# Introduction to Within-Person Analysis and RM ANOVA

- Today's Class:
  - From between-person to within-person
  - ANOVAs for longitudinal data
  - Variance model comparisons using −2ΔLL

# 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, $\beta_0$, $\beta_1$, $\beta_2$, and $\beta_3$)
  - The number of fixed effects will show up in formulas as $k$ (so $k$ = 4 here)

- **Model for the Variance ("Piles" of Variance):**
  - $e_i \sim N(0, \sigma_e^2)$ → **ONE** source of residual (unexplained) deviation
  - $e_i$ has a mean of 0 with some estimated constant residual variance $\sigma_e^2$, is normally distributed, is unrelated to $X$ and $Z$, and is unrelated across people (across all observations, just people here)
  - **Contains residual variance only in above BP model**

# Review: Variances and Covariances

**Variance:**
Dispersion of y

$$\text{Variance}\ (y_t) = \frac{\sum_{i=1}^{N} (y_{ti} - \hat{y}_{ti})^2}{N - k}$$

**Covariance:**
How y's go together, unstandardized

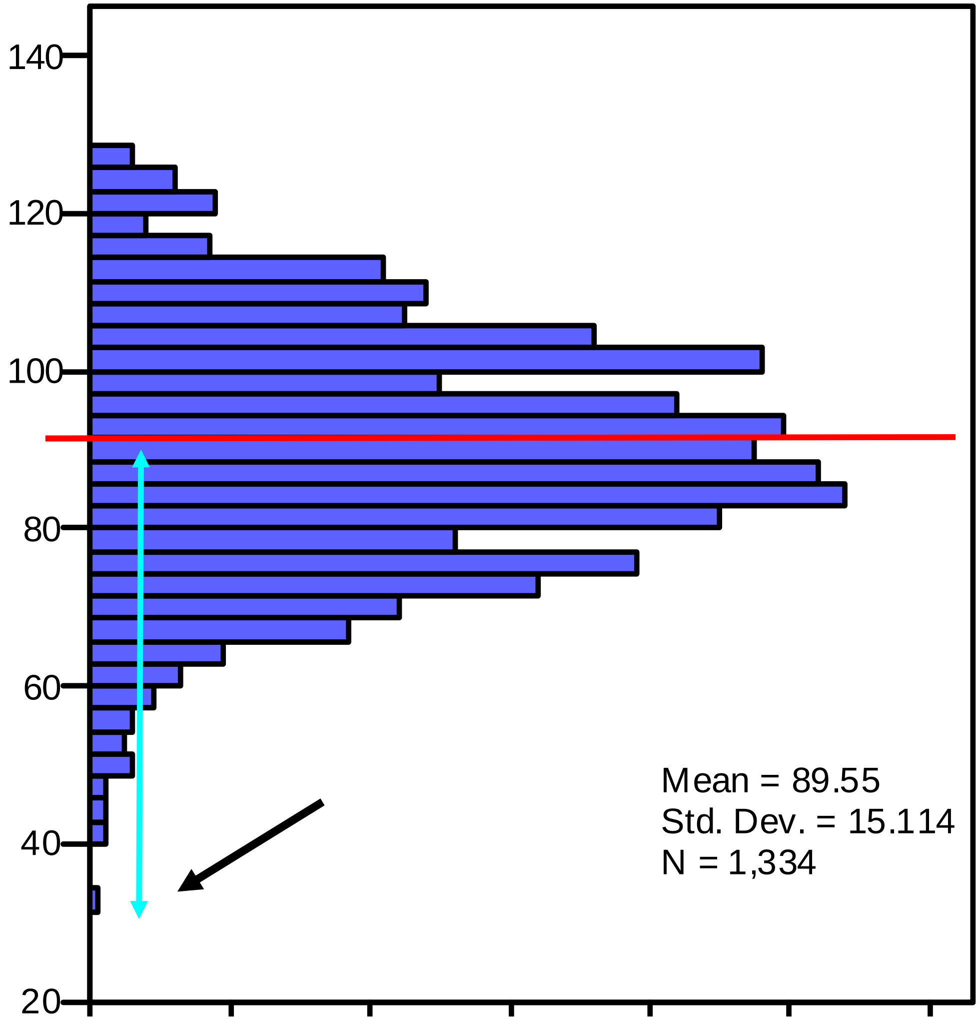
$$\text{Covariance}\ (y_1, y_2) = \frac{\sum_{i=1}^{N} (y_{1i} - \hat{y}_{1i})(y_{2i} - \hat{y}_{2i})}{N - k}$$

**Correlation:**
How y's go together, standardized (−1 to 1)

$$\text{Correlation}\ (y_1, y_2) = \frac{\text{Covariance}(y_1, y_2)}{\sqrt{\text{Variance}(y_1)} * \sqrt{\text{Variance}(y_2)}}$$

N = # people, t = time, i = person
k = # fixed effects, $\hat{y}_{ti}$ = y predicted from fixed effects

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



Mean = 89.55
Std. Dev. = 15.114
N = 1,334

$$y_i = \beta_0 + e_i$$

**Filling in values:**

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

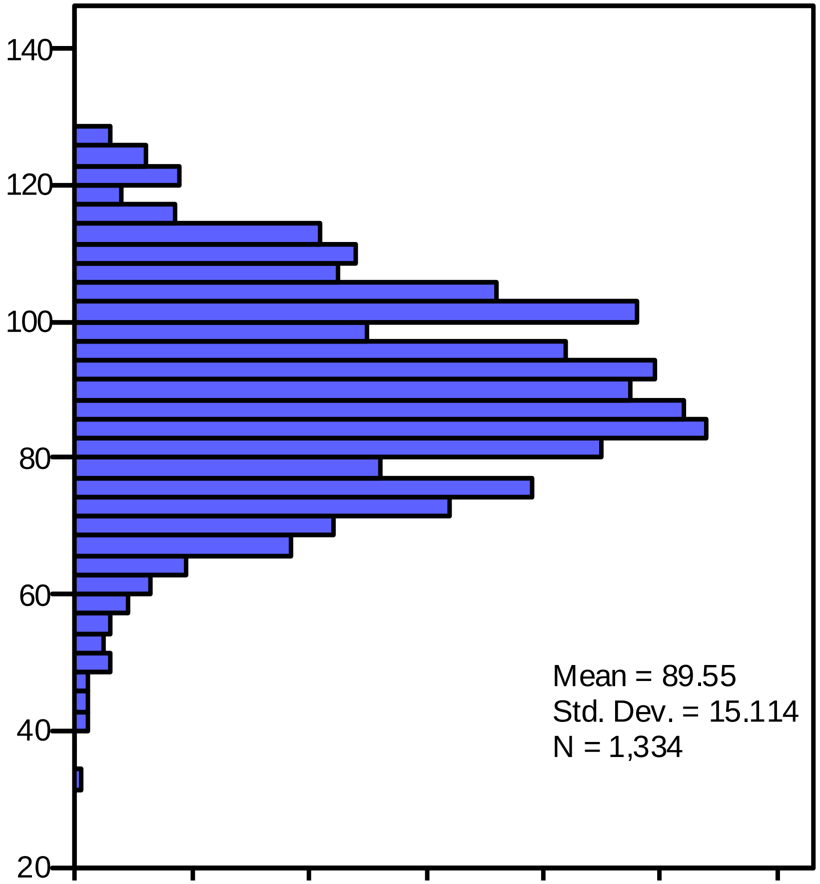
Model for the Means

Y Error Variance:

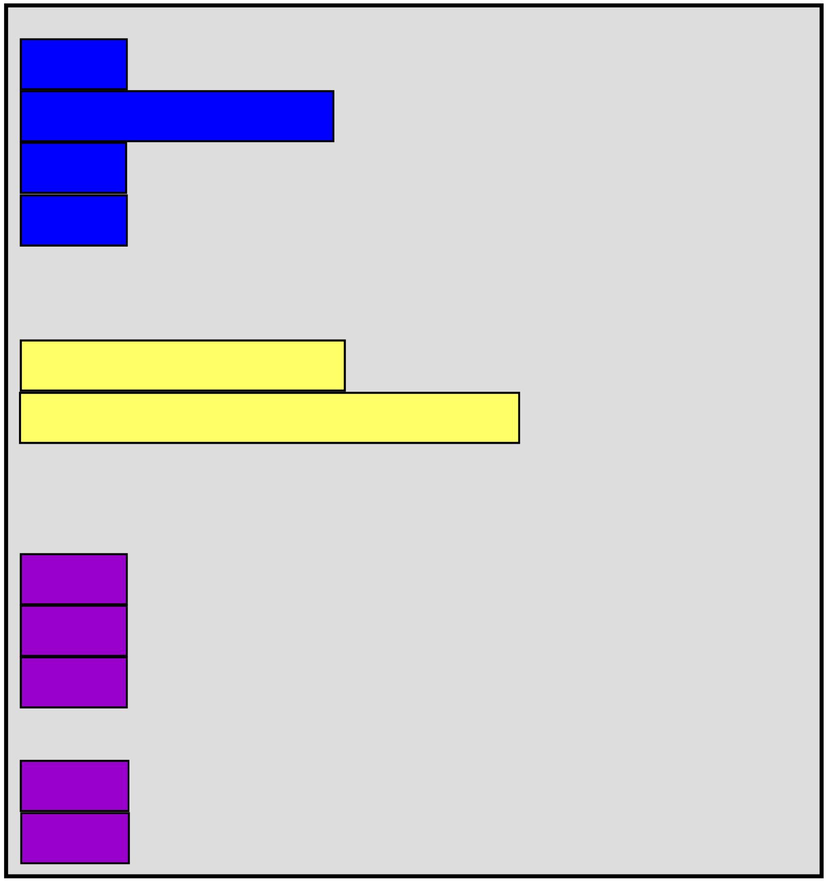
$$\frac{\Sigma (y - y_{pred})^2}{N - 1}$$

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

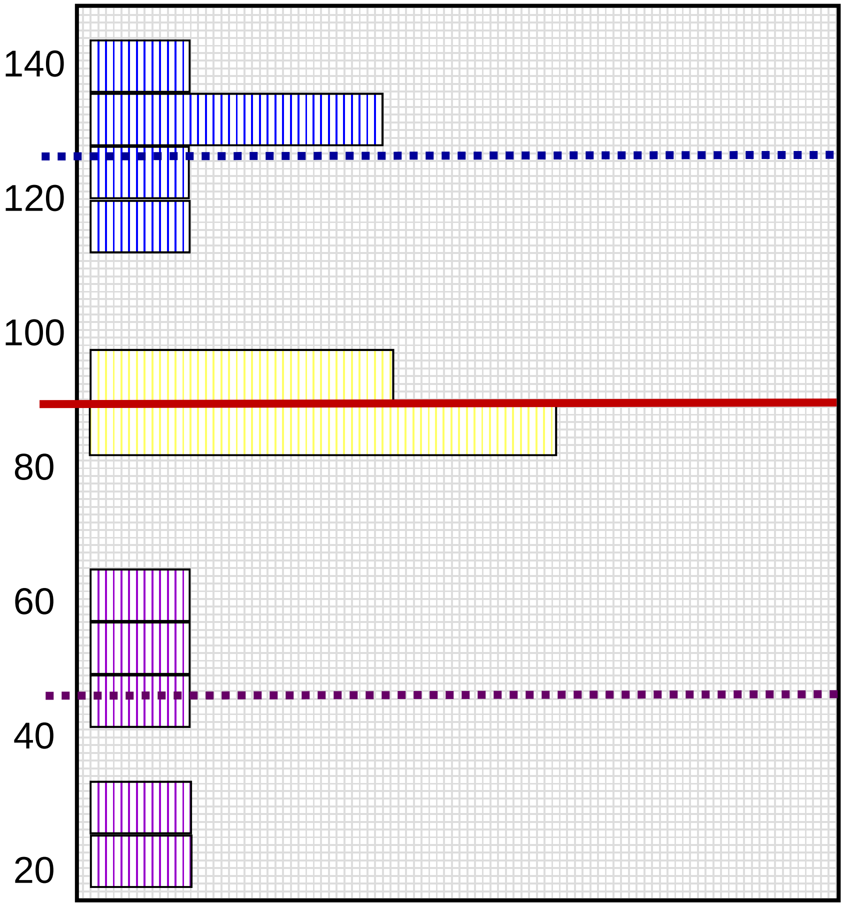
Full Sample Distribution



3 People, 5 Occasions each



# Empty +Within-Person Model



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

= Grand Mean

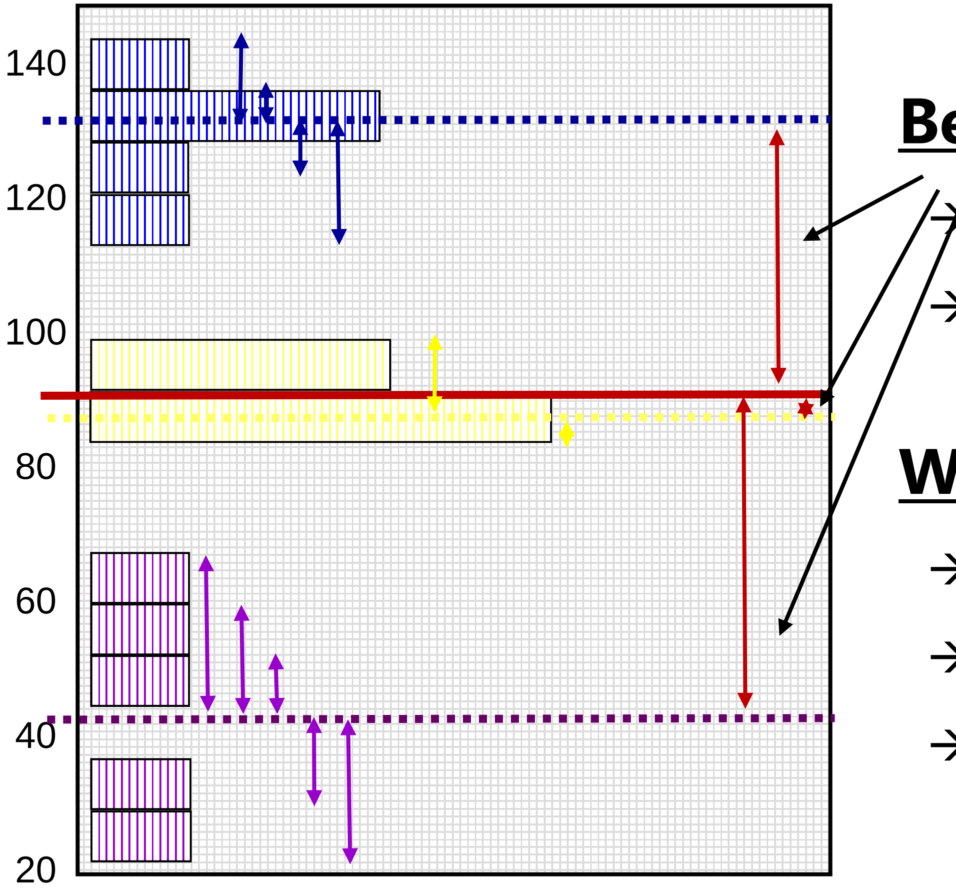
= Fixed Intercept

**Can make better guess by taking advantage of repeated observations:**

= Person Mean

→ Random Intercept

# Empty +Within-Person Model



Variance of Y → 2 sources:

**Between-Person (BP) Variance:**

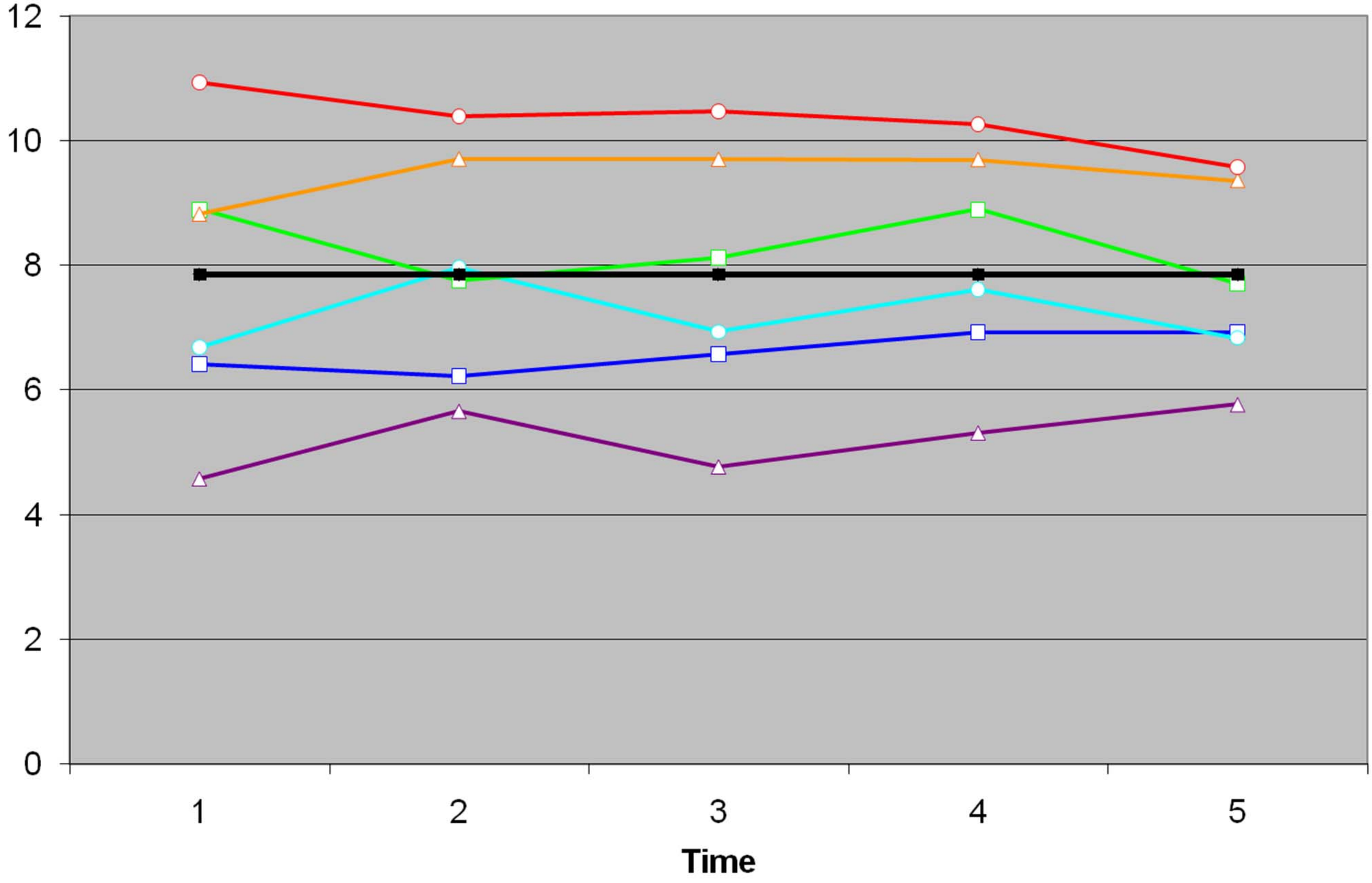
- → Differences from **GRAND** mean
- → **INTER**-Individual Differences

**Within-Person (WP) Variance:**

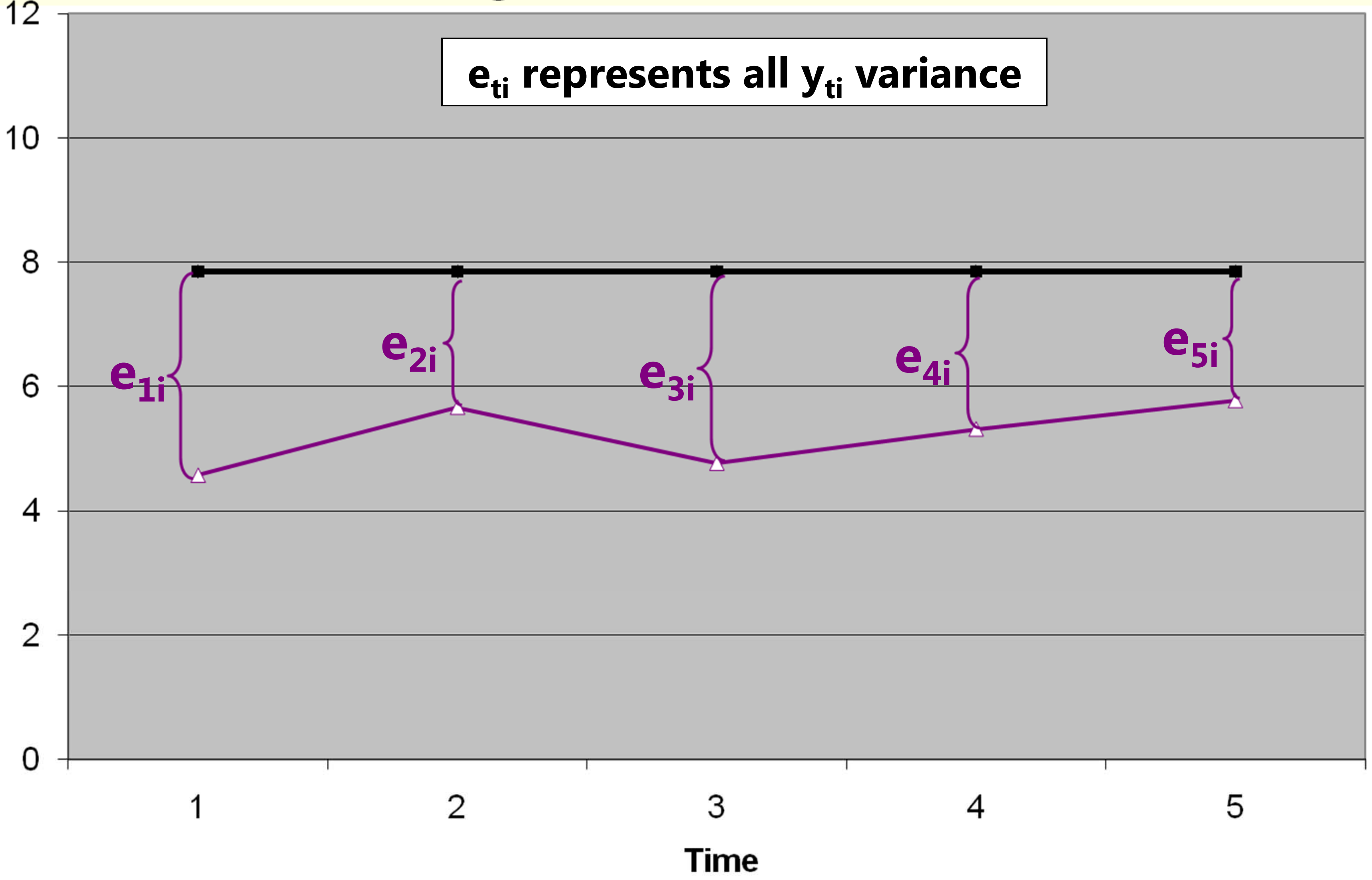
- → Differences from **OWN** mean
- → **INTRA**-Individual Differences
- → This part is only observable through longitudinal data.

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

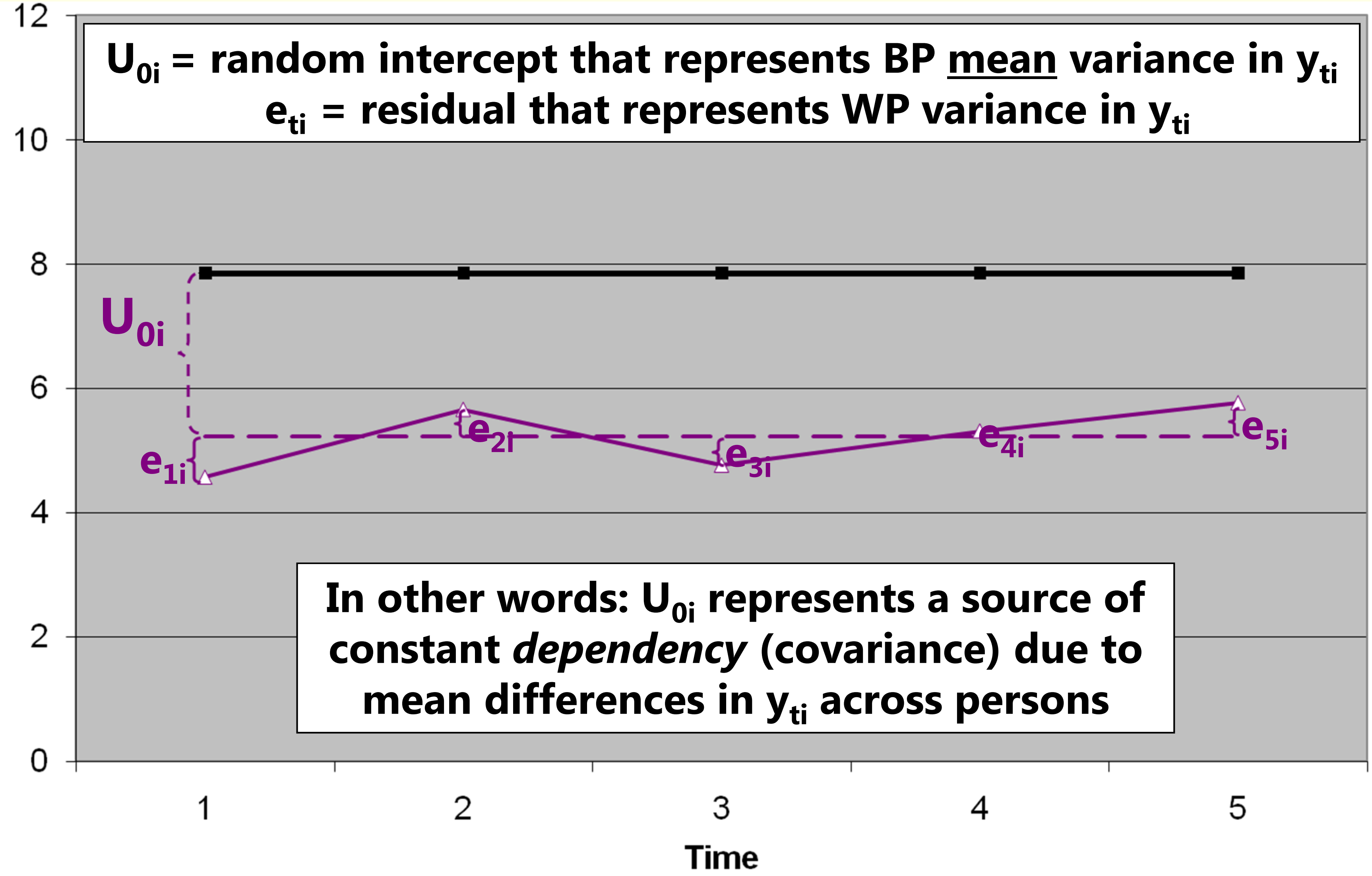
# Hypothetical Longitudinal Data



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



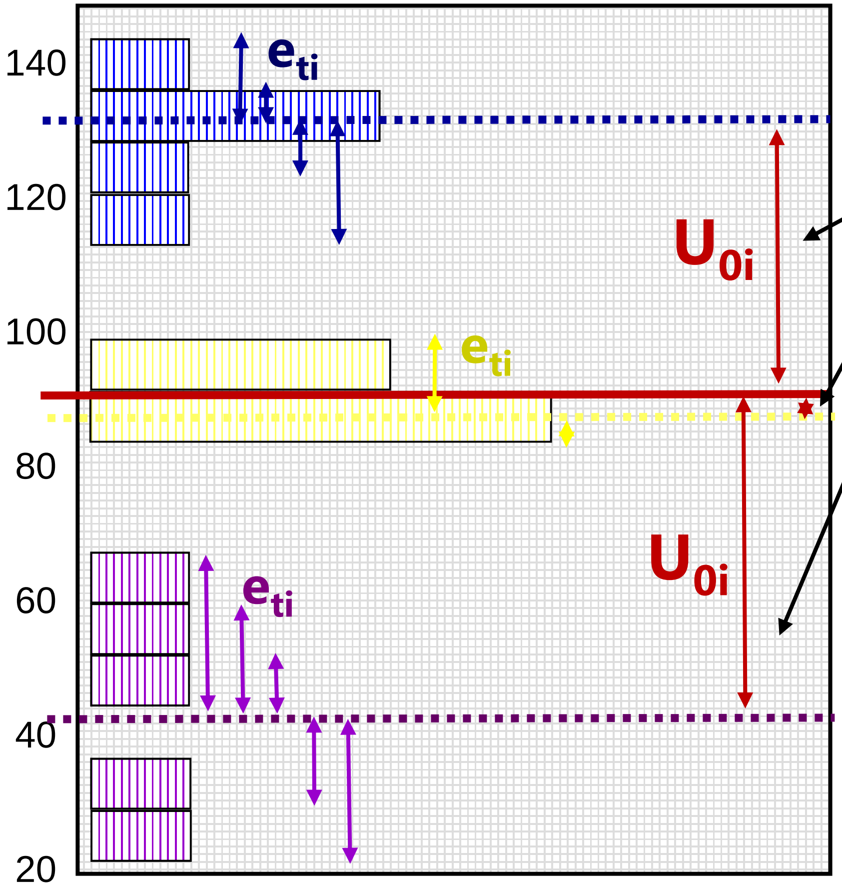
# “Error” in a +WP Model for the Variance: Multilevel Model



**$U_{0i}$ = random intercept that represents BP <u>mean</u> variance in $y_{ti}$**
**$e_{ti}$ = residual that represents WP variance in $y_{ti}$**

**In other words: $U_{0i}$ represents a source of constant *dependency* (covariance) due to mean differences in $y_{ti}$ across persons**

# Empty +Within-Person Model



**$y_{ti}$ variance → 2 sources:**

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

- → **Between**-Person Variance
- → Differences from **GRAND** mean
- → **INTER**-Individual Differences

**<u>Level 1 Residual Variance</u> (of $e_{ti}$, as $\sigma^2_e$):**

- → **Within**-Person Variance
- → Differences from **OWN** mean
- → **INTRA**-Individual Differences

# BP vs. +WP Empty Models

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

$$y_i = \beta_0 + e_i$$

  - $\beta_0$ = fixed intercept = grand mean
  - $e_i$ = residual deviation from GRAND mean

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

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

  - $\beta_0$ = 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):**

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

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

**R matrix** **R CORR Matrix**

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

- 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

# Example Data for BP and WP Models

- 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)* | 54.90 *(1.13)* | 51.99 *(0.89)* |
| Treatment | 50.76 *(0.91)* | 58.62 *(0.99)* | 54.70 *(0.87)* |
| Marginal | 49.92 *(0.73)* | 56.76 *(0.79)* | 53.34 *(0.64)* |

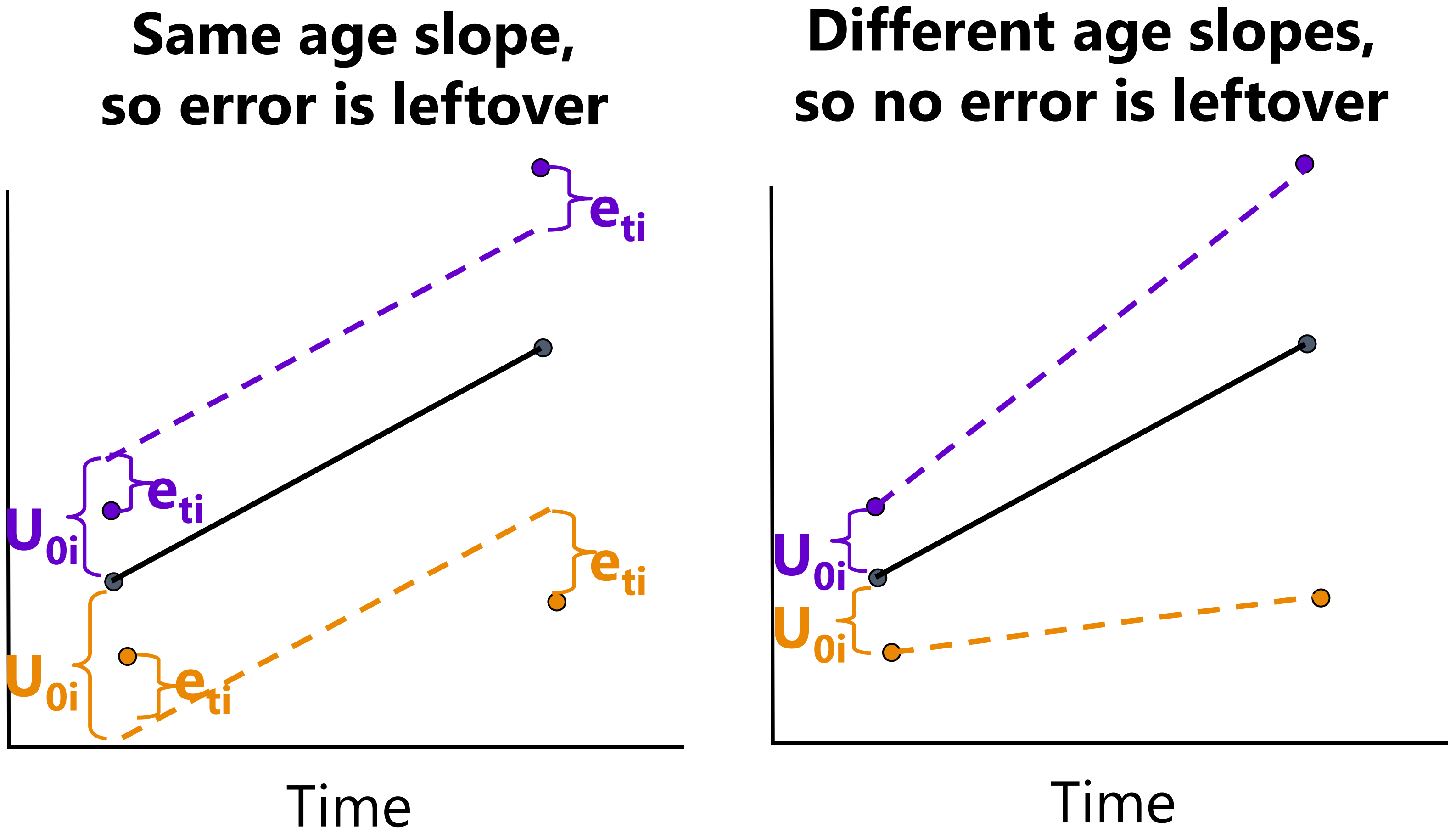
**5.82**

**1.68** **2.04** **3.72** **2.71**

**7.86**

**6.84**

# 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**

# 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": $\beta_0 + \beta_1(T_1) + \beta_2(T_2) + \beta_3(T_3)$**
  - **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 $\sigma_e^2$)
- **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**

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

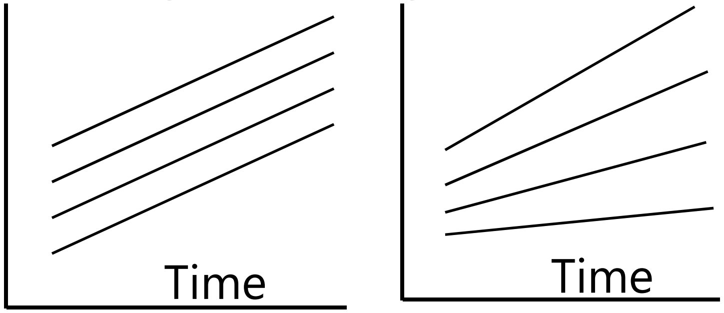
# 2a. Univariate Repeated Measures

- 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 $\left(\tau_{U_0}^2 + \sigma_e^2\right)$ 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 ($\sigma_e^2$ 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

# Comparing Models for the Variance

- Choosing a model for the variance requires assessment of **relative model fit**: how well does the model fit relative to other possible models?

- 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 main questions in choosing a model for the variance:**
    - How does the residual variance differ across occasions?
    - How are the residuals from the same unit correlated?

- Nested models are compared using a **"likelihood ratio test": −2ΔLL test** (aka, "$\chi^2$ 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 $\chi^2$ distribution with df = Δdf**
4. Get *p*-value from CHIDIST in excel or LRTEST option in STATA

# 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

# Flavors of Maximum Likelihood

- Remember that Maximum likelihood comes in 2 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) → $\frac{\sum(\mathrm{y_i} - \mathrm{y_{pred}})^2}{\mathrm{N} - \mathrm{k}}$
- **“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 → $\frac{\sum(\mathrm{y_i} - \mathrm{y_{pred}})^2}{\mathrm{N}}$
- **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

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

- 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

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

# ML vs. REML in a nutshell

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

**"Population"**

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

**"Sample"**

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

| **All comparisons must have same N!!!** | **ML** | **REML** |
|---|---|---|
| To select, type… | METHOD=ML<br>(-2 log likelihood) | METHOD=REML *default*<br>(-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<br>-- *OR* --<br>ML −2ΔLL only (NO REML −2ΔLL) | NO *p*-values<br>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