Introduction to Multilevel Models

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

What is a Multilevel Model (MLM)?

- Same as other terms you have heard of:
  - **General 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
  - **Hierarchical Linear Model** (if you are from education)
    - Not the same as hierarchical regression

- Special cases of MLM:
  - Random Effects ANOVA or Repeated Measures ANOVA
  - (Latent) Growth Curve Model (where "Latent" → 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)

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)*

Dimensions for Organizing Models

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

Note: Least Squares is only for GLM

- **<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)

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

How We Will Learn MLM

- "Levels" are defined by the context of a study
  - **Level ≈ a dimension of sampling** (can be nested or crossed)

- We will start with MLM for longitudinal data…
  - Level 1 = variation over time, Level 2 = variation over persons
  - More complex case because of the time dimension

- …We will follow with MLM for clustered data…
  - Level 1 = variation over persons, Level 2 = variation over groups

- … and conclude with MLM for clustered+longitudinal data
  - Time (Level 1) within persons (Level 2) within groups (Level 3)
  - Persons (Level 1) within occasions (Level 2) within groups (Level 3)

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, person within group, school within country
  - 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"

Dependency comes from…

- Mean differences across sampling units (persons, groups)
  - Creates constant dependency over time (or persons)
  - Will be represented by a random intercept in our models

- Individual/group differences in effects of predictors
  - Longitudinal: individual differences in growth, stress reactivity
  - Clustered: group differences in slopes of person predictors
  - Creates non-constant 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

- Longitudinal data: 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 Between-Person predictors
  - Within-Person variation needs Within-Person predictors
  - Between-Group variation needs Between-Group predictors

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

- ANOVA: test differences among discrete IV factor levels
  - Between-Groups: Gender, Intervention Group, Age Groups
  - Within-Subjects (Repeated Measures): Condition, Time
  - Test main effects of continuous covariates (ANCOVA)

- Regression: test whether slopes relating predictors to outcomes are different from 0
  - Persons measured once, differ categorically or continuously on a set of time-invariant (person-level) covariates

- What if a predictor is assessed repeatedly (time-varying predictors) but can't be characterized by 'conditions'?
  - ANOVA or Regression won't work → need MLM

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

- Some things <u>don’t change</u> over measurements…
  - Sex, Ethnicity
  - → <u>Time-Invariant</u> Predictor = <u>Person</u> Level
- Some things <u>do change</u> over measurements…
  - Health Status, Stress Levels, Living Arrangements
  - → <u>Time-Varying</u> Predictor = <u>Time</u> Level
- Some predictors might be measured at <u>higher levels</u>
  - Family SES, length of marriage, school size, country size
- Interactions between levels may be included, too
  - Does the effect of <u>health status</u> differ by <u>gender</u> and <u>SES</u>?

Level: Time Person Family

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

100 uses 4 cases

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")?

Introduction to Multilevel Models

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

Options for Longitudinal Models

- Although models and software are logically separate, longitudinal data can be analyzed via multiple analytic frameworks:
  - "Multilevel/Mixed Models"
    - Dependency over time, persons, groups, etc. is modeled via random effects (multivariate through "levels" using stacked/long data)
    - Builds on GLM, generalizes easier to additional levels of analysis
  - "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

  - Because random effects and latent variables are the same thing, many longitudinal models can be specified/estimated either way
    - And now "Multilevel Structural Equation Models" can do it all…

Data Requirements for Our Models

- A useful outcome variable:
  - Has an interval scale*
    - A one-unit difference means the same thing across all scale points
    - In subscales, each contributing item has an equivalent scale
    - **Other kinds of outcomes can be analyzed using generalized multilevel models instead, but estimation is more challenging*
  - Has scores with the same meaning over observations
    - Includes meaning of construct
    - Includes how items relate to the scale
    - Implies measurement invariance
- FANCY MODELS CANNOT SAVE BADLY MEASURED VARIABLES OR CONFOUNDED RESEARCH DESIGNS.

Requirements for Longitudinal Data

- Multiple OUTCOMES 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

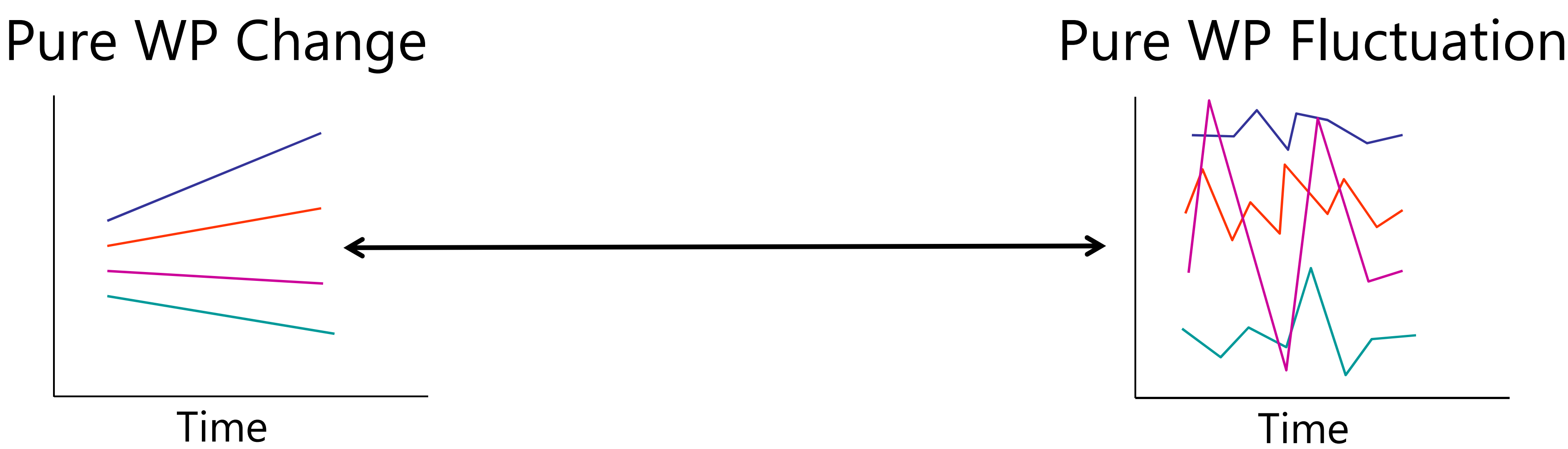
Levels of Analysis in Longitudinal Data

- 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 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 Models

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

The Two Sides of a (BP) Model

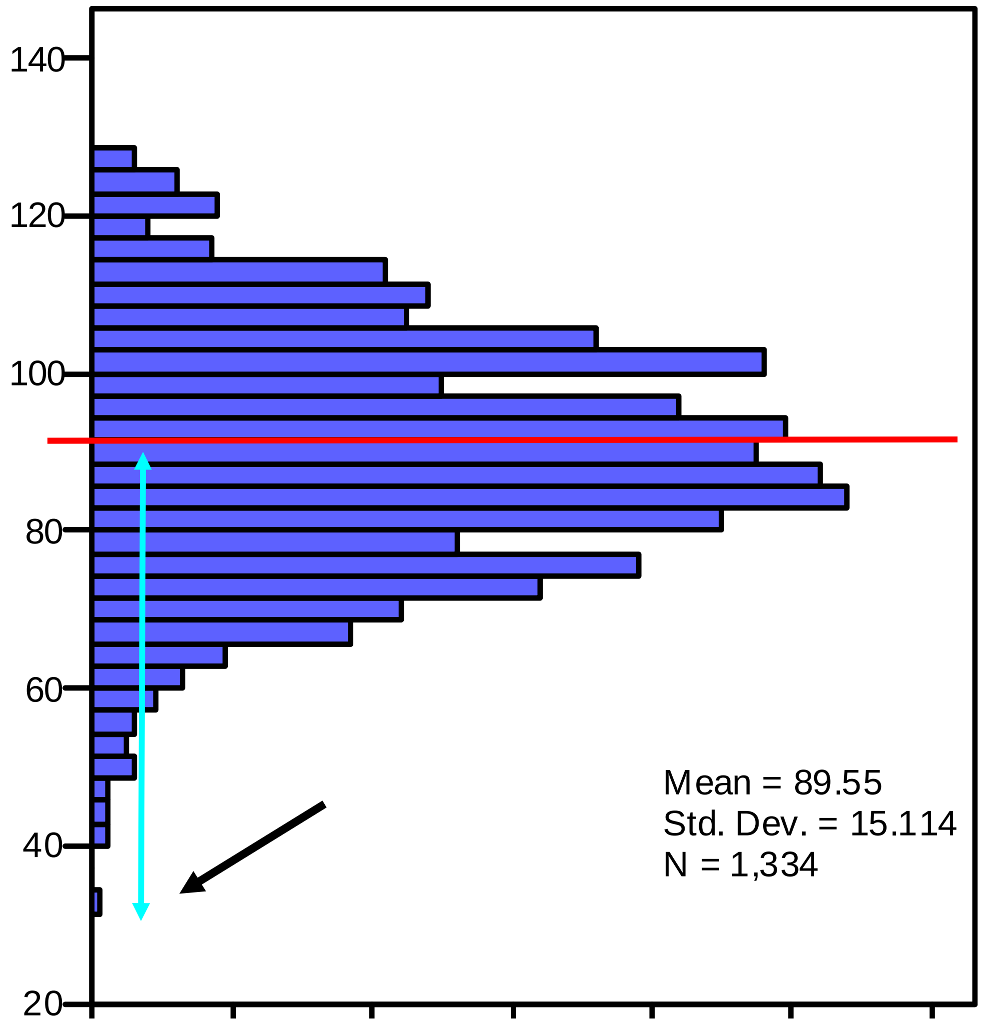
$$y_i = \boxed{\beta_0 + \beta_1 X_i + \beta_2 Z_i + \beta_3 X_i Z_i} + \boxed{e_i}$$

Our focus today

- **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)



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

Filling in values:

$$\mathbf{32 = \underbrace{90}_{\text{Y pred}} + -58}$$

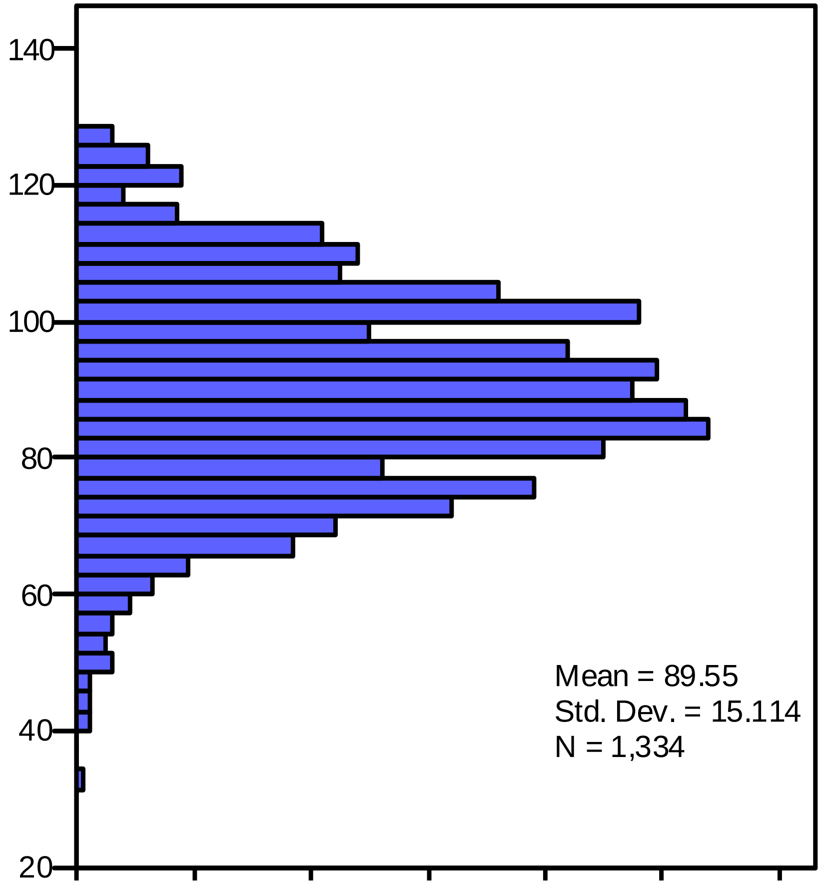
Model for the Means

Y Error Variance:

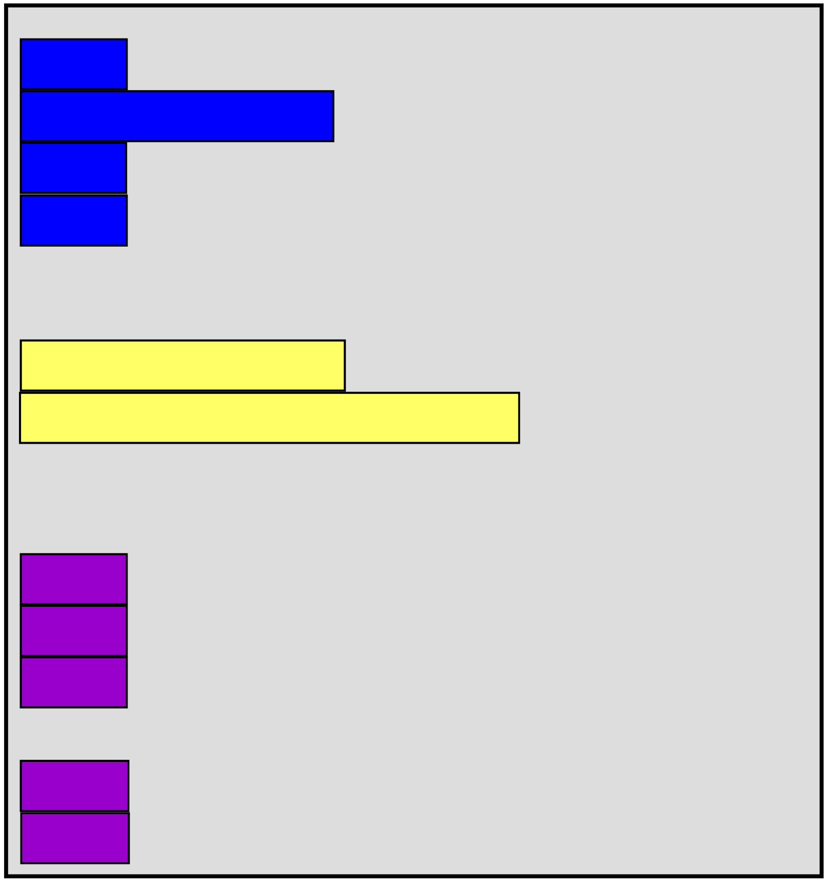
$$\frac{\Sigma\,(y - y_{pred})^2}{N - 1}$$

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

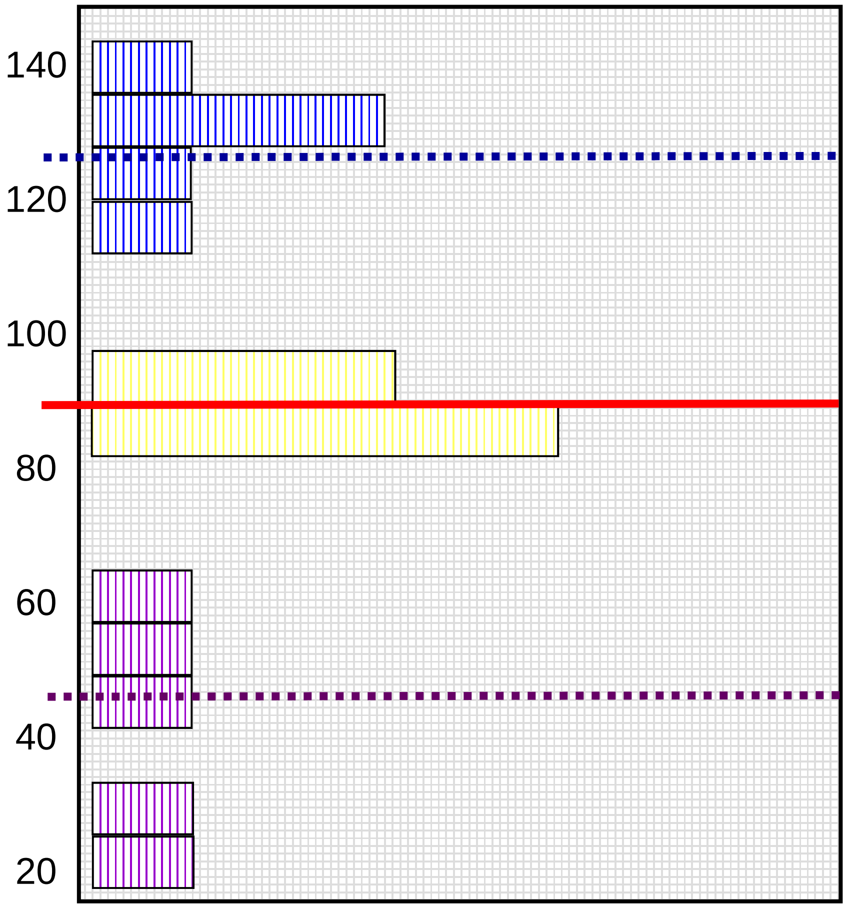
Full Sample Distribution



3 People, 5 Occasions each



Empty +Within-Person Model



Start off with Mean of Y as "best guess" for any value:

= Grand Mean

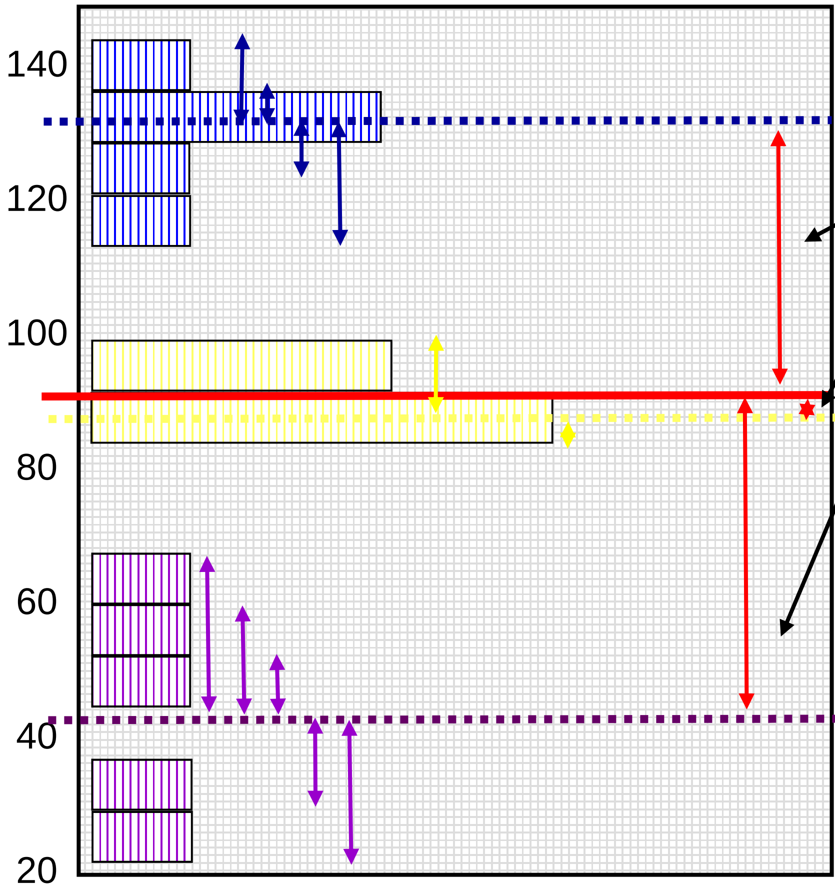
= Fixed Intercept

Can make better guess by taking advantage of repeated observations:

= Person Mean

→ Random Intercept

Empty +Within-Person Model



Variance of Y → 2 sources:

<u>Between-Person (BP) Variance</u>:

→ Differences from **GRAND** mean

→ **INTER**-Individual Differences

<u>Within-Person (WP) Variance</u>:

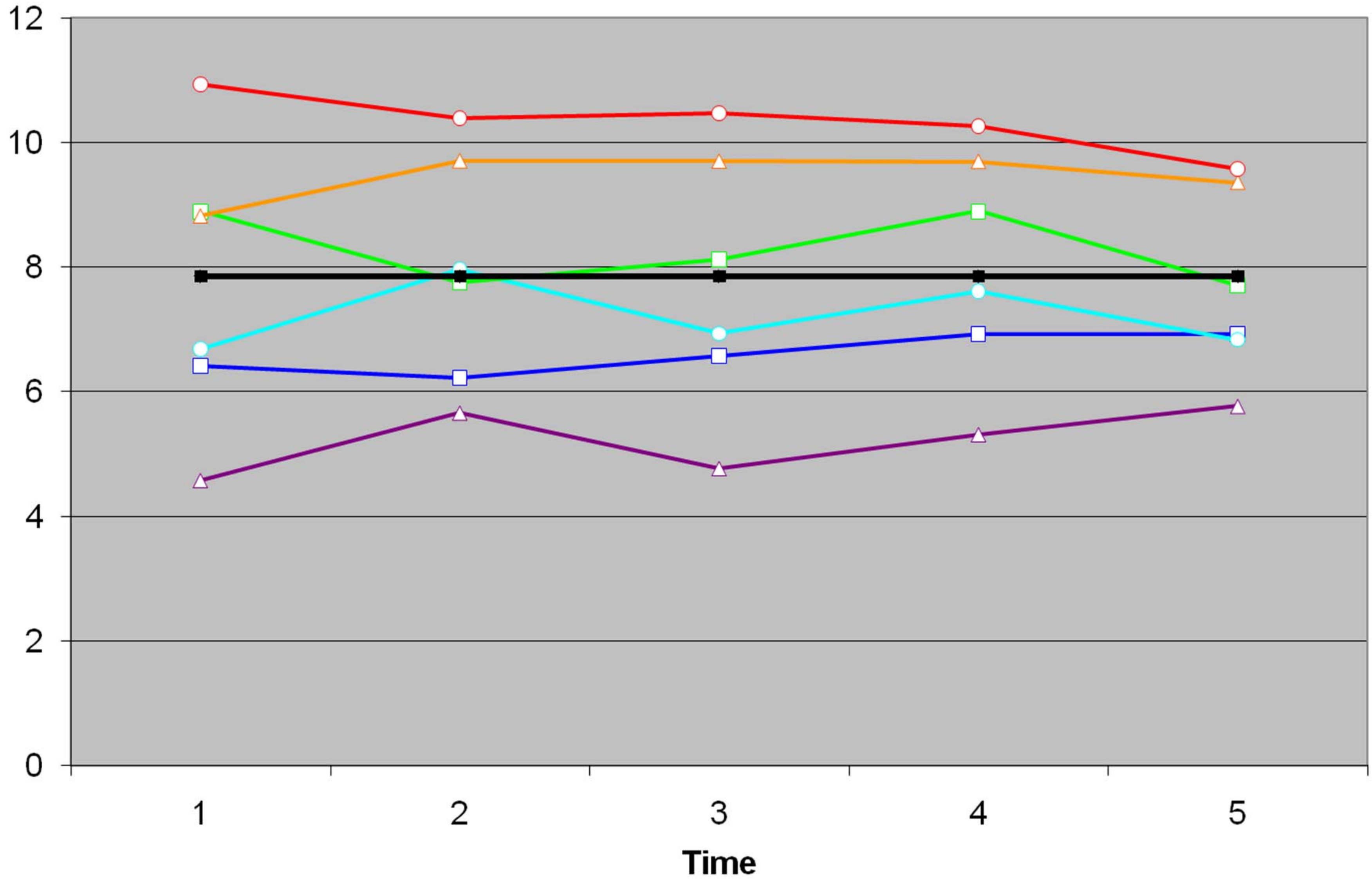
→ Differences from **OWN** mean

→ **INTRA**-Individual Differences

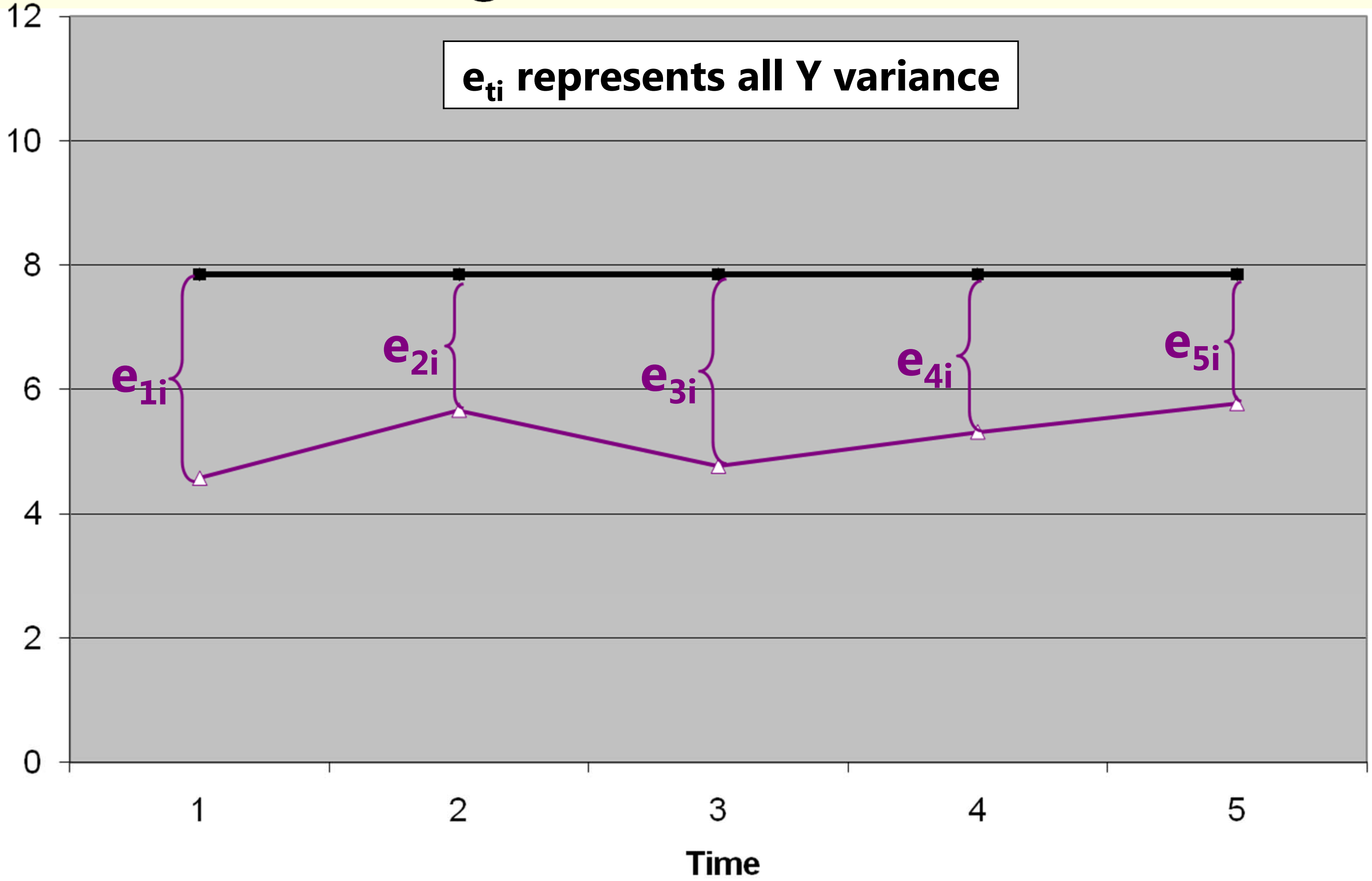
→ This part is only observable through longitudinal data.

Now we have 2 piles of variance in Y to predict.

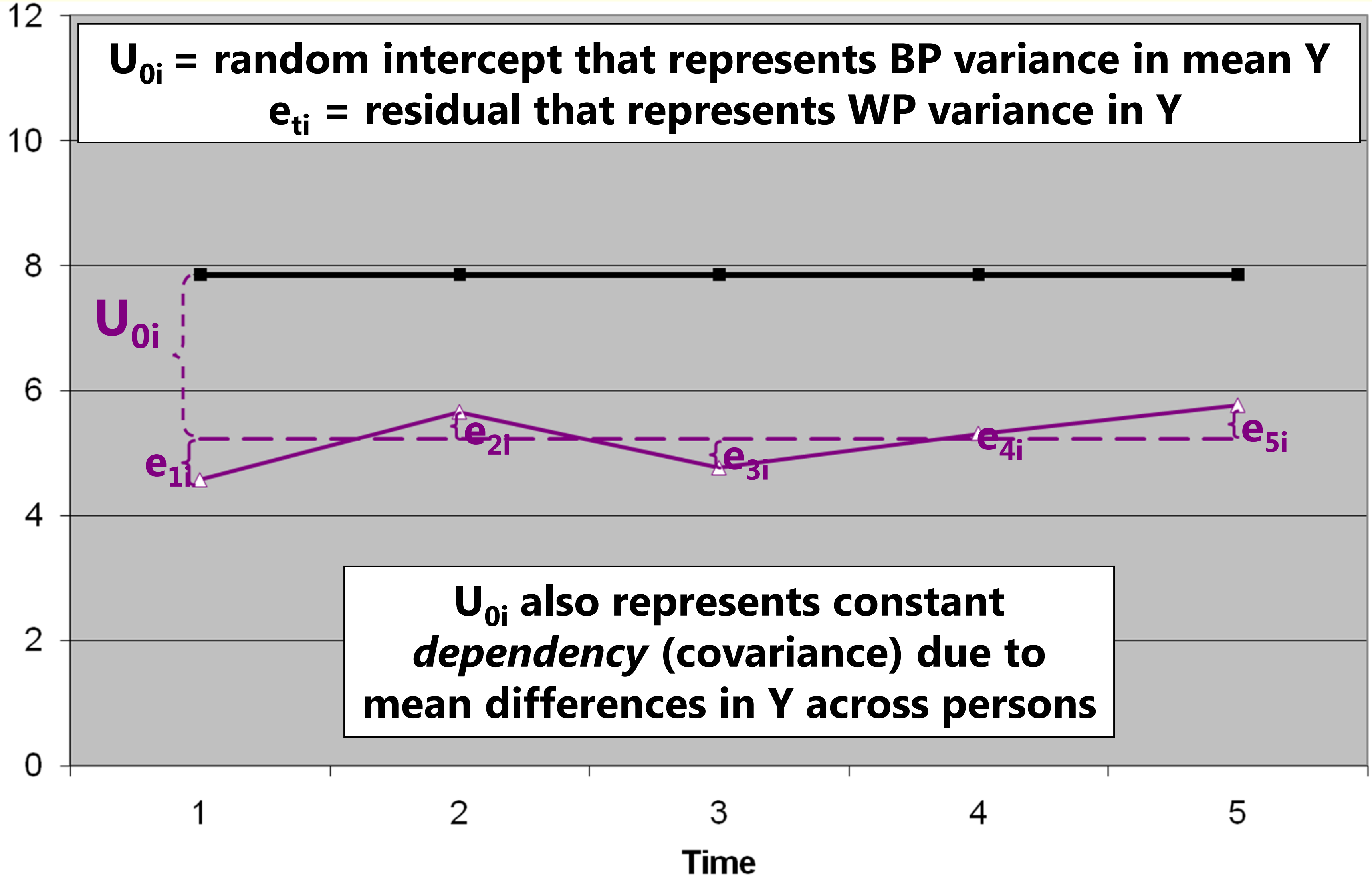
Hypothetical Longitudinal Data



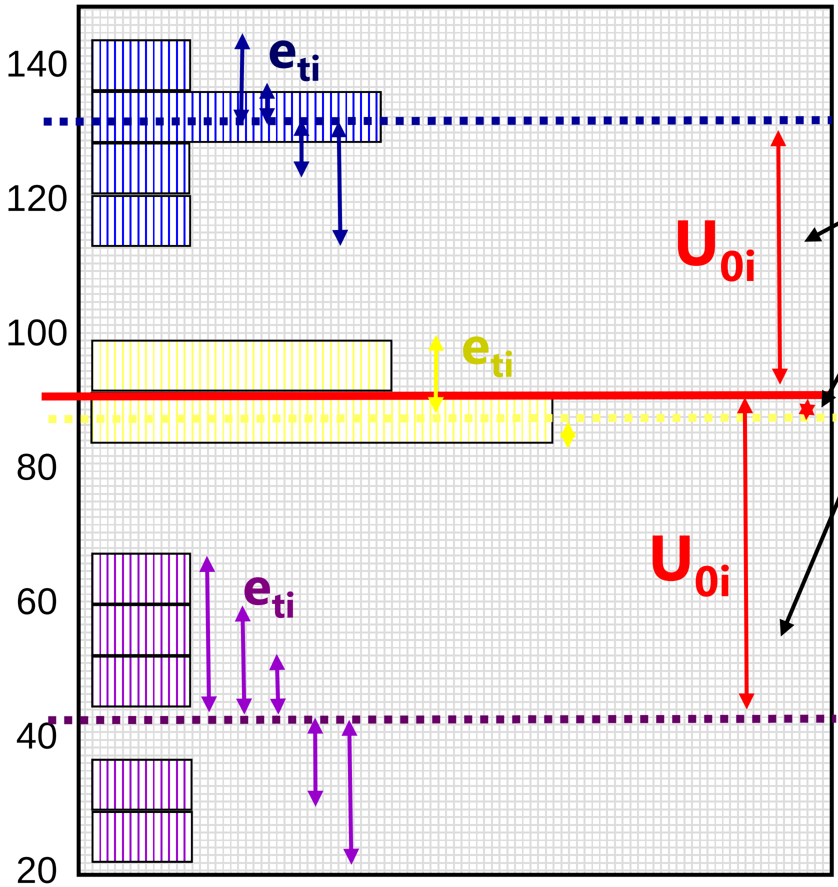
"Error" in a BP Model for the Variance: Single-Level Model



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



Empty +Within-Person Model



Variance of Y → 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):

$$y_i = \beta_0 + e_i$$

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

- Empty **+Within-Person** Model (>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

Intraclass Correlation (ICC)

Intraclass Correlation (ICC):

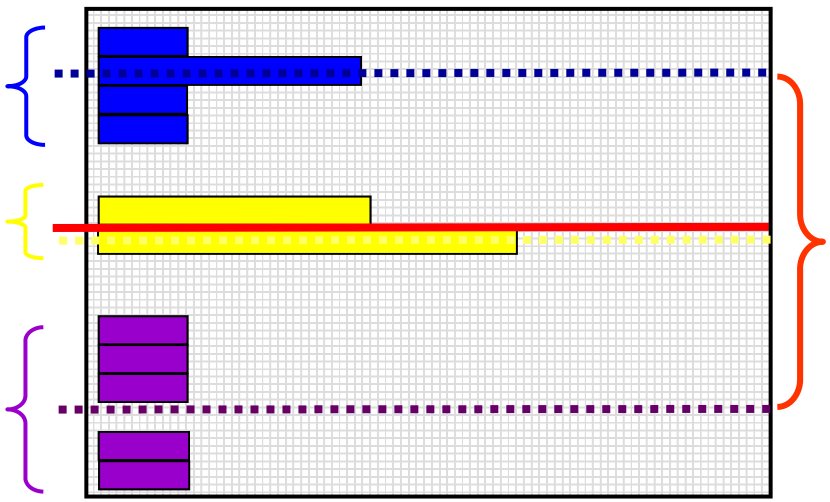
$$\mathrm{ICC} = \frac{\mathrm{BP}}{\mathrm{BP} + \mathrm{WP}} = \frac{\text{Intercept Variance}}{\text{Intercept Variance} + \text{Residual Variance}}$$

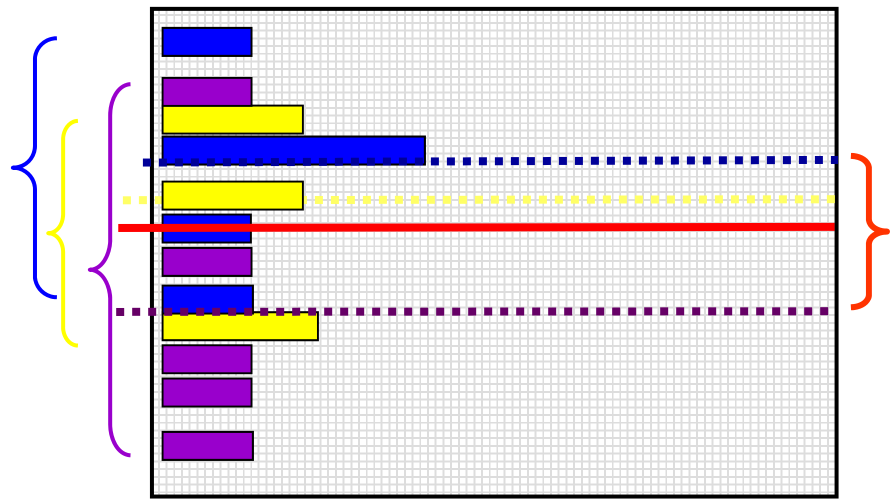
$$= \frac{\text{BP Variance in Mean Outcome}}{\text{Total Outcome Variance}} = \frac{\boldsymbol{\tau}^2_{\mathbf{U_0}}}{\boldsymbol{\tau}^2_{\mathbf{U_0}} + \boldsymbol{\sigma}^2_{\mathbf{e}}}$$

- ICC = Proportion of total variance that is between persons
- ICC = Average correlation among occasions
- 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 *constant* person dependency)**

$$ICC = \frac{Between-Person}{Between-Person + Within-Person}$$

<u>Counter-Intuitive:</u> Between-Person Variance is in the numerator, but the ICC is the correlation over time!





<u>ICC</u> = BTW / BTW + within

→ Large ICC

→ Large correlation over time

<u>ICC</u> = btw / btw + WITHIN

→ Small ICC

→ Small correlation over time

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**" = σ^2_e is also now "leftover" after predictors

Introduction to Multilevel Models

- Topics:
  - What is multilevel modeling?
  - Concepts in longitudinal data
  - From between-person to within-person models
  - **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 due to 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 variances…
  - 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**":

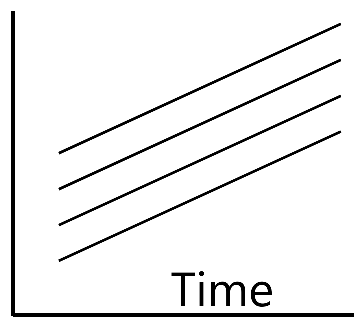
$$\begin{bmatrix} \sigma_e^2 & 0 & 0 & 0 \\ 0 & \sigma_e^2 & 0 & 0 \\ 0 & 0 & \sigma_e^2 & 0 \\ 0 & 0 & 0 & \sigma_e^2 \end{bmatrix}$$

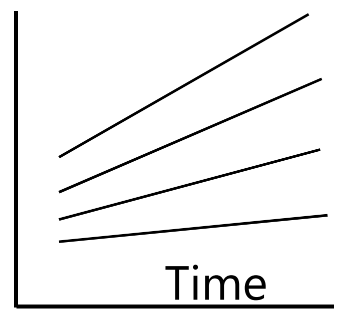
2a. Univariate Repeated Measures

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

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

- 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 error 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, 4 occasions), called "**Unstructured**"—it's not a model, it IS the data reproduced directly:

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

  - Because it can never be wrong, UN can be useful for **complete and balanced longitudinal data** with few occasions (e.g., 2-4)
  - Parameters = $\dfrac{\#\text{occasions} * (\#\text{occasions} + 1)}{2}$ so can be hard to estimate
  - Unstructured can also be specified to include random intercept variance $\tau_{U_0}^2$
  - Every other model for the variances is nested within Unstructured (we can do model comparisons to see if all other models are 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 Variances** (complete data required)
  - 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