Introduction to Multilevel Longitudinal Models

- Topics:
  - **What is multilevel modeling?**
  - **Concepts in longitudinal data**
  - From between-person to within-person models
  - Example 1: Introduction to multilevel modeling software
  - Kinds of ANOVAs for longitudinal data

What is a Multilevel Model (MLM)?

- <u>Same as other terms you have heard of:</u>
  - **General(ized) Linear Mixed Model** (if you are from statistics)
    - *Mixed* = Fixed and Random effects; "ized" =not-normal residuals
  - **Random Coefficients Model** (also if you are from statistics)
    - Random coefficients = Random effects = latent variables/factors
  - **Hierarchical Linear Model** (if you are from education)
    - *Not* the same thing as hierarchical (step-wise) regression

- <u>Special cases of MLM:</u>
  - 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 are 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)*

What can MLM do for you?

1. **Model dependency across observations**
   - Longitudinal, clustered, and/or cross-classified data? No problem!
   - Tailor your model of sources of correlation to your data

2. **Include categorical or continuous predictors at any level**
   - Time-varying, person-level, group-level predictors for each variance
   - Explore reasons for dependency, don't just control for dependency

3. **Does not require same data structure for each person**
   - Unbalanced or missing data? No problem!

4. **You already know how (or you will soon)!**
   - Use SPSS Mixed, **SAS Mixed**, Stata, Mplus, R, HLM, MlwiN…
   - What's an intercept? What's a slope? What's a pile of variance?

1. Model Dependency

- Sources of dependency depend on the sources of **variation** created by your sampling design: residuals for outcomes from the same unit are likely to be related, which violates the GLM "independence" assumption

- **"Levels" for dependency** ="levels of random effects"
  - Sampling dimensions can be **nested**
    - e.g., time within person, time within group, trial within person
  - If you can't figure out the direction of your nesting structure, odds are good you have a **crossed sampling design** instead
    - e.g., persons crossed with items, raters crossed with targets
  - To have a "level", there must be random outcome variation due to sampling that **remains** after including the model's fixed effects
    - e.g., treatment vs. control does not create another level of "group" (but it would if you had multiple treatment and multiple control groups)

Longitudinal Dependency comes from…

- Mean differences across sampling units (e.g., persons)
  - Creates <u>constant</u> dependency over time
  - Will be represented by a random intercept in our models

- Individual differences in effects of predictors
  - Individual differences in change over time, stress reactivity
  - Creates <u>non-constant</u> dependency, the size of which depends on the value of the predictor at each occasion or for each person
  - Will be represented by random slopes in our models

- Non-constant within-person correlation for unknown reasons (time-specific autocorrelation)
  - Can add other patterns of correlation as needed for this (AR, TOEP)

Why care about dependency?

- In other words, what happens if we have the wrong model for the variance (assume independence instead)?

- **Validity of the tests of the predictors** depends on having the “most right” model for the variance
  - Estimates will usually be ok → come from model for the means
  - Standard errors (and thus *p*-values) can be inaccurate

- The sources of variation that exist in your outcome will dictate **what kinds of predictors** will be useful
  - Between-Person variation needs Time-Invariant predictors
  - Within-Person variation needs Time-Varying predictors
  - Between-Item variation needs Item-Specific predictors

2. Include categorical or continuous predictors at any level of analysis

- "ANOVA" test differences among discrete groups
- "Regression" tests slopes for continuous predictors
- What if a predictor is assessed repeatedly but can't be characterized by discrete "conditions"?
  - ANOVA or Regression won't work → you need MLM

- Some things don't change over time → time-invariant
- Some things do change over time → time-varying
- Some things are measured at higher levels
- Interactions are possible at same level or across levels

3. Does not require same data structure per person (by accident or by design)

RM ANOVA: uses **multivariate** (wide) data structure:

ID	Sex	T1	T2	T3	T4
100	0	5	6	8	12
101	1	4	7	.	11

People missing any data are excluded (data from ID 101 are not included at all)

MLM: uses **stacked** (long) data structure:

Only rows missing data are excluded

ID 100 uses 4 cases

ID 101 uses 3 cases

ID	Sex	Time	Y
100	0	1	5
100	0	2	6
100	0	3	8
100	0	4	12
101	1	1	4
101	1	2	7
101	1	3	.
101	1	4	11

Time can also be **unbalanced** across people such that each person can have his or her own measurement schedule: Time “0.9” “1.4” “3.5” “4.2”…

4. You already know how!

- If you can do GLM, you can do MLM
  (*and if you can do generalized linear models, you can do generalized multilevel models, too*)

- How do you interpret an estimate for…
  - the intercept?
  - the effect of a continuous variable?
  - the effect of a categorical variable?
  - a variance component ("pile of variance")?

Options for Longitudinal Models

- Models and software are separate entities. Longitudinal data can be analyzed via multiple analytic frameworks:
  - **"Multilevel/Mixed Models"**
    - Dependency over time, persons, groups, etc. are modeled via random effects (multivariate → univariate through "levels" of stacked/long data)
    - Builds on GLM, generalizes more easily to additional levels of nesting and/or crossed dimensions of sampling
  - **"Structural Equation Models"**
    - Dependency over time *only* is modeled via latent variables (single-level analysis using multivariate/wide data)
    - Generalizes easier to broader analysis of latent constructs, mediation, and multivariate multilevel models in general (aka, "Multilevel SEM")
  - Because random effects and latent variables are the same thing, many longitudinal models can be specified/estimated either way

Requirements for Longitudinal Data

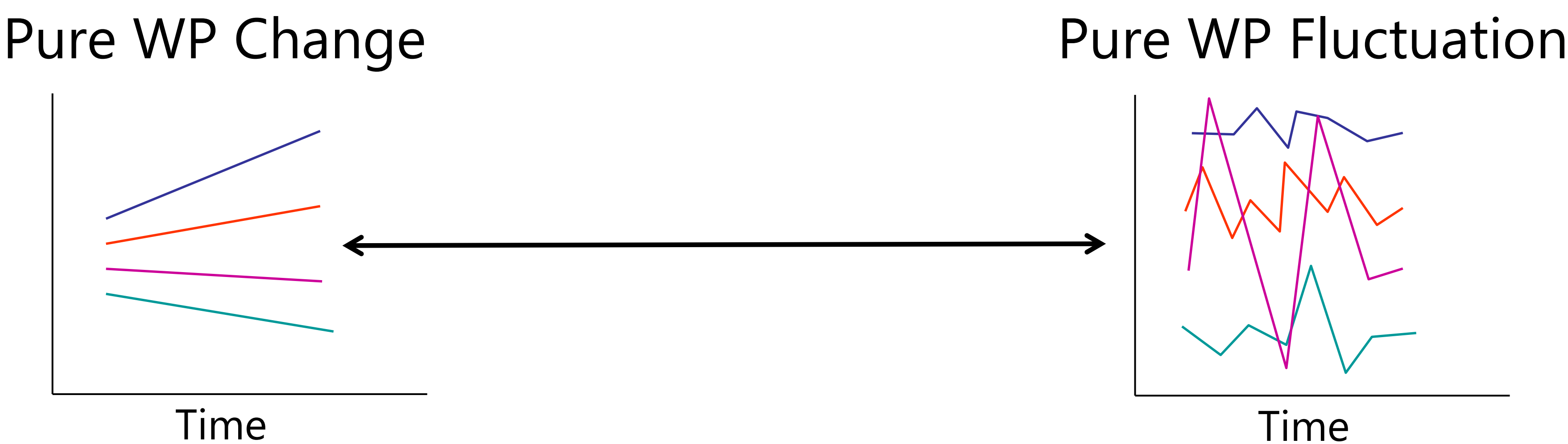
- Multiple <u>OUTCOMES</u> from the same sampling unit!
  - 2 is the minimum, but just 2 can lead to problems:
    - Only 1 kind of change is observable (1 difference)
    - Can't distinguish "real" individual differences in change from "error"
    - Repeated measures ANOVA is just fine for 2 observations
      - Necessary assumption of "sphericity" is satisfied with only 2 observations even if compound symmetry doesn't hold
  - More data is better (with diminishing returns)
    - More occasions → better description of the form of change
    - More persons → better estimates of amount of individual differences in change; better prediction of those individual differences
    - More items/stimuli/groups → more power to show effects of differences between items/stimuli/groups

Levels of Analysis in 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:
  - **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 units can be anything that is measured repeatedly (like animals, schools, countries…)

A Longitudinal Data Continuum

- **Within-Person Change**: Systematic change
  - Magnitude or direction of change can be different across individuals
  - "Growth curve models" → Time is meaningfully sampled
- **Within-Person Fluctuation**: No systematic change
  - Outcome just varies/fluctuates over time (e.g., emotion, stress)
  - Time is just a way to get lots of data per individual



Introduction to Multilevel Longitudinal Models

- Topics:
  - What is multilevel modeling?
  - Concepts in longitudinal data
  - **From between-person to within-person models**
  - Example 1: Introduction to multilevel modeling software
  - Kinds of ANOVAs for longitudinal data

The Two Sides of a (BP) Model

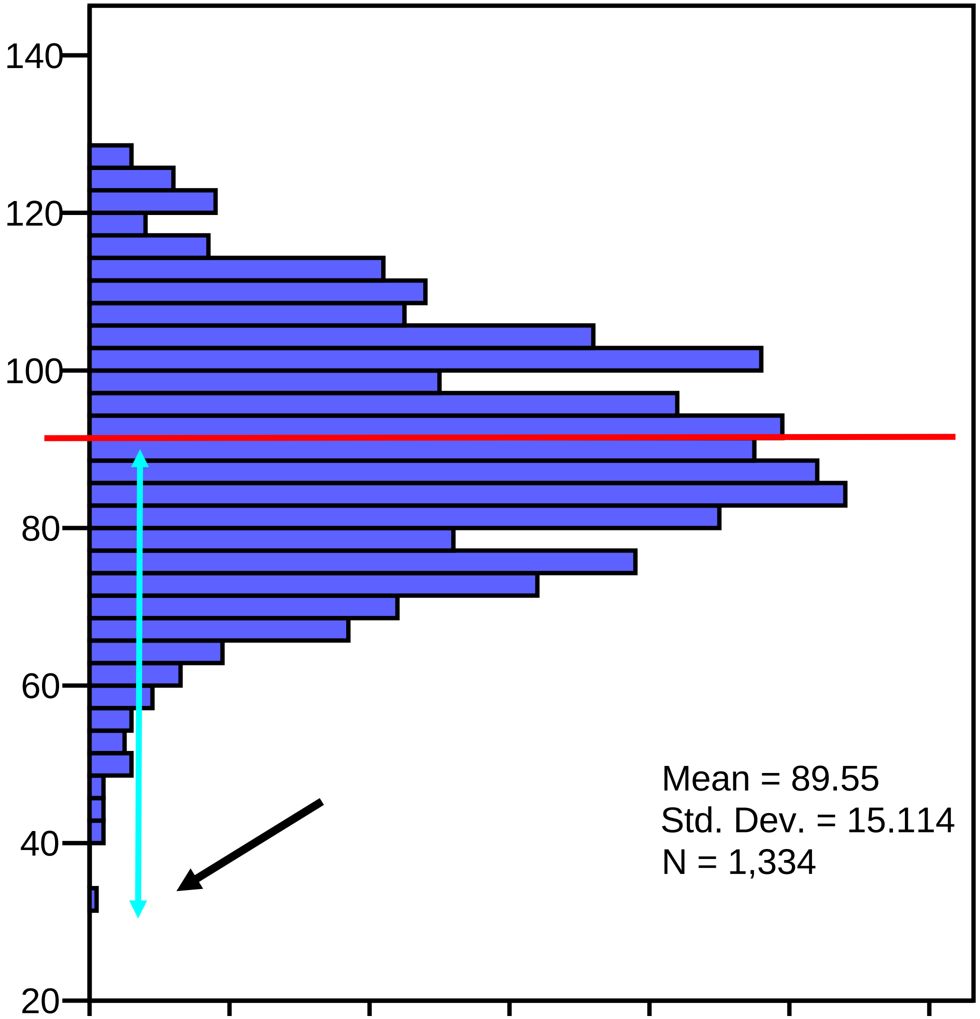
$$y_i = \boxed{\beta_0 + \beta_1 X_i + \beta_2 Z_i + \beta_3 X_i Z_i} + \boxed{e_i}$$

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 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
  - e_i has a mean of 0 with some estimated constant variance σ_e^2, is normally distributed, is unrelated to X and Z, and is unrelated across people (across all observations, just people here)
  - **Estimated parameter is residual variance only in above BP model**

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



$$\mathbf{y_i} = \color{red}{\beta_0} + \color{blue}{e_i}$$

Filling in values:

$$\mathbf{32} = \underbrace{\color{red}{\mathbf{90}}}_{y_{pred}} + \color{blue}{\mathbf{-58}}$$

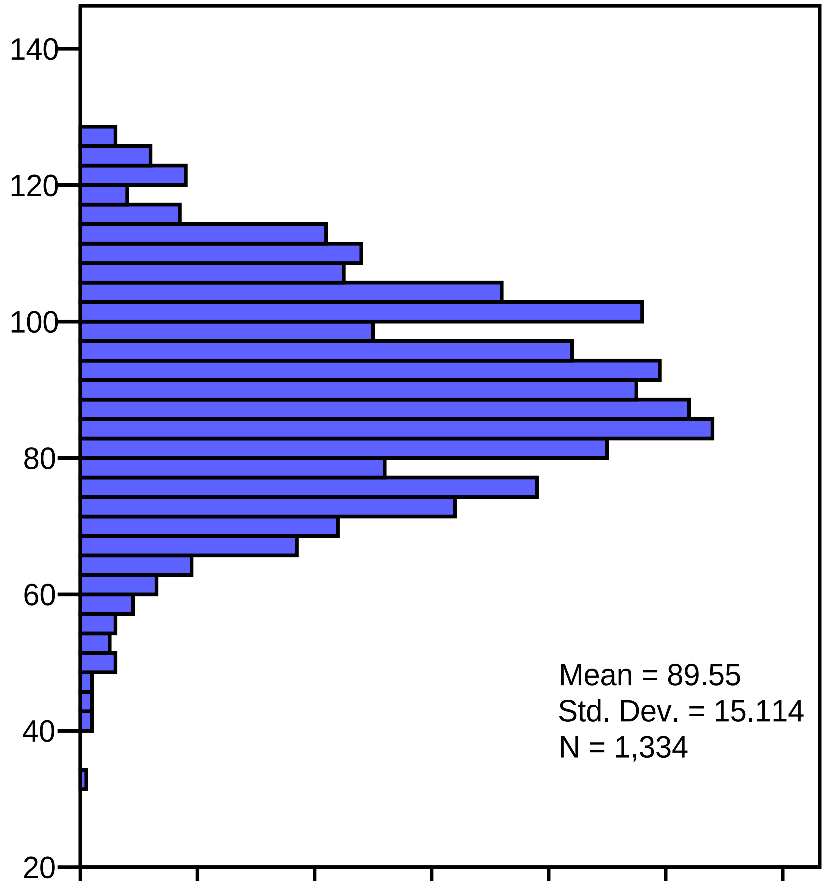
Model for the Means

y_i error variance:

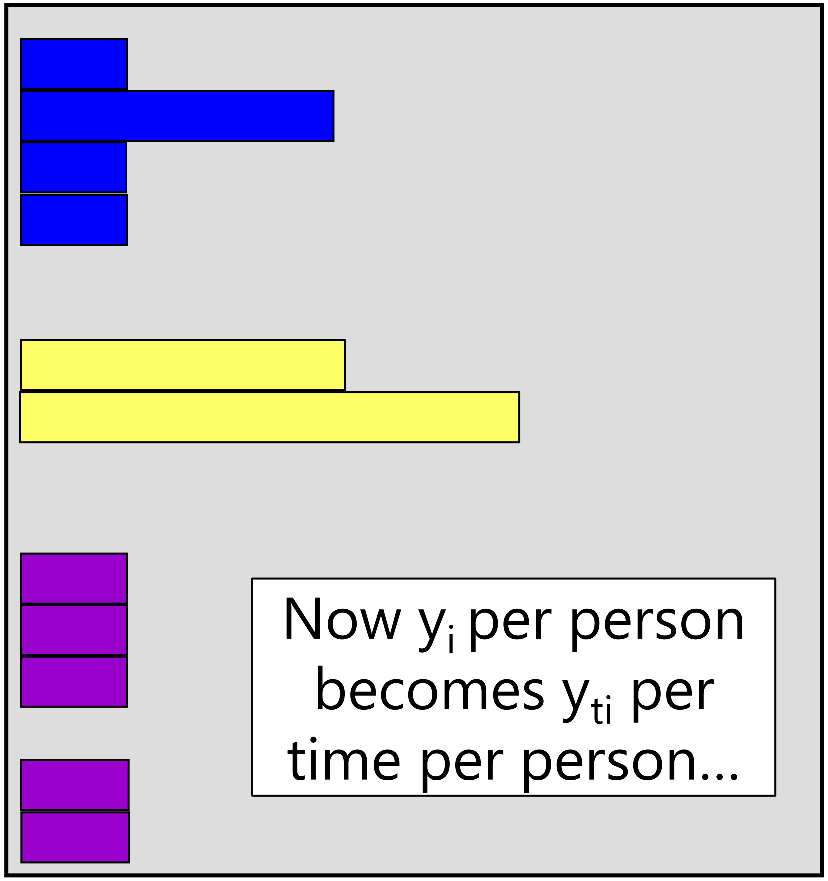
$$\frac{\sum\left(y_i - y_{pred}\right)^2}{N-1}$$

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

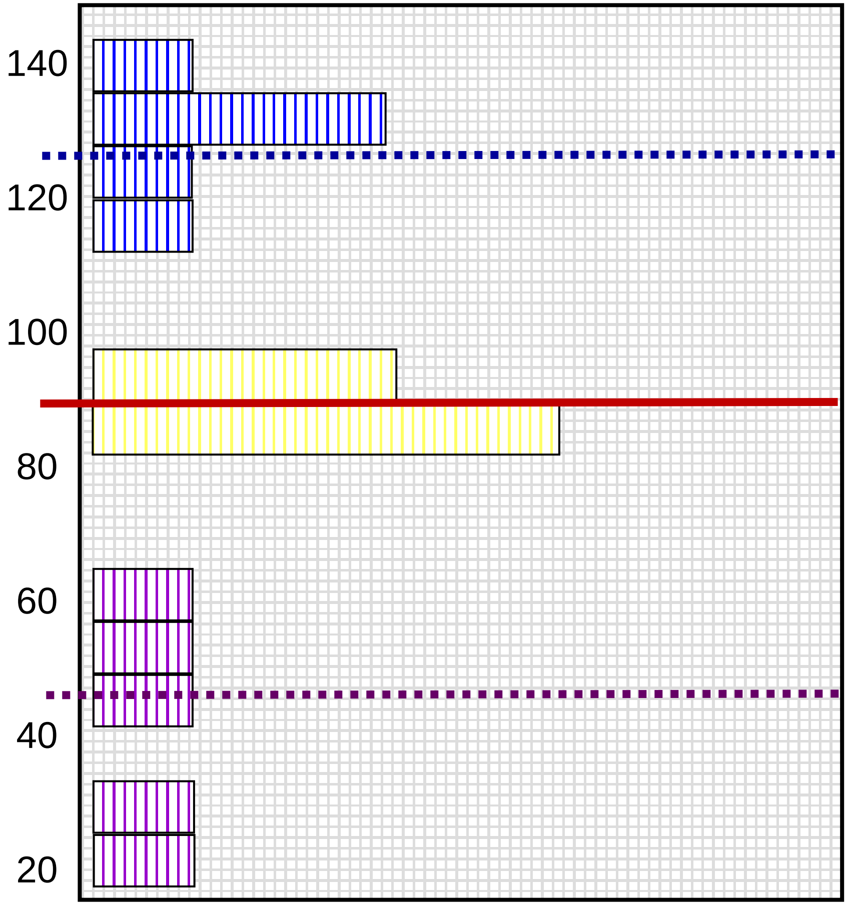
Full Sample Distribution



3 People, 5 Occasions each



Empty +Within-Person Model for y_{ti}



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

= Grand Mean

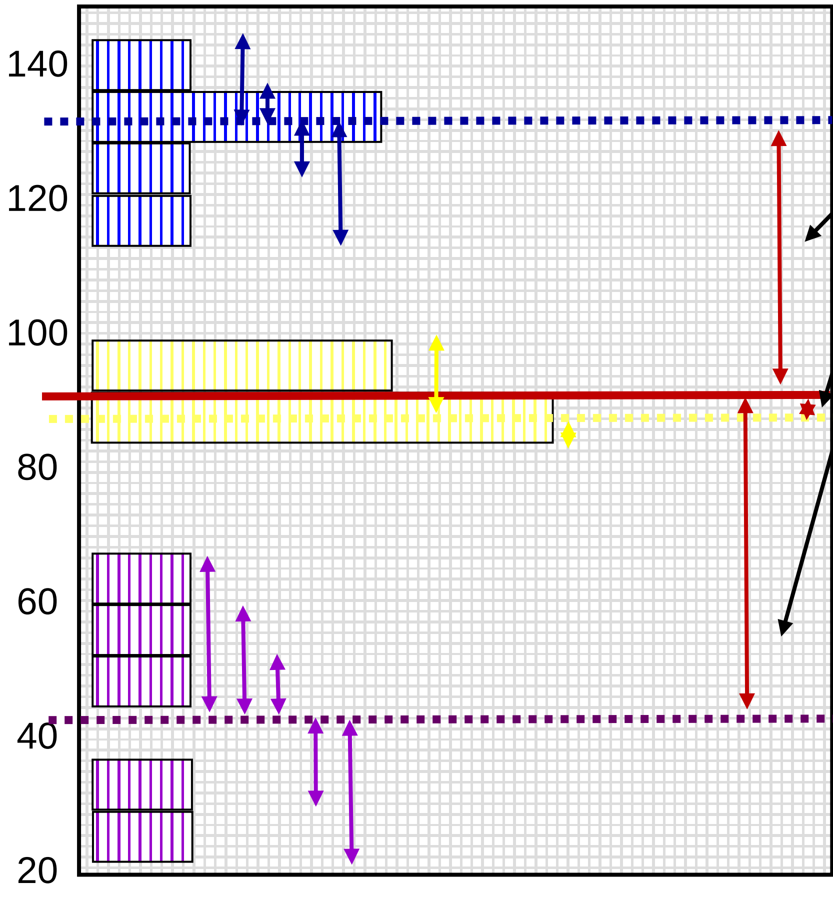
= Fixed Intercept

Can make better guess by taking advantage of repeated observations:

= Person Mean

→ Random Intercept

Empty +Within-Person Model



y_{ti} variance → 2 sources:

Between-Person (BP) Variance:

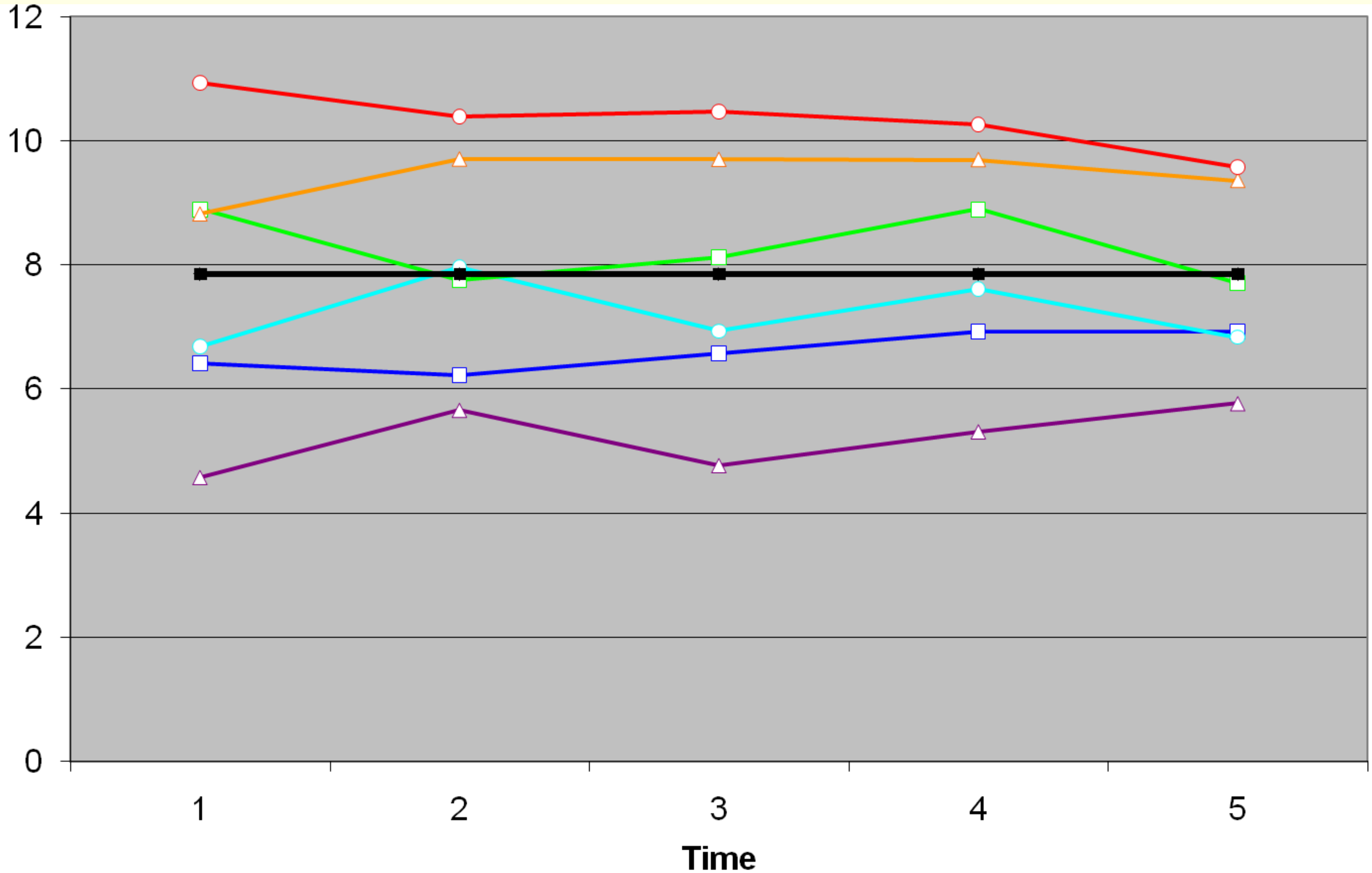
- → Differences from **GRAND** mean
- → **INTER**-Individual Differences

Within-Person (WP) Variance:

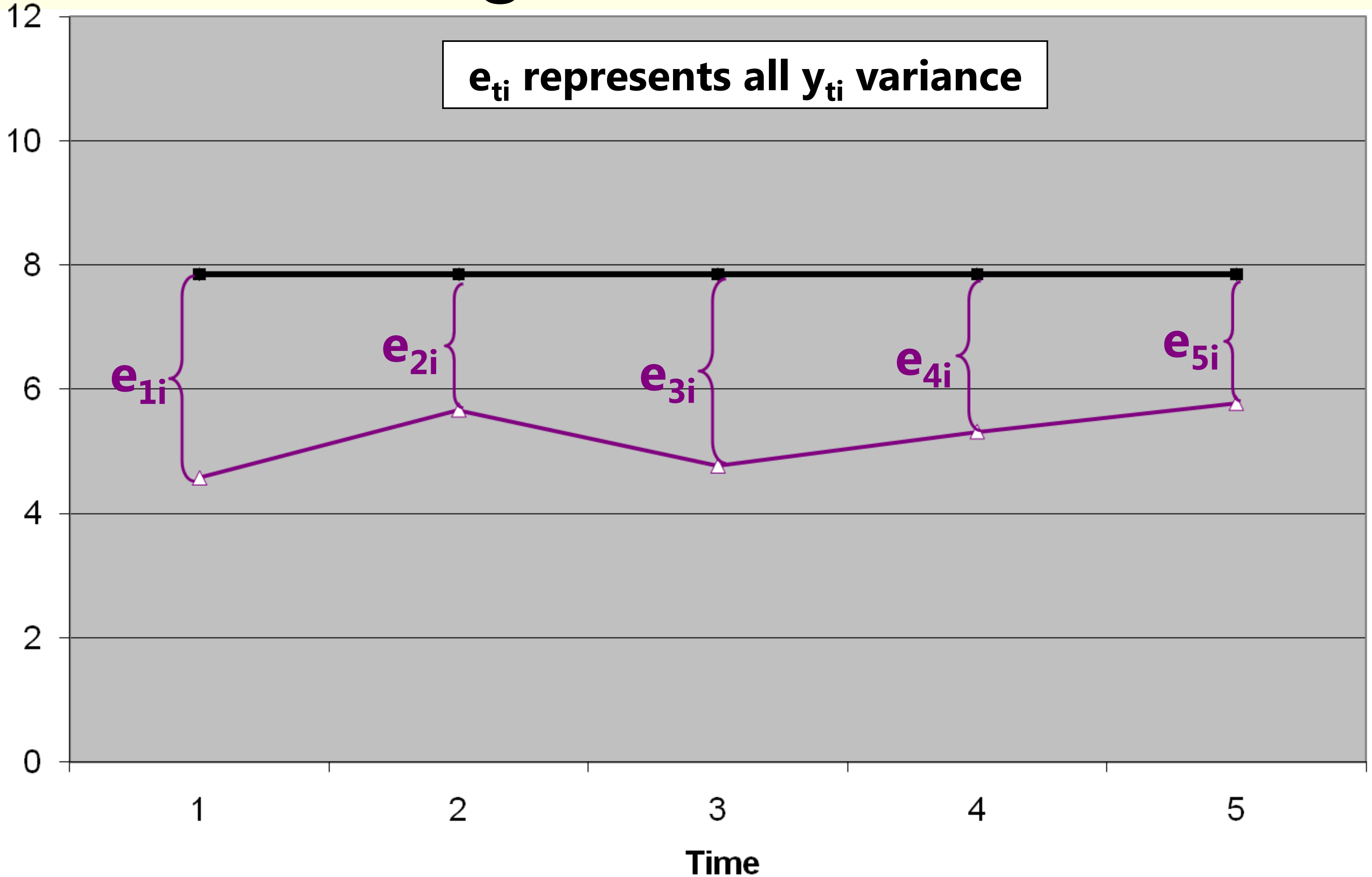
- → Differences from **OWN** mean
- → **INTRA**-Individual Differences
- → This part is only observable through longitudinal data.

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

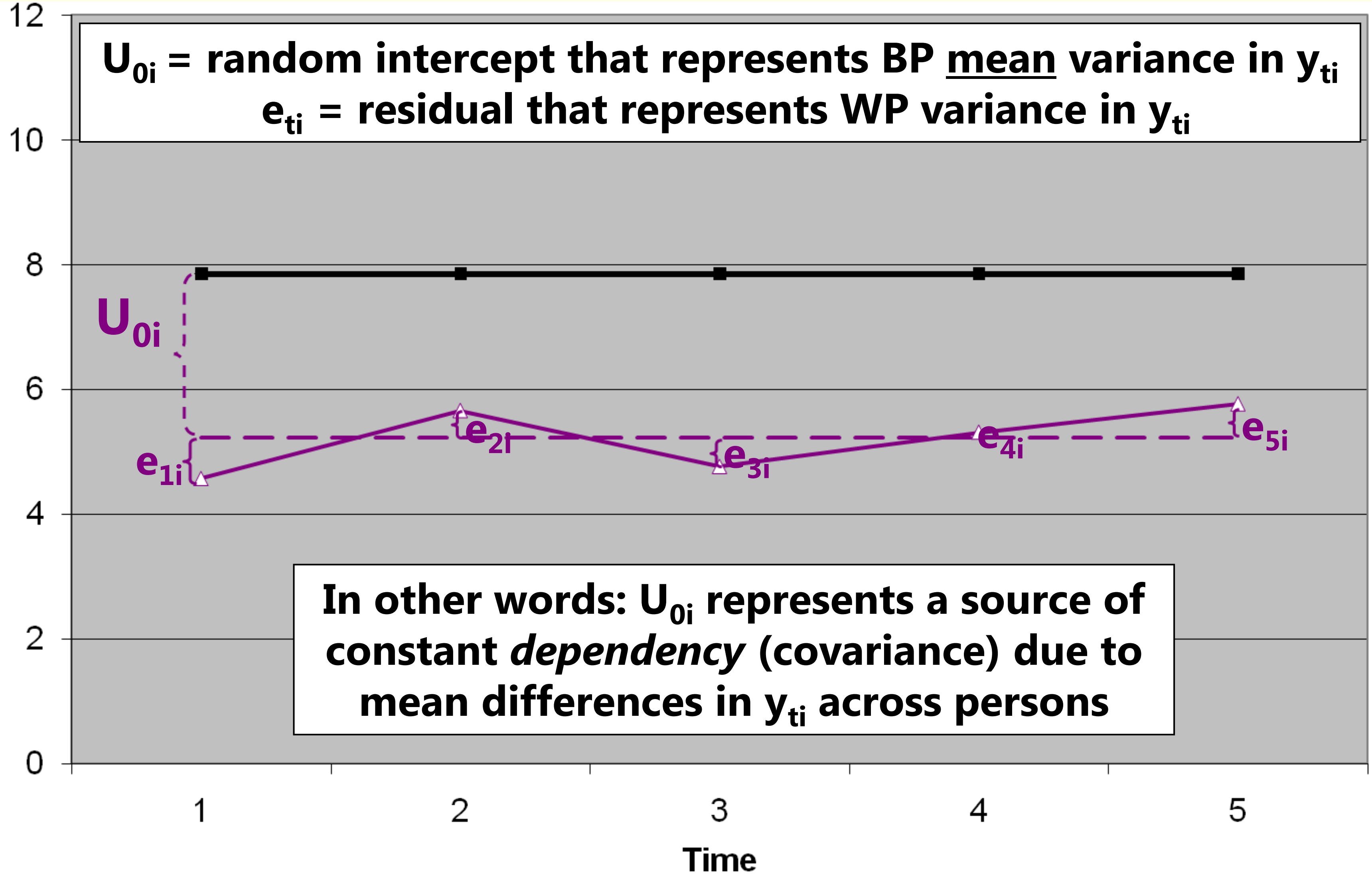
Hypothetical Longitudinal Data (black line = sample mean)



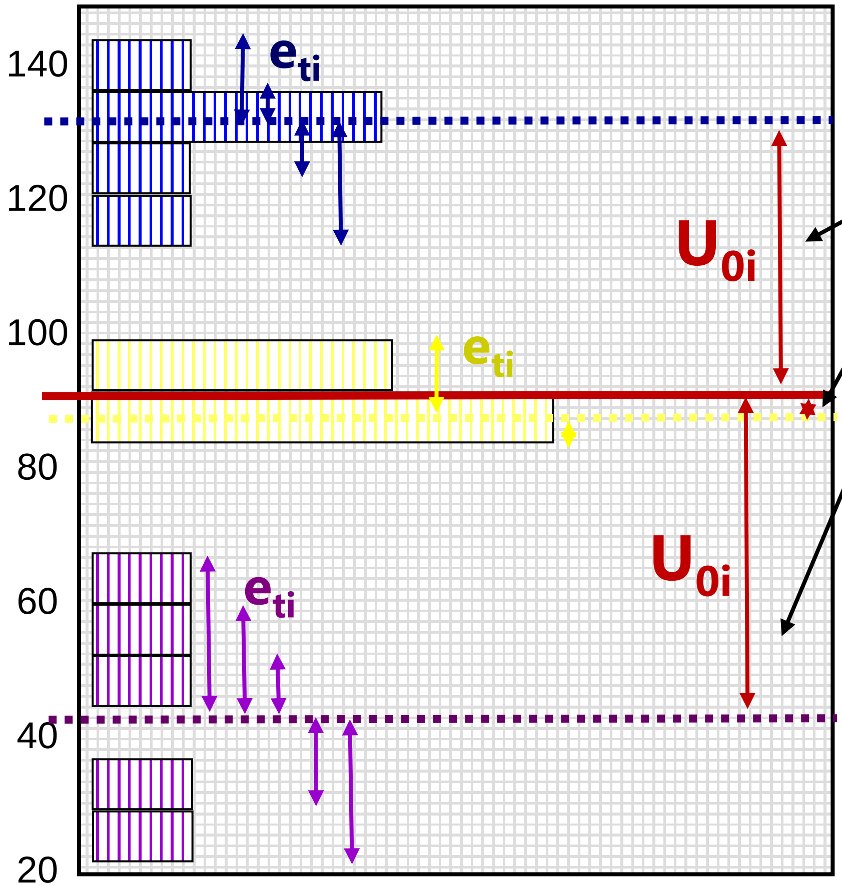
“Error” in a BP Model for the Variance: Single-Level Model



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



Empty +Within-Person Model



y_{ti} variance → 2 sources:

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

→ **Between**-Person Variance

→ Differences from **GRAND** mean

→ **INTER**-Individual Differences

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

→ **Within**-Person Variance

→ Differences from **OWN** mean

→ **INTRA**-Individual Differences

BP vs. +WP Empty Models

- Empty **Between-Person** Model (used for 1 occasion):

$$\mathbf{y_i} \quad = \quad \beta_0 \quad + \quad e_i$$

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

- Empty **+Within-Person** Model (for >1 occasions):

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

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

Intraclass Correlation (ICC)

Intraclass Correlation (ICC):

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

R matrix = residual variances and covariances by occasion

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

R matrix

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

R CORR 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 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)**

BP and +WP Conditional Models

- Multiple Regression, **Between-Person** ANOVA: **1 PILE**
  - $y_i = (\beta_0 + \beta_1 X_i + \beta_2 Z_i \ldots) + e_i$
  - e_i → ONE residual, assumed uncorrelated with equal variance across observations (here, just persons) → "**BP (all) variation**"

- Repeated Measures, **Within-Person** ANOVA: **2 PILES**
  - $y_{ti} = (\beta_0 + \beta_1 X_i + \beta_2 Z_i \ldots) + U_{0i} + e_{ti}$
  - U_{0i} → A random intercept for differences in person means, assumed uncorrelated with equal variance across persons → "**BP (mean) variation**" $= \tau^2_{U_0}$ is now "leftover" after predictors
  - e_{ti} → A residual that represents remaining time-to-time variation, usually assumed uncorrelated with equal variance across observations (now, persons and time) → "**WP variation**" $= \sigma^2_e$ is also now "leftover" after predictors

Introduction to Multilevel Longitudinal Models

- Topics:
  - What is multilevel modeling?
  - Concepts in longitudinal data
  - From between-person to within-person models
  - **Example 1: Introduction to multilevel modeling software**
  - Kinds of ANOVAs for longitudinal data

Example Data for BP and WP Models: GLM using Multilevel Modeling Software

- 50 kids in a control or treatment group each measured twice
- Hypothesis: Learning outcome should be higher at post-test than pre-test, with a greater difference in the treatment group

Means *(SE)*	Pre-Test		Post-Test	Marginal
Control	**49.08** *(1.14)*	**5.82**	54.90 *(1.13)*	51.99 *(0.89)*
	1.68	**2.04**	**3.72**	**2.71**
Treatment	50.76 *(0.91)*	**7.86**	58.62 *(0.99)*	54.70 *(0.87)*
Marginal	49.92 *(0.73)*	**6.84**	56.76 *(0.79)*	53.34 *(0.64)*

Likelihood-Based Model Comparisons for Deciding Between Alternative Models for the Variance (briefly)

- **Relative model fit** is indexed by a "**deviance**" statistic → **−2LL**
  - Log of likelihood (LL = total data height) of observing the data given model parameters; −2*LL so that the differences between model LL values follow $\sim\chi^2$
  - **−2LL is a measure of BADNESS of fit, so smaller values = better models**
  - Two flavors (labeled as −2 log likelihood in SAS, SPSS, but given as LL instead in STATA and Mplus): Maximum Likelihood (**ML**) or Restricted (Residual) ML (**REML**)

- **Nested models are compared using their deviance values: −2ΔLL Test** (i.e., Likelihood Ratio Test, Deviance Difference Test)

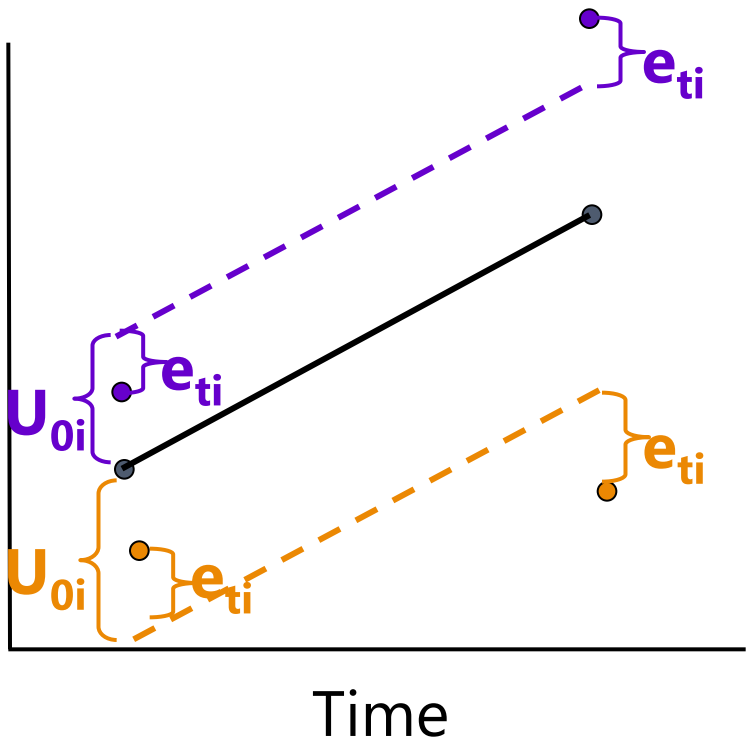
1. Calculate −2ΔLL: $(-2LL_{fewer}) - (-2LL_{more})$
2. Calculate Δdf: $(\#\ Parms_{more}) - (\#\ Parms_{fewer})$
3. Compare −2ΔLL to χ^2 distribution with df = Δdf
   CHIDIST in excel will give exact p-values for the difference test; so will STATA

1. & 2. must be positive values!

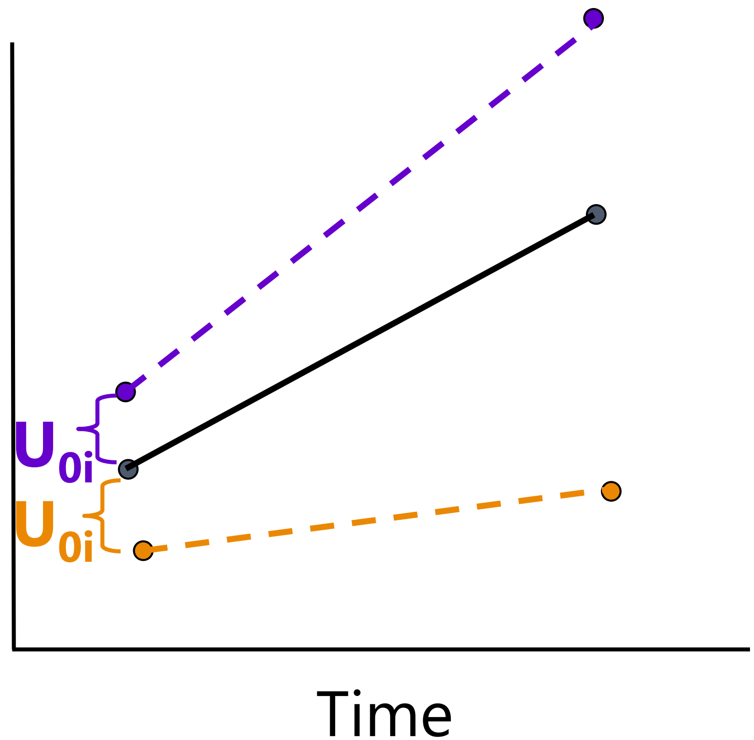
- **Add** parameters? Model fit can be **BETTER** (signif) or **NOT BETTER**
- **Remove** parameters? Model fit can be **WORSE** (signif) or **NOT WORSE**

Why error and person*time are the same thing in two-occasion data

Same age slope, so error is leftover



Different age slopes, so no error is leftover



Introduction to Multilevel Longitudinal Models

- Topics:
  - What is multilevel modeling?
  - Concepts in longitudinal data
  - From between-person to within-person models
  - Example 1: Introduction to multilevel modeling software
  - **Kinds of ANOVAs for longitudinal data**

ANOVA for 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":** $\boldsymbol{\beta_0 + \beta_1(T_1) + \beta_2(T_2) + \beta_3(T_3)}$
  - **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…

1. Between-Groups ANOVA

- **Uses e_{ti} only** (total variance = a single variance term of σ_e^2)
- **Assumes no covariance** at all among observations from the same person: *Dependency? What dependency?*
- Will usually be **very, very wrong** for longitudinal data
  - WP effects tested against wrong residual variance (significance tests will often be way too conservative)
  - Will also tend to be wrong for clustered data, but less so *(because the correlation among persons from the same group is not as strong as the correlation among occasions from the same person)*
- Predicts a variance-covariance matrix over time (here, 4 occasions) like this, called "**Variance Components**" (**R** matrix is TYPE=VC on REPEATED):

R matrix

$$\begin{bmatrix} \sigma_e^2 & 0 & 0 & 0 \\ 0 & \sigma_e^2 & 0 & 0 \\ 0 & 0 & \sigma_e^2 & 0 \\ 0 & 0 & 0 & \sigma_e^2 \end{bmatrix}$$

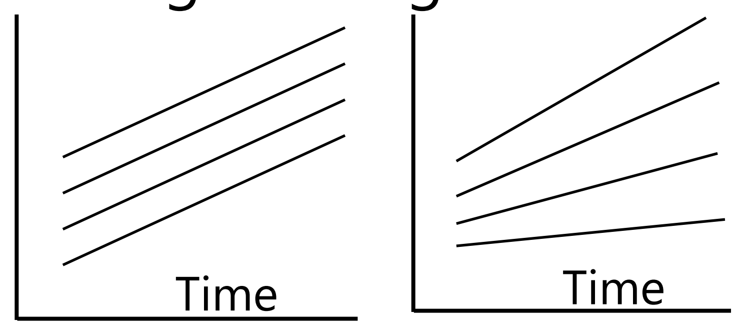
2a. Univariate Repeated Measures

- Separates total variance into **two** sources:
  - **Between-Person** (mean differences due to U_{0i}, or $\boldsymbol{\tau^2_{U_0}}$ across persons)
  - **Within-Person** (remaining variance due to e_{ti}, or $\boldsymbol{\sigma^2_e}$ across time, person)
- Predicts a variance-covariance matrix over time (here, 4 occasions) like this, called "**Compound Symmetry**" (**R** matrix is TYPE=**CS** on REPEATED):
  - **Mean differences from U_{0i} are the <u>only</u> reason why occasions are correlated**

R matrix

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

- Will usually be at least somewhat wrong for longitudinal data
  - If people change at different rates, the variances and covariances over time have to change, too



The Problem with Univariate RM ANOVA

- Univ. RM ANOVA $\left(\tau^2_{U_0} + \sigma^2_e\right)$ predicts **compound symmetry:**
  - All variances and all covariances are equal across occasions
  - In other words, the amount of error observed should be the same at any occasion, so a single, pooled residual variance term makes sense
  - If not, tests of fixed effects may be biased (i.e., sometimes tested against too much or too little error, if error is not really constant over time)
  - **COMPOUND SYMMETRY RARELY FITS FOR LONGITUDINAL DATA**
- But to get the correct tests of the fixed effects, the data must only meet a less restrictive assumption of **sphericity**:
  - In English → **pairwise differences** between adjacent occasions have equal variance and covariance (satisfied by default with only 2 occasions)
  - If compound symmetry is satisfied, so is sphericity (but see above)
  - Significance test provided in ANOVA for where data meet sphericity assumption
  - **Other RM ANOVA approaches are used when sphericity fails…**

The Other Repeated Measures ANOVAs…

- 2b. **Univariate RM ANOVA with sphericity corrections**
  - Based on ε → how far off sphericity (from 0-1, 1=spherical)
  - Applies an overall correction for model df based on estimated ε, but it doesn't really address the problem that data ≠ model

- 3. **Multivariate Repeated Measures ANOVA**
  - All variances and covariances are estimated separately over time (here, *n* = 4 occasions), called "**Unstructured**" (**R** matrix is TYPE=UN on REPEATED)—it's not a model, it IS the data:

R matrix

$$\begin{bmatrix} \sigma_{11}^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{21} & \sigma_{22}^2 & \sigma_{23} & \sigma_{24} \\ \sigma_{31} & \sigma_{32} & \sigma_{33}^2 & \sigma_{43} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_{44}^2 \end{bmatrix}$$

  - Because it can never be wrong, UN can be useful for **complete and balanced longitudinal data** with few (e.g., 2-4) occasions (*n*)
  - Parameters = $\frac{n*(n+1)}{2}$ so it can be hard to estimate with many occasions
  - Unstructured can also be specified to include random intercept variance $\tau_{U_0}^2$
  - All other models for the variance are nested under Unstructured, so we can do LRT model comparisons to see if any other model is NOT WORSE

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

- **Saturated Model for the Means** (balanced time required)
  - All possible mean differences
  - Unparsimonious, but best-fitting (is a description, not a model)

- **3 kinds of Models for the Variance** (need complete data in least squares)
  - BP ANOVA (σ_e^2 only) → assumes independence and constant variance over time
  - Univ. RM ANOVA $\left(\tau_{U_0}^2 + \sigma_e^2\right)$ → assumes constant variance and covariance
  - Multiv. RM ANOVA (whatever) → no assumptions; is a description, not a model

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 (polynomials, pieces)
  - Random intercepts and slopes and/or alternative covariance structures that *predict* intermediate patterns of variance and covariance over time