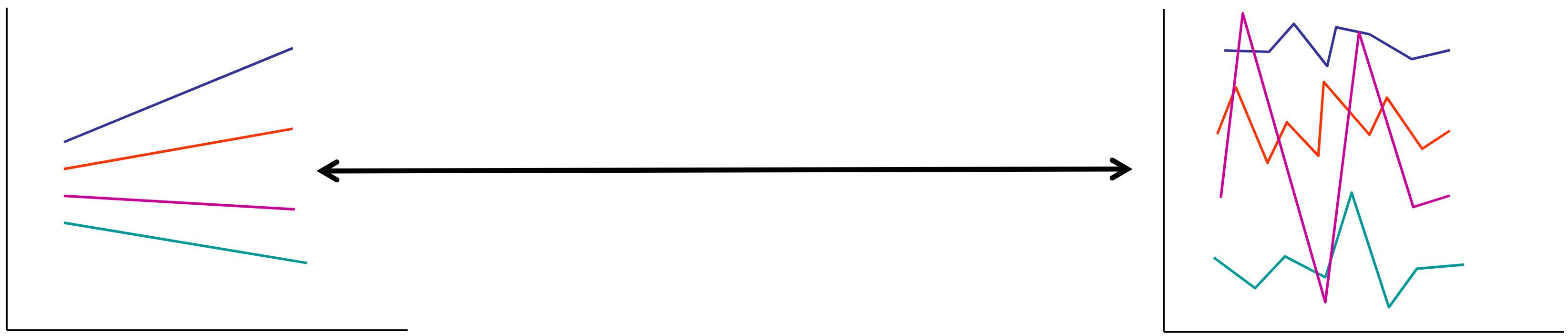
- Does time vary both **within-** AND **between-persons**?
  - Model will need to differentiate each level of time effect
  - Often known as "accelerated" longitudinal designs

- Is time balanced or unbalanced?
  - **Balanced** = everyone has a shared measurement schedule
    - Some people may miss occasions, making their data "incomplete"
  - **Unbalanced** = people have different possible "time" values
    - By definition, observations are "incomplete" across persons
    - This is a consequence of any time metric varying between persons

Characterizing Longitudinal Data

Pure WP Change

Pure WP Fluctuation



Time

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 → describe pattern of variance and covariance over time

ANOVA for two-level longitudinal data?

- There are 3 possible "kinds" of ANOVAs we could use:
  - Between-Persons/Groups, Univariate RM, and Multivariate RM
- **NONE OF THEM ALLOW:**
  - **Missing occasions** (do listwise deletion when using least squares)
  - **Time-varying predictors** (covariates are BP predictors only)
- Each includes the same model for the means for time: all possible mean differences (so 4 parameters to get to 4 means)
  - **"Saturated means" model for *Time*:** $\beta_0 + \beta_1(T_1) + \beta_2(T_2) + \beta_3(T_3)$
  - **The *Time* variable must be balanced and discrete in ANOVA!**
- These ANOVAs differ by what they predict for the correlation across outcomes from the same person in the model for the variance…
  - i.e., **how they "handle dependency"** due to persons, or what they says the variance and covariance of the y_{ti} residuals should look like…

Summary: ANOVA models for longitudinal data are like "one size fits most"

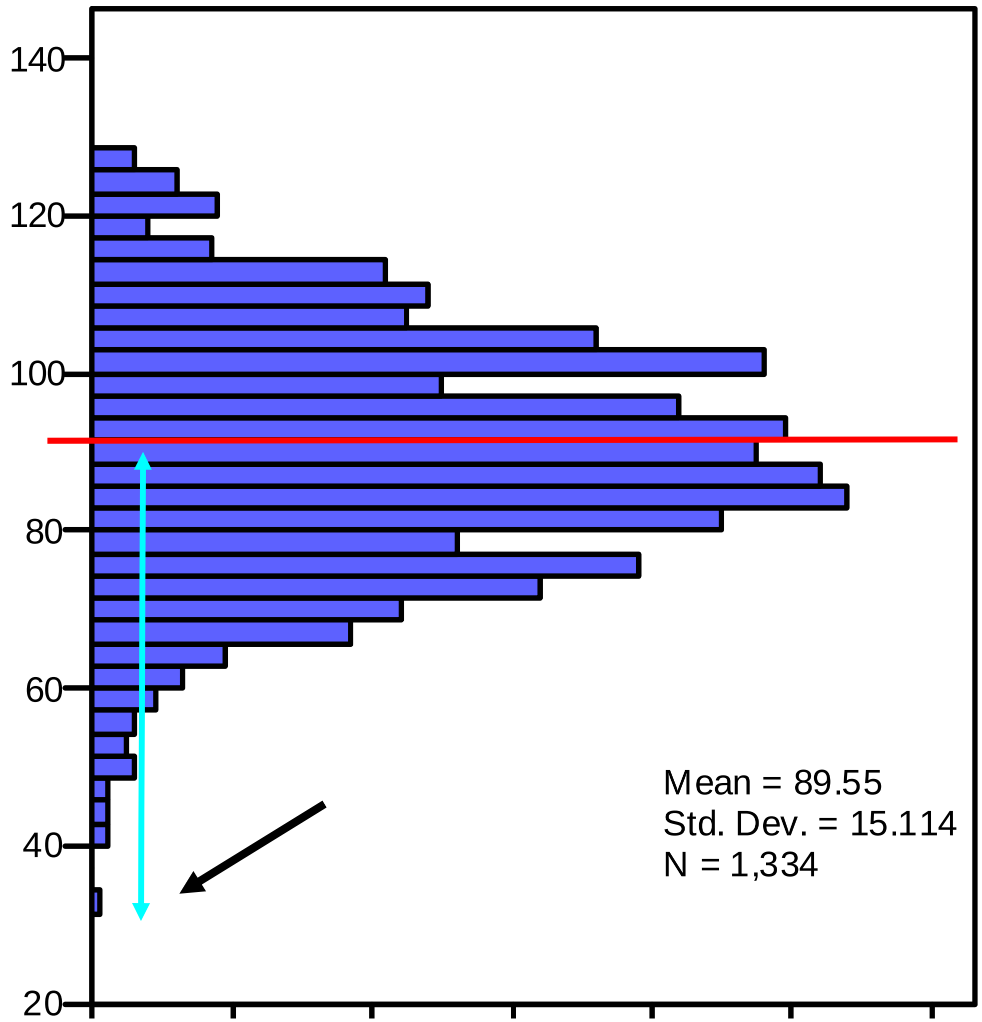
- **Saturated Model for the Means** (balanced time required)
  - All possible mean differences across time
  - Unparsimonious, but best-fitting (is a description of the complete data)

- **3 kinds of Models for the Variance** (need complete data in least squares)
  - BP ANOVA (σ_e^2 only; **VC**) → independence and constant variance over time
  - Univ. RM ANOVA $\left(\tau_{U_0}^2 + \sigma_e^2; \mathbf{CS}\right)$ → constant variance and covariance over time
  - Multiv. RM ANOVA (**unstructured**) → is a description of the (complete) data

there is no structure that shows up in a scalar equation (i.e., the way U_{0i} + e_{ti} does)

- **MLM will give us more flexibility in both parts of the model:**
  - Fixed effects that *predict* the pattern of means over time (polynomials, pieces)
  - Random intercepts and slopes and/or alternative covariance structures that *predict* intermediate patterns of variance and covariance over time

An Empty Between-Person Model (i.e., Single-Level)



$y_i = \beta_0 + e_i$

Filling in values:

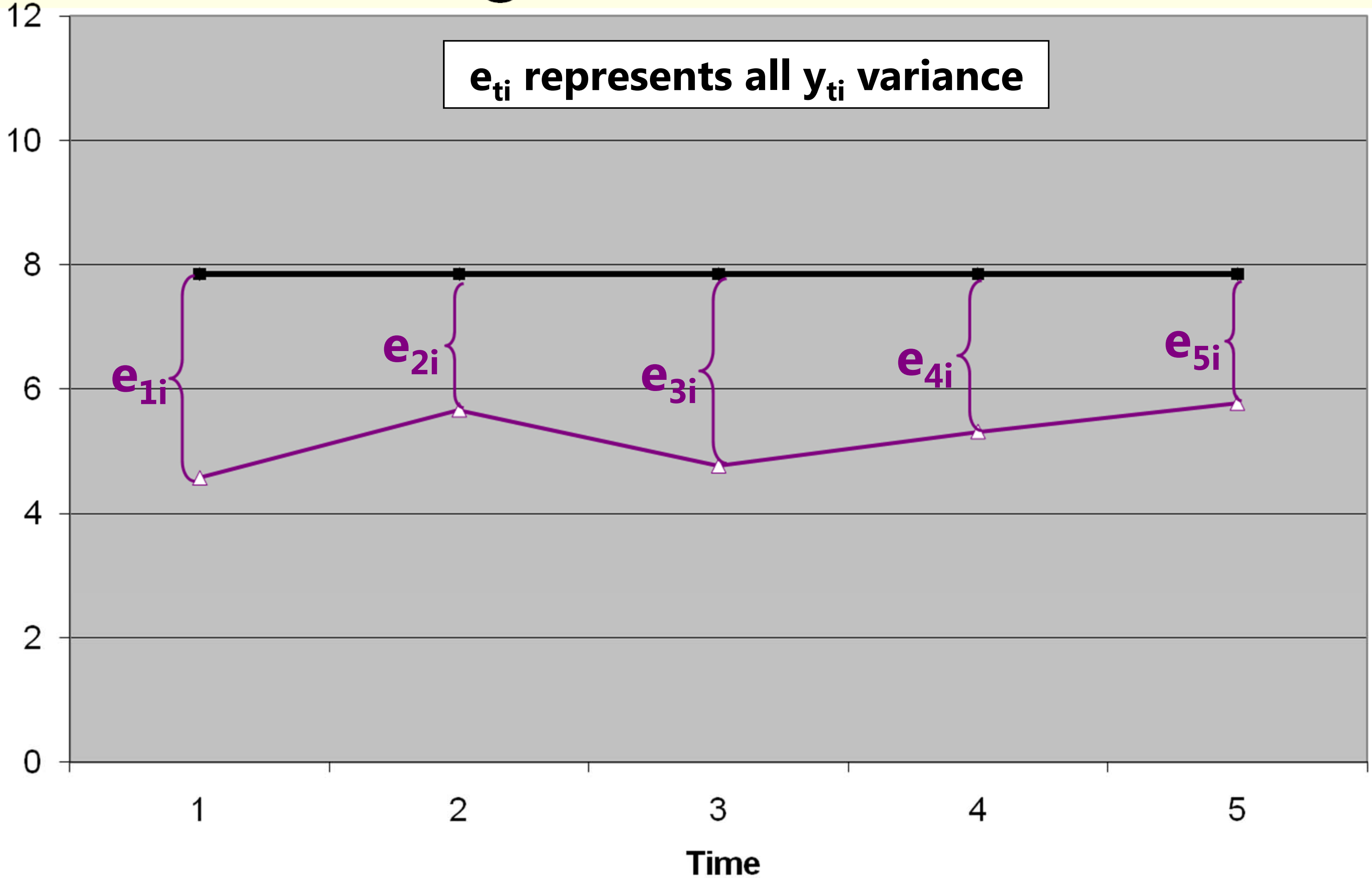
$32 = \underbrace{90}_{y_{pred}} + -58$

Model for the Means

y_i error variance:

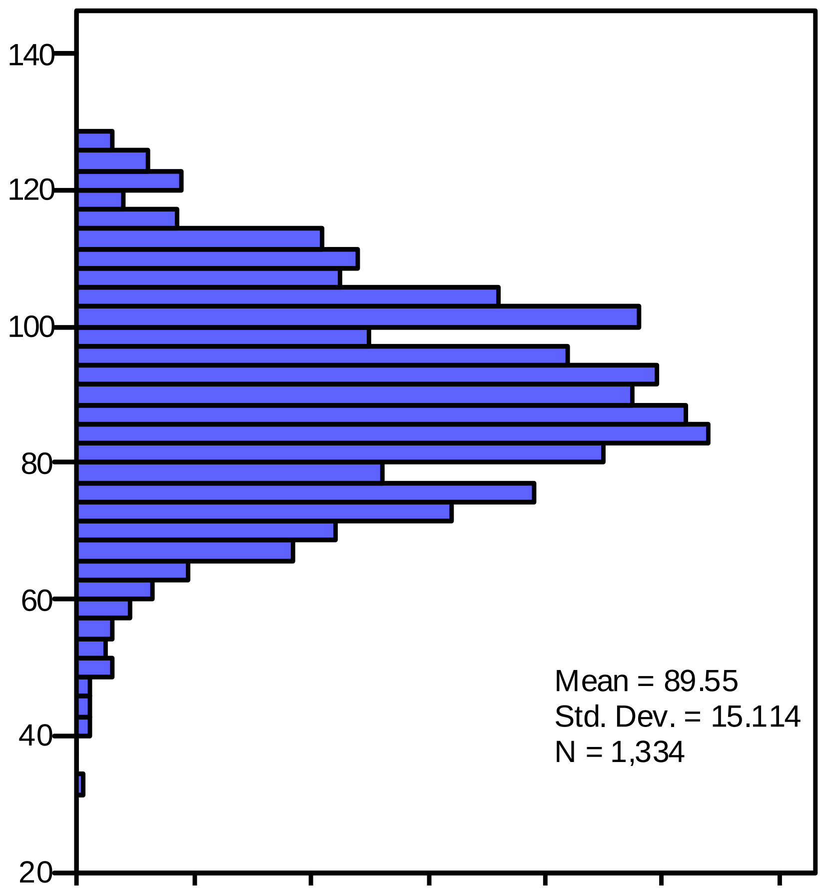
$$\frac{\sum\left(y_i - y_{pred}\right)^2}{N-1}$$

“Error” in a BP Model for the Variance: Single-Level Model

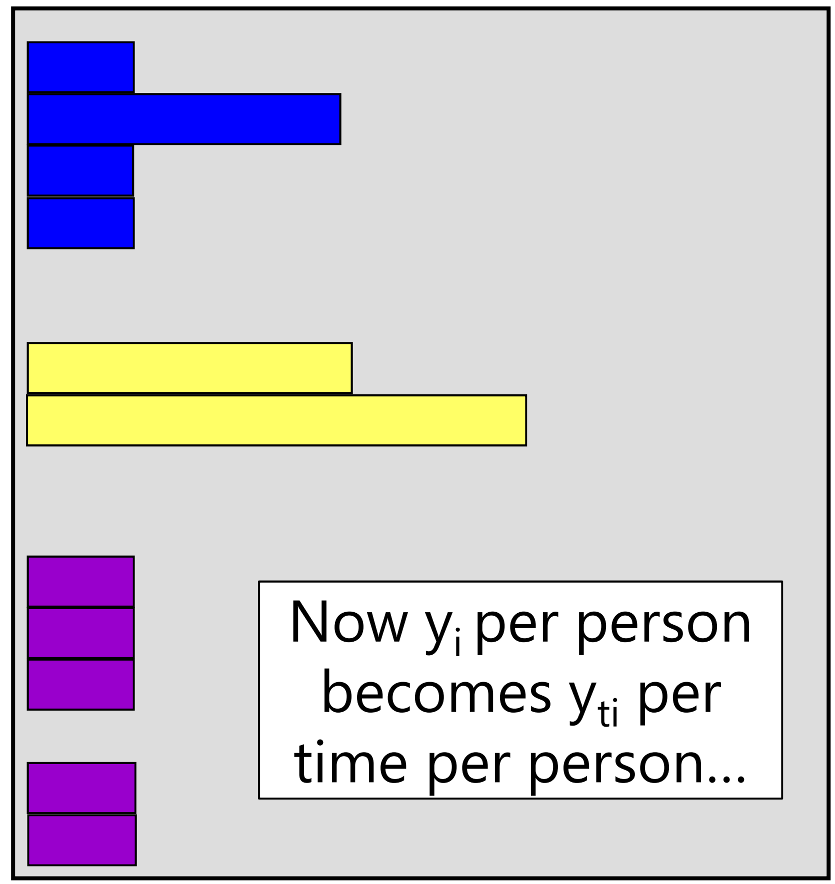


Adding Within-Person Information… (i.e., to become a Two-Level Model)

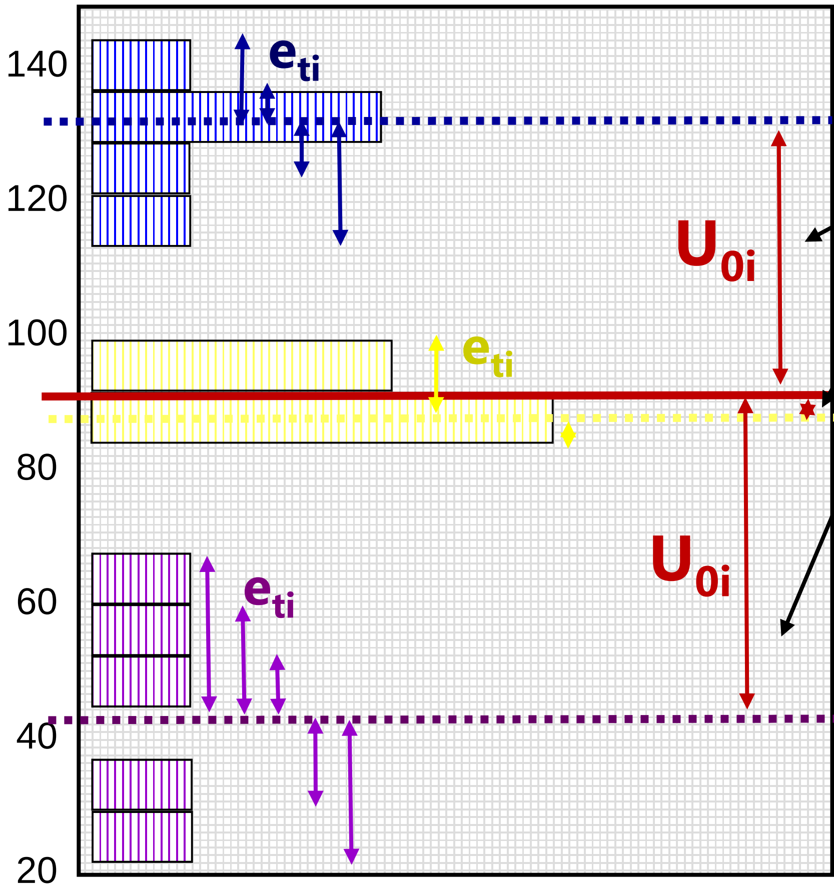
Full Sample Distribution



3 People, 5 Occasions each



Empty +Within-Person Model



y_{ti} variance (V) → 2 sources:

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

→ **Between**-Person Variance in **G**

→ Differences from **GRAND** mean

→ **INTER**-Individual Differences

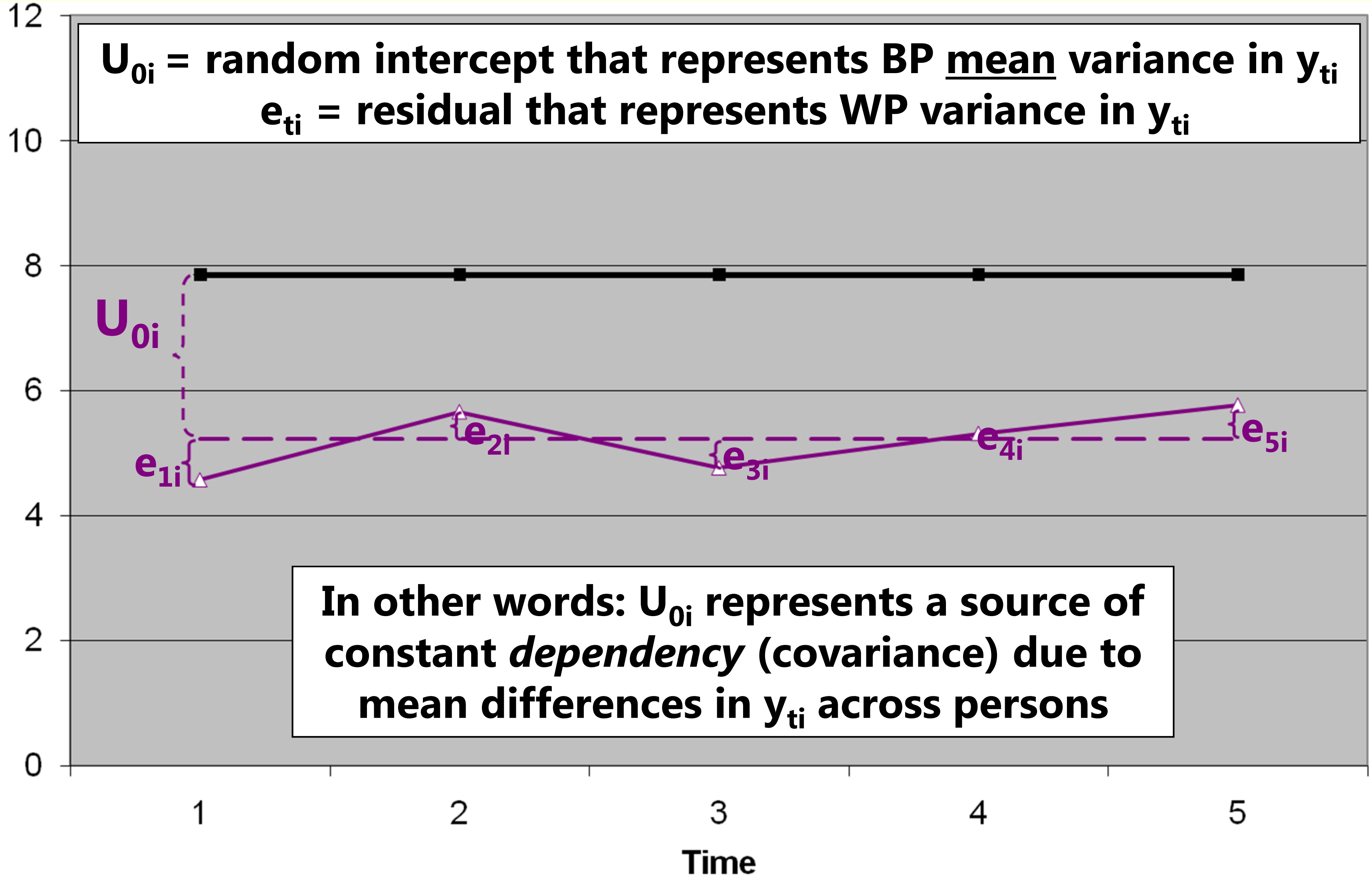
Level-1 Residual Variance (of e_{ti}, as σ^2_e):

→ **Within**-Person Variance in **R**

→ Differences from **OWN** mean

→ **INTRA**-Individual Differences

"Error" in a +WP Model for the Variance: Multilevel Model



BP vs. +WP Empty Models

- Empty **Between-Person** 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 **+Within-Person** Model (for >1 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

Same 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 Variance of e_{ti} → σ_e^2
- Level-2 Variance of U_{0i} → $\tau_{U_0}^2$

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

Fixed Intercept = mean of means (=mean 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):

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

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

V matrix

$$\begin{bmatrix} \sigma_e^2 + \tau_{u_0}^2 & \tau_{u_0}^2 & \tau_{u_0}^2 \\ \tau_{u_0}^2 & \sigma_e^2 + \tau_{u_0}^2 & \tau_{u_0}^2 \\ \tau_{u_0}^2 & \tau_{u_0}^2 & \sigma_e^2 + \tau_{u_0}^2 \end{bmatrix}$$

VCORR Matrix

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

- ICC = Proportion of total variance that is between persons
- ICC = Correlation of occasions from same person (in VCORR)
- 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*

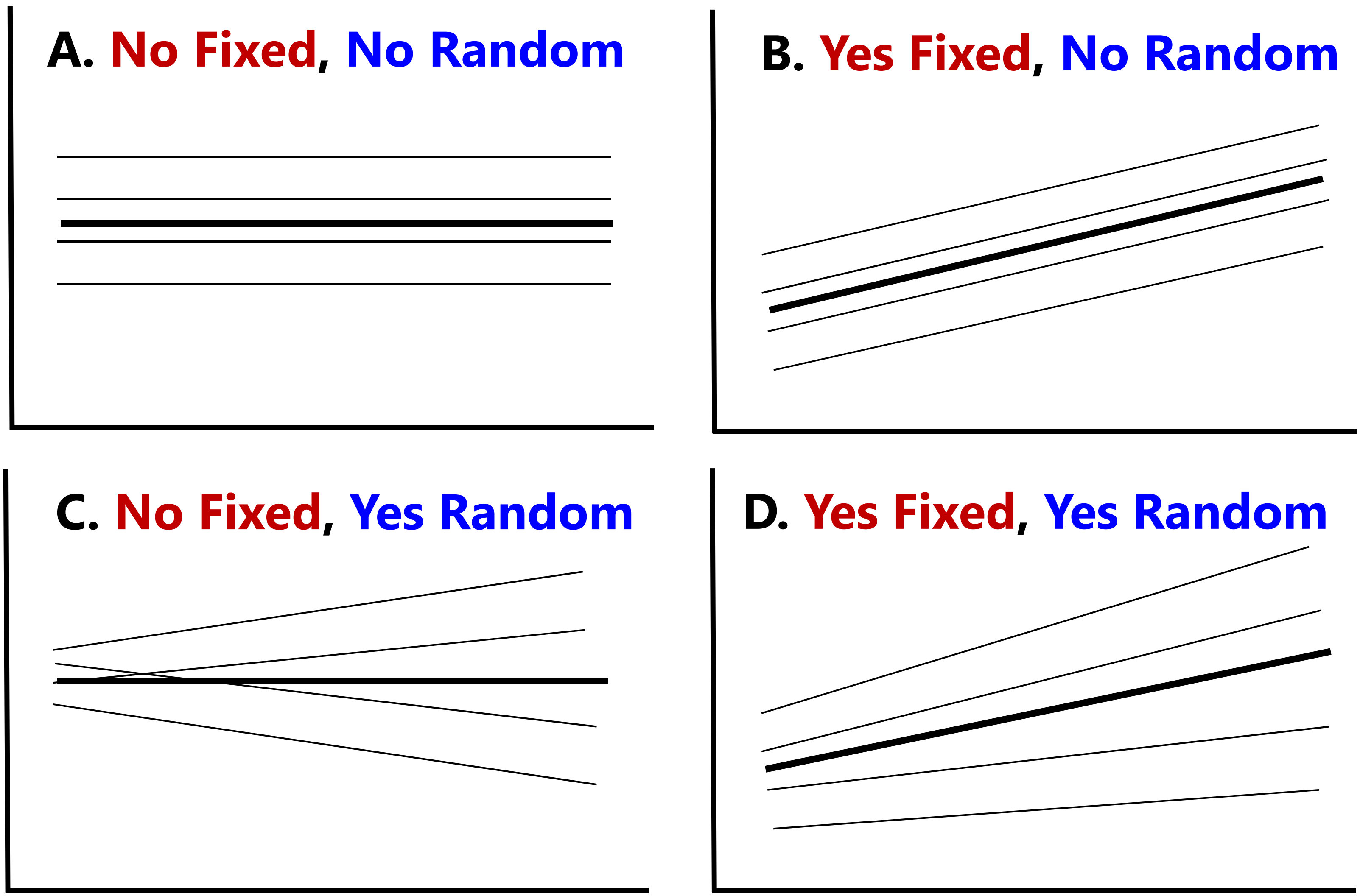
- 2 questions about the possible effects of *time*:

1. **Is there an effect of time on average?**
   - If the line describing the sample means not flat?
   - Significant **FIXED** effect of time

2. **Does the average effect of time vary across individuals?**
   - Does each individual need his or her own line?
   - Significant **RANDOM** effect of time

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

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

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 $\tau_{U_{01}}$

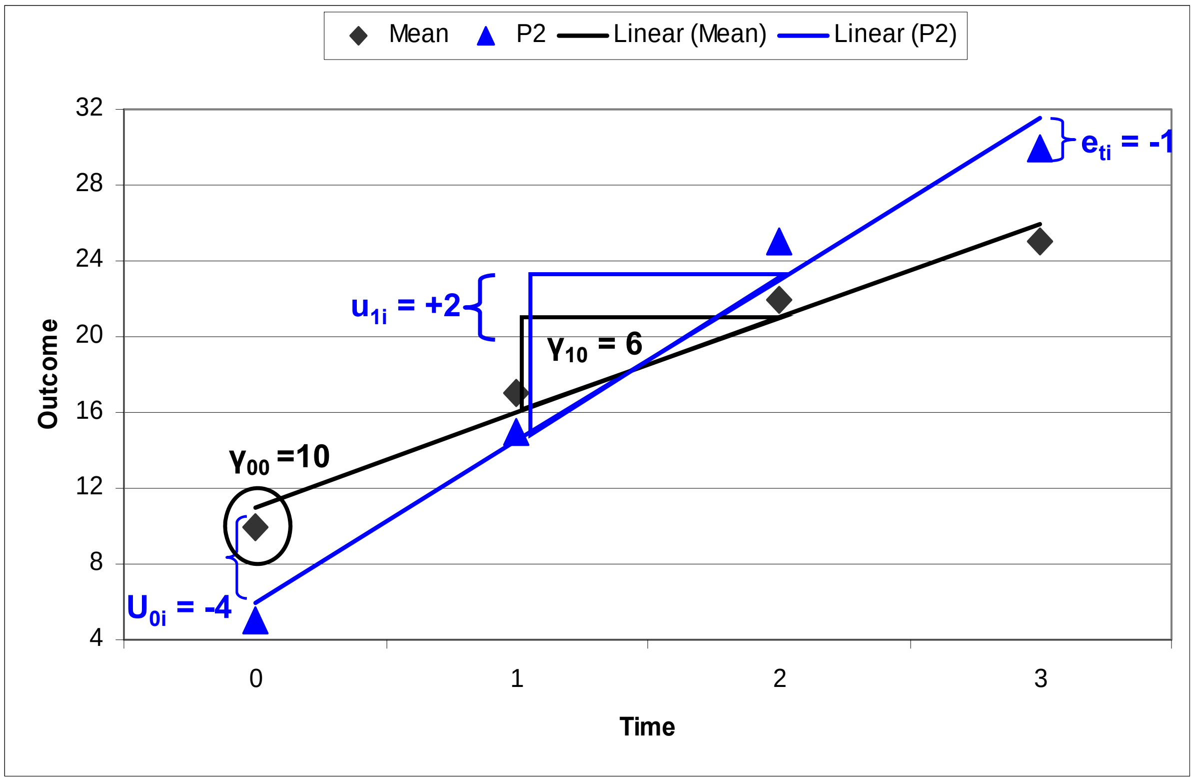
Composite Model

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

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

2 Random Effects Variances:

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

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

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

1 e_{ti} Residual Variance $= \sigma^2_e$

Summary: Sequential Models for Effects of Time

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

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

Composite: $y_{ti} = \gamma_{00} + U_{0i} + e_{ti}$

Empty Means, Random Intercept Model: 3 parms = γ_{00}, σ^2_e, $\tau^2_{U_0}$

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

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

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

Fixed Linear Time, Random Intercept Model: 4 parms = γ_{00}, γ_{10}, σ^2_e, $\tau^2_{U_0}$

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

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

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

Random Linear Time Model: 6 parms = γ_{00}, γ_{10}, σ^2_e, $\tau^2_{U_0}$, $\tau^2_{U_1}$, $\tau_{U_{01}}$ (→ cov of U_{0i} and U_{1i})

Random Linear Time Models Imply:

- **People differ from each other systematically in TWO ways**—in intercept (U_{0i}) and 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 $\mathbf{e_{ti}}$ **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)

What Does Each Side of the Model Need?

- Nested models (i.e., in which one is a subset of the other) can now differ from each other in two important ways

- **Model for the Means** → which predictors and which **fixed effects** of them are included in the model
  - Does not require assessment of relative model fit using LL or −2LL (can use univariate or multivariate Wald tests for this)

- **Model for the Variance** → what the pattern of variance and covariance of residuals from the same unit should be
  - DOES require assessment of relative model fit using LL or −2LL
  - Cannot use the Wald test *p*-values that show up on the output for testing significance of variances because those *p*-values are use a two-sided sampling distribution for what the variance could be (but variances cannot be negative, so those *p*-values are not valid)

Testing Significance of Fixed Effects (of Predictors) in the Model for the Means

- Any single-df **fixed effect** has 4-5 relevant pieces of output:
  - **Estimate** = best guess for the fixed effect from our data
  - **Standard Error** = precision of fixed effect estimate (quality of most likely estimate)
  - ***t*-value or *z*-value** = Estimate / Standard Error
  - ***p*-value** = probability that fixed effect estimate is ≠ 0
  - **95% Confidence Interval** = Estimate ± 1.96*SE = range in which true (population) value of estimate is expected to fall 95% of the time

- Compare test statistic (*t* or *z*) to critical value at chosen level of significance (known as alpha): this is a "**univariate Wald test**"
- Whether the *p*-value is based on *t* or *z* varies by program…

Evaluating Significance of 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 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 (Mplus, STATA)	use ***t*** distribution (SAS, SPSS)
Numerator DF > 1 *(test 2+ fixed effects)* is **Multivariate Wald Test**	use χ^2 distribution (Mplus, STATA)	use ***F*** distribution (SAS, SPSS)
Denominator DF options	not applicable, so DDF is not given	SAS, STATA 14: BW, KR SAS, STATA 14, SPSS: Satterthwaite

Evaluating Effect Size of Fixed Effects

- Most common measure of effect size in MLM is Pseudo-R^2
  - Is supposed to be variance accounted for by predictors
  - Multiple piles of variance mean multiple possible values of pseudo R^2 (can be calculated per variance component or per model level)
  - A fixed linear effect of time will reduce level-1 residual variance σ_e^2 in **R**
  - By how much is the residual variance σ_e^2 reduced?

$$\text{Pseudo } R_e^2 = \frac{\text{residual variance}_{\text{fewer}} - \text{residual variance}_{\text{more}}}{\text{residual variance}_{\text{fewer}}}$$

  - If time varies between persons, then level-2 random intercept variance $\tau_{U_0}^2$ in **G** may also be reduced:

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

  - But you are likely to see a (net) INCREASE in $\tau_{U_0}^2$ instead…. Here's why:

Increases in Random Intercept Variance

- Level-2 random intercept variance $\tau^2_{U_0}$ will often increase as a consequence of reducing level-1 residual variance σ^2_e
- Observed level-2 $\tau^2_{U_0}$ is NOT just between-person variance
  - Also has a small part of within-person variance (level-1 σ^2_e), or: **Observed $\boldsymbol{\tau^2_{U_0}}$ = True $\boldsymbol{\tau^2_{U_0}}$ + ($\boldsymbol{\sigma^2_e/n}$)**
    - As *n* occasions increases, bias of level-1 σ^2_e is minimized
  - Likelihood-based estimates of "true" $\tau^2_{U_0}$ use (σ^2_e/n) as correction factor: **True $\boldsymbol{\tau^2_{U_0}}$ = Observed $\boldsymbol{\tau^2_{U_0}}$ – ($\boldsymbol{\sigma^2_e/n}$)**
- For example: observed level-2 $\tau^2_{U_0}$=4.65, level-1 σ^2_e=7.06, *n*=4
  - True $\tau^2_{U_0}$= 4.65 −(**7.60**/4) = **2.88** in empty means model
  - Add fixed linear time slope → reduce σ^2_e from 7.06 to 2.17 (R^2 = .69)
  - But now True $\tau^2_{U_0}$= 4.65 −(**2.17**/4) = **4.10** in fixed linear time model

Significance Tests for Choosing Models for the Variance

- Requires assessment of **relative model fit**: how well does the model fit relative to other possible models?
  - Assessment of *absolute* model fit is only possible for balanced data

- Relative fit is indexed by overall model **log-likelihood** (**LL**):
  - Log of likelihood for each person's outcomes given model parameters
  - Sum log-likelihoods across all independent persons = **model LL**
  - Two flavors: Maximum Likelihood (ML) or Restricted ML (REML)

- What you get for this on your output varies by software…

- Given as −2*log likelihood (−2LL) in SAS or SPSS MIXED:
  −2LL gives BADNESS of fit, so **smaller** value = better model

- Given as just log-likelihood (LL) in STATA MIXED and Mplus:
  LL gives GOODNESS of fit, so **bigger** value = better model

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 =** $(\# \text{ Parms}_{more}) - (\# \text{ Parms}_{fewer})$
3. **Compare −2ΔLL to χ^2 distribution with df = Δdf**
4. Get *p*-value from CHIDIST in excel or LRTEST option in STATA

Comparing Models for the Variance

- <u>What your *p*-value for the −2ΔLL test means:</u>
  - 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

Flavors of Maximum Likelihood

- Remember that Maximum likelihood comes in two flavors:
- **"Restricted (or residual) maximum likelihood"**
  - Only available for general linear models or general linear mixed models (that assume normally distributed residuals)
  - Is same as LS given complete outcomes, but it doesn't require them
  - Estimates variances the same way as in LS (accurate) → $\frac{\sum(y_i - y_{pred})^2}{N - k}$
- **"Maximum likelihood" (ML; also called FIML*)**
  - Is more general, is available for the above plus for non-normal outcomes and latent variable models (CFA/SEM/IRT)
  - Is NOT the same as LS: it under-estimates variances by not accounting for the # of estimated fixed effects → $\frac{\sum(y_i - y_{pred})^2}{N}$
- **FI = Full information → it uses all original data (they both do)*

Flavors of Full-Information Maximum Likelihood

- Restricted maximum likelihood (**REML**; used in MIXED)
  - Provides unbiased variances
  - Especially important for small N (< 100 units)
  - **−2ΔLL test** cannot be used to compare models differing in fixed effects (no biggee; we can do this using univariate or multivariate Wald tests)
  - **−2ΔLL test** MUST be used to compare different models for the variance

$$\frac{\sum\left(y_i - y_{pred}\right)^2}{N - k}$$

- Maximum likelihood (**ML**; also used in MIXED)
  - Variances (and SEs) are too small in small samples
  - Is only option in most software for path models and SEM
  - **−2ΔLL test** can be used to compare **any** nested model; must be used to compare different models for the variance

$$\frac{\sum\left(y_i - y_{pred}\right)^2}{N}$$

ML vs. REML in a nutshell

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

"Population"

$$\frac{\sum(y_i - y_{pred})^2}{N}$$

"Sample"

$$\frac{\sum(y_i - y_{pred})^2}{N - k}$$

All comparisons must have same N!!!	**ML**	**REML**
To select, type…	METHOD=ML (-2 log likelihood)	METHOD=REML *default* (-2 res log likelihood)
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)

Rules for Comparing Models

All observations must be the same across models!

Compare Models Differing In:

Type of Comparison:	**Means Model (Fixed) Only**	**Variance Model (Random) Only**	**Both Means and Variances Model (Fixed and Random)**
Nested? YES, can do significance tests via…	Fixed effect *p*-values from ML or REML -- *OR* -- ML −2ΔLL only (NO REML −2ΔLL)	NO *p*-values REML −2ΔLL (ML −2ΔLL is ok if big N)	ML −2ΔLL only (NO REML −2ΔLL)
Non-Nested? NO signif. tests, instead see…	ML AIC, BIC (NO REML AIC, BIC)	REML AIC, BIC (ML ok if big N)	ML AIC, BIC only (NO REML AIC, BIC)

Nested = one model is a direct subset of the other

Non-Nested = one model is not a direct subset of the other

3 Decision Points for Model Comparisons

1. Are the models **nested** or **non-nested**?
   - Nested: have to add OR subtract effects to go from one to other
     - Can conduct significance tests for improvement in fit
   - Non-nested: have to add AND subtract effects
     - No significance tests available for these comparisons

2. Differ in model for the **means**, **variances**, or **both**?
   - Means? Can only use −2ΔLL tests if ML (or *p*-value of each fixed effect)
   - Variances? Can use ML (or preferably REML) −2ΔLL tests, no *p*-values
   - Both sides? Can only use −2ΔLL tests if ML

3. Models estimated using **ML** or **REML**?
   - ML: All model comparisons are ok
   - REML: Model comparisons are ok for the variance parameters only

Effect Size for Random Effects Variances

- We can test if a random effect variance is significant, but the variance estimates are not likely to have inherent meaning
  - e.g., "I have a significant fixed linear time effect of γ_{10} **= 1.72**, so people increase by 1.72/time on average. I also have a significant random linear time slope variance of $\tau^2_{U_1}$ **= 0.91**, so people need their own slopes (people change differently). But how much is a variance of **0.91**, really?"
- **95% Random Effects Confidence Intervals** can tell you
  - Can be calculated for each effect that is random in your model
  - Provide range around the fixed effect within which 95% of your sample is predicted to fall, based on your random effect variance:

$$\text{Random Effect 95\% CI} = \text{fixed effect} \pm \left(1.96 * \sqrt{\text{Random Variance}}\right)$$

$$\text{Linear Time Slope 95\% CI} = \gamma_{10} \pm \left(1.96 * \sqrt{\tau^2_{U_1}}\right) \rightarrow 1.72 \pm \left(1.96 * \sqrt{0.91}\right) = -0.15 \text{ to } 3.59$$

  - So although people improve on average, individual slopes are predicted to range from −0.15 to 3.59 (so some people may actually decline)

Another Variance Model Effect Size: Intercept/Slope Reliability

- Another measure of effect size for random effects variances is **Intercept Reliability (IR) or Slope Reliability (SR)**

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

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

- IR formula is the same, just replacing $\boldsymbol{\sigma^2_{L1}}$ with 1
- SR is known as growth rate reliability in context of time (Willett, 1989)

The Big Picture of Longitudinal Data: **Models for the Means for Time**

- What kind of change occurs on average over "time"? There are two baseline models to consider:
  - "**Empty**" → only a fixed intercept (predicts no change)
  - "**Saturated**" → all occasion mean differences from time 0 (ANOVA model that uses # fixed effects= *n*) *** *may not be possible in unbalanced data*

Parsimony ← → Good fit

Empty Model:
Predicts NO change over time
1 Fixed Effect

In-between options:
polynomial slopes, piecewise slopes, nonlinear slopes…

Saturated Means:
Reproduces mean at each occasion
 # Fixed Effects = # Occasions

Name… that… Trajectory!

The Big Picture of Longitudinal Data:
Models for the Variance for Time

Compound Symmetry (CS)

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

Most useful model: likely somewhere in between!

Unstructured (UN)

$$\begin{bmatrix} \sigma^2_{T1} & \sigma_{T12} & \sigma_{T13} & \sigma_{T14} \\ \sigma_{T21} & \sigma^2_{T2} & \sigma_{T23} & \sigma_{T24} \\ \sigma_{T31} & \sigma_{T32} & \sigma^2_{T3} & \sigma_{T43} \\ \sigma_{T41} & \sigma_{T42} & \sigma_{T43} & \sigma^2_{T4} \end{bmatrix}$$

Parsimony ⟵⟶ **Good fit**

Univariate RM ANOVA

Name …that … Structure!

Multivariate RM ANOVA

What is the pattern of variance and covariance over time?

CS and UN are just two of the many, many options available within MLM, including ***random effects models*** (for change) and ***alternative covariance structure models*** (for fluctuation).

Summary: Unconditional Longitudinal Models

Model for the Means for Time:

- What kind of **fixed effects of time** are needed to create a function with which to parsimoniously **represent the pattern of saturated means** across time?
  - Continuous or discontinuous? This choice likely comes from the design!
  - Polynomials? Pieces? Log time? Truly nonlinear? This comes from the means plot!
  - Use obtained *p*-values to test significance of fixed effects (Wald test)
  - Use pseudo-R^2 values to describe effect size (just for residual if time is WP only)

Model for the Variance for Time (building V):

- What kind of **random effects of time** in G are needed :
  - To account for **individual differences in each aspect of change**?
  - To describe any non-constant variance and covariance across occasions?
  - Do the residuals in R show any covariance after accounting for random effects?
  - Use REML −2ΔLL tests to test significance of new effects (or ML if big upper-level *N*)
  - Use random effects CIs and intercept/slope reliability to describe effect size

Summary of Unconditional Time Models

- 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 ($\tau_{U_1}^2$)

Multiple BP time slope variances are possible…

$\tau_{U_{01}}$ covariance

Next we will add predictors to account for each pile!