Introduction to Within-Person Analysis and Kinds of ANOVAs

- Topics:
  - From between-person to within-person models for the variance (from cross-sectional to longitudinal outcomes)
  - Kinds of analyses of variance (ANOVAs) for longitudinal data
  - Comparisons of different models for the variance using likelihood ratio tests: welcome (back) to −2ΔLL!

The Two Sides of a General Linear Model

$$y_i = \beta_0 + \beta_1(x1_i) + \beta_2(x2_i) + \cdots + e_i$$

Our focus now

- **Model for the Means (→ Predicted Values):**
  - Each person's expected (predicted) outcome is a weighted linear function of his/her values on $x1_i$ and $x2_i$ (and any other predictors); each variable is measured once per person
  - **Estimated constants are called fixed effects** (here, β_0, β_1, and β_2)
- **Model for the Variance (→ "Piles" of Variance):**
  - $e_i \sim \mathrm{N}(0, \sigma_e^2)$→ ONE (BP) source of residual (unexplained) error
  - In GLMs, e_i has a mean of 0 with some estimated constant variance σ_e^2, is normally distributed, is unrelated to $x1_i$ and $x2_i$, and is **independent** across all observations (which is just one outcome per person here)
  - **There is only ONE source of residual variance in the above GLM because it was designed for only ONE (BP) dimension of sampling!**

Means, Variances, Covariances, and Correlations

Using population notation: N = # persons, i = person

(Arithmetic) Mean (μ):
Central tendency of y_i

$$\mu_i = \frac{\sum_{i=1}^{N} y_i}{N}$$

Variance (Var):
Dispersion of y_i
in squared units

$$Var(y_i) = \sigma_y^2 = \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N}$$

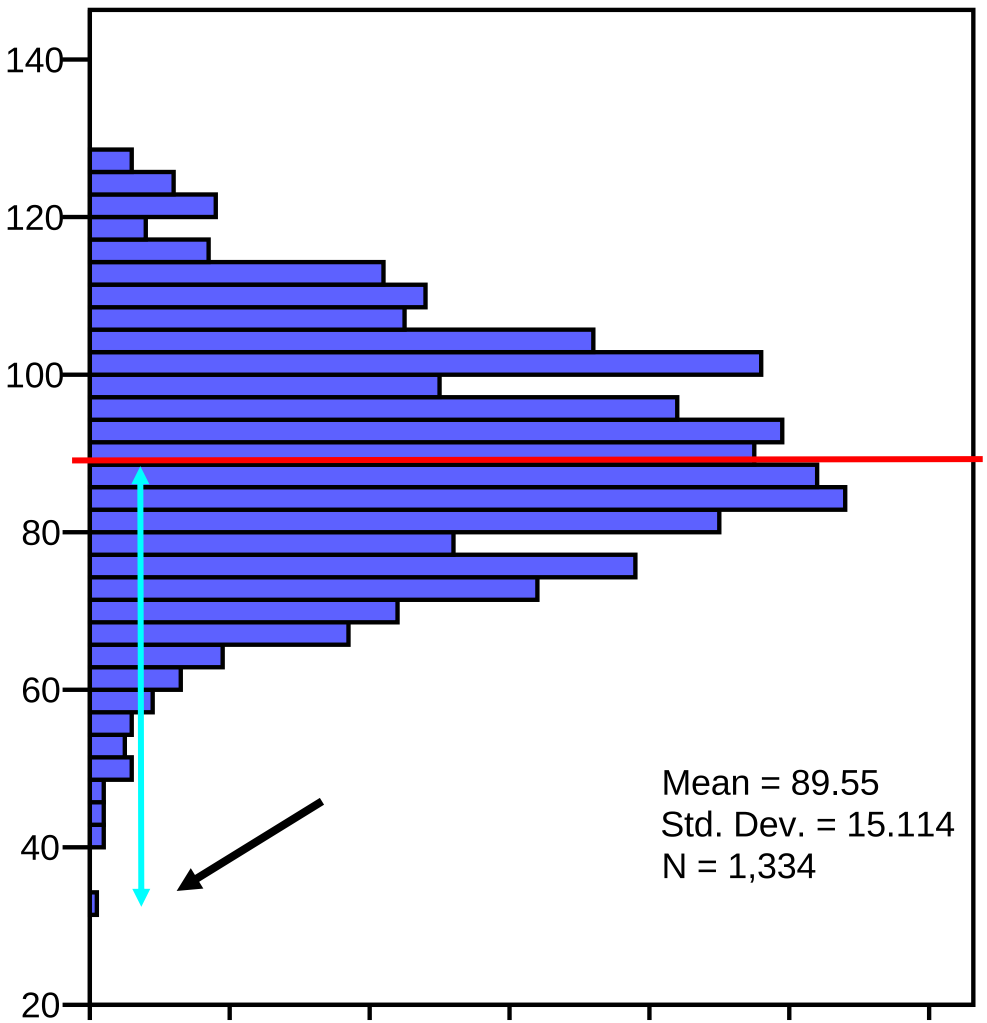
Covariance (Cov):
How outcomes (e.g., y_{1i} and y_{2i}) go together in original metrics (unstandardized)

$$Cov(y_{1i}, y_{2i}) = \sigma_{y_1,y_2} = \frac{\sum_{i=1}^{N}[(y_{1i}-\widehat{y_1})(y_{2i}-\widehat{y_2})]}{N}$$

Correlation (r):
Covariance that has been standardized: -1 to 1

$$r(y_{1i}, y_{2i}) = \frac{Cov(y_{1i}, y_{2i})}{\sqrt{Var(y_{1i})}\sqrt{Var(y_{2i})}}$$

An "Empty Means" General Linear Model → Single-Level Model for the Variance



$$y_i = \beta_0 + e_i$$

Filling in values:

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

Model for the Means

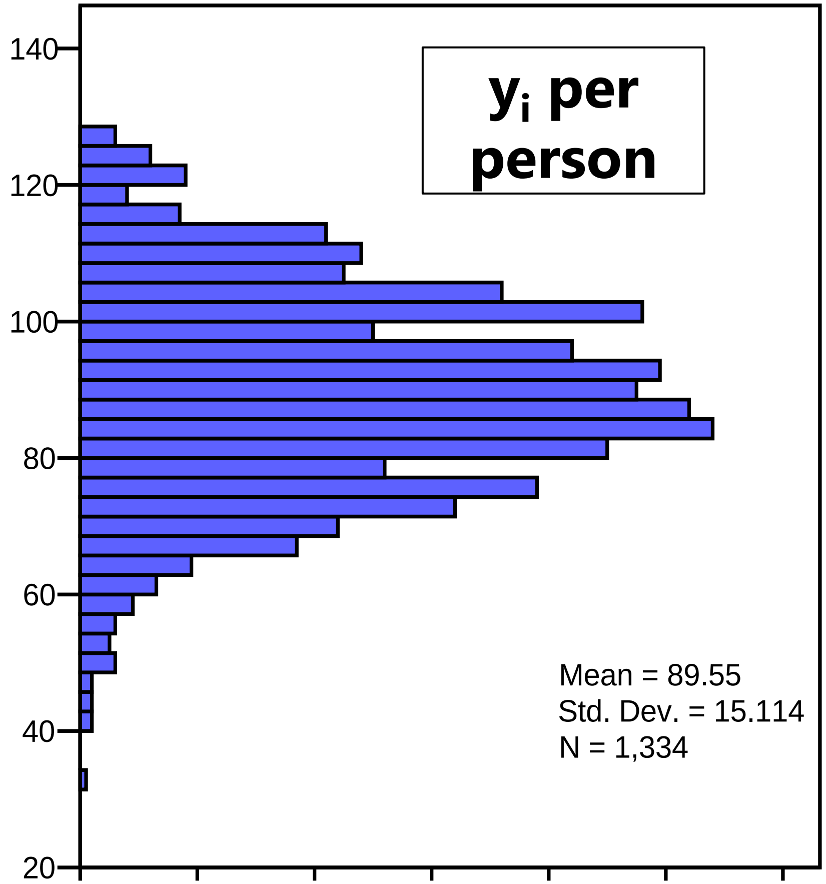
$\hat{y}_i$ **= "y-hat" model-predicted outcome**

y_i residual ("error") variance:

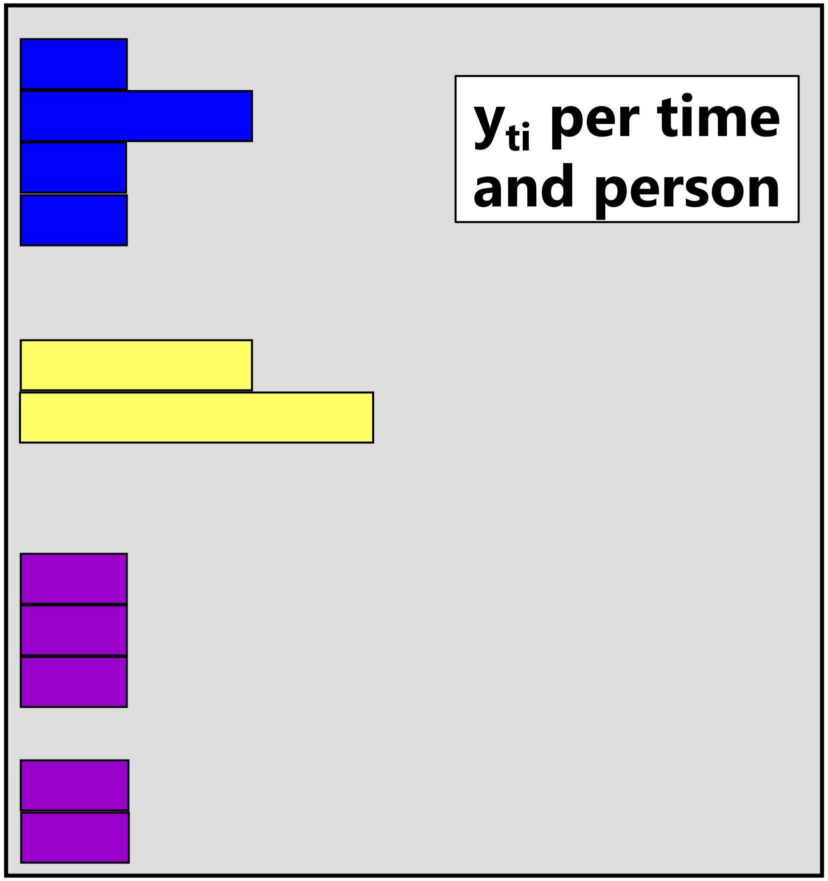
$$\frac{\sum (y_i - \hat{y}_i)^2}{N - 1}$$

Adding Repeated Occasions → Two-Level Model for the Variance

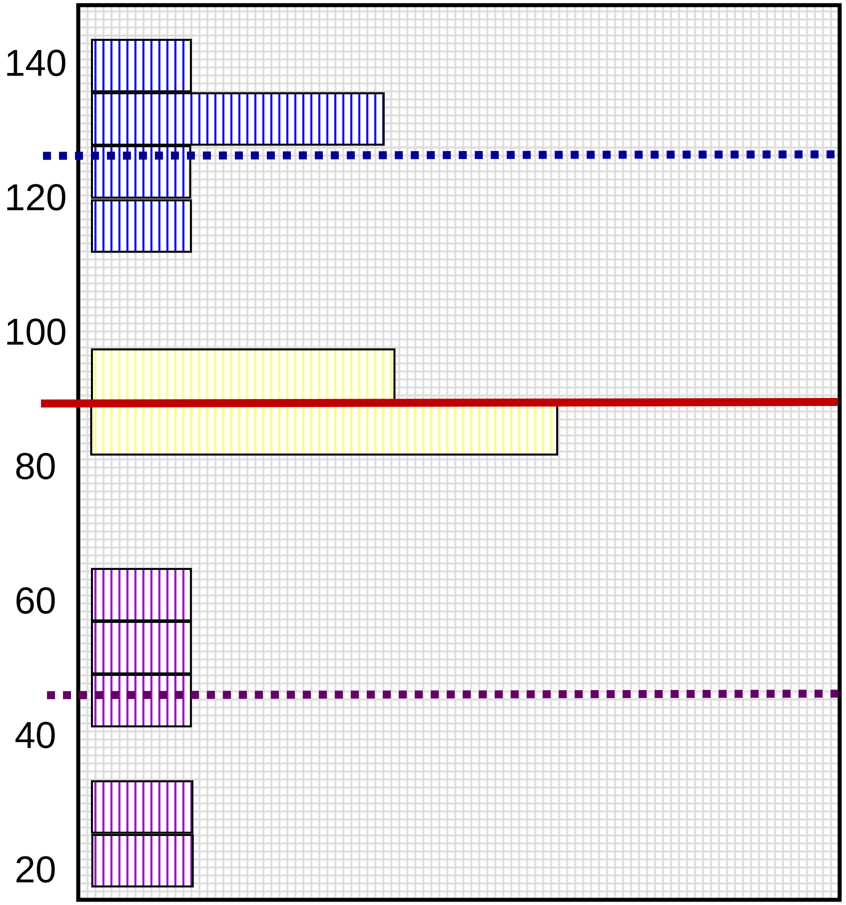
Full Sample Distribution



5 Occasions (t); 3 People (i)



Empty Means Two-Level Model



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

= Grand Mean

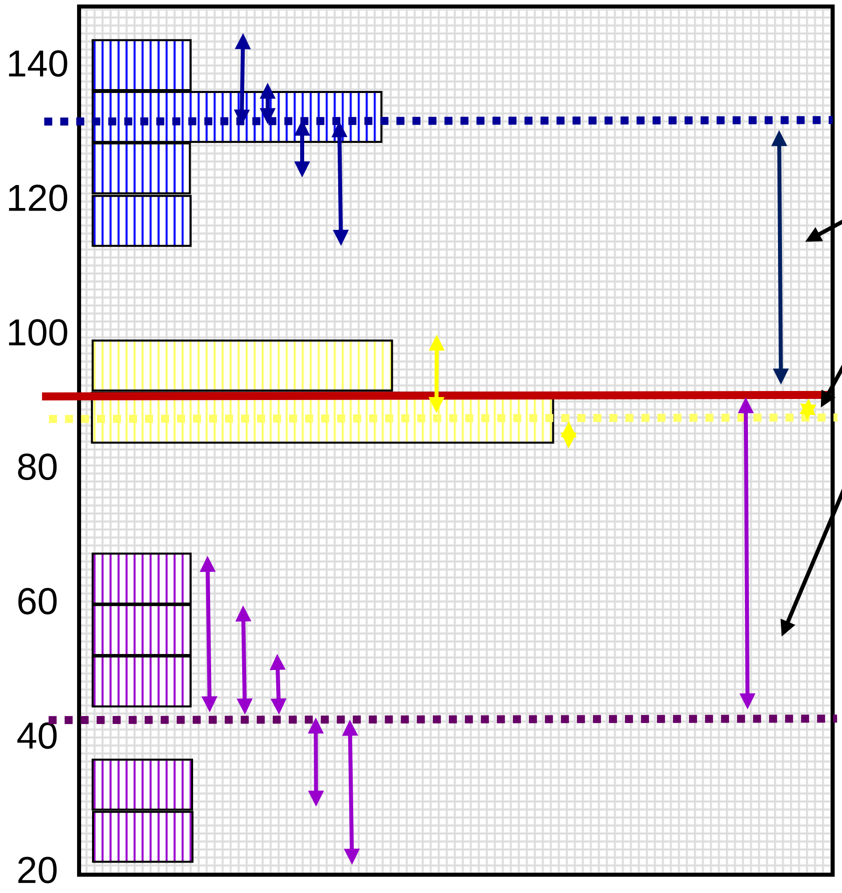
= **Fixed Intercept**

Can make better guess by taking advantage of repeated observations:

= Person Mean

→ Random Intercept

Empty Means Two-Level Model



Variance of y_{ti} → 2 sources:

Between-Person (BP) Variance:

→ Differences from **GRAND** mean

→ **INTER**-Individual Differences

Within-Person (WP) Variance:

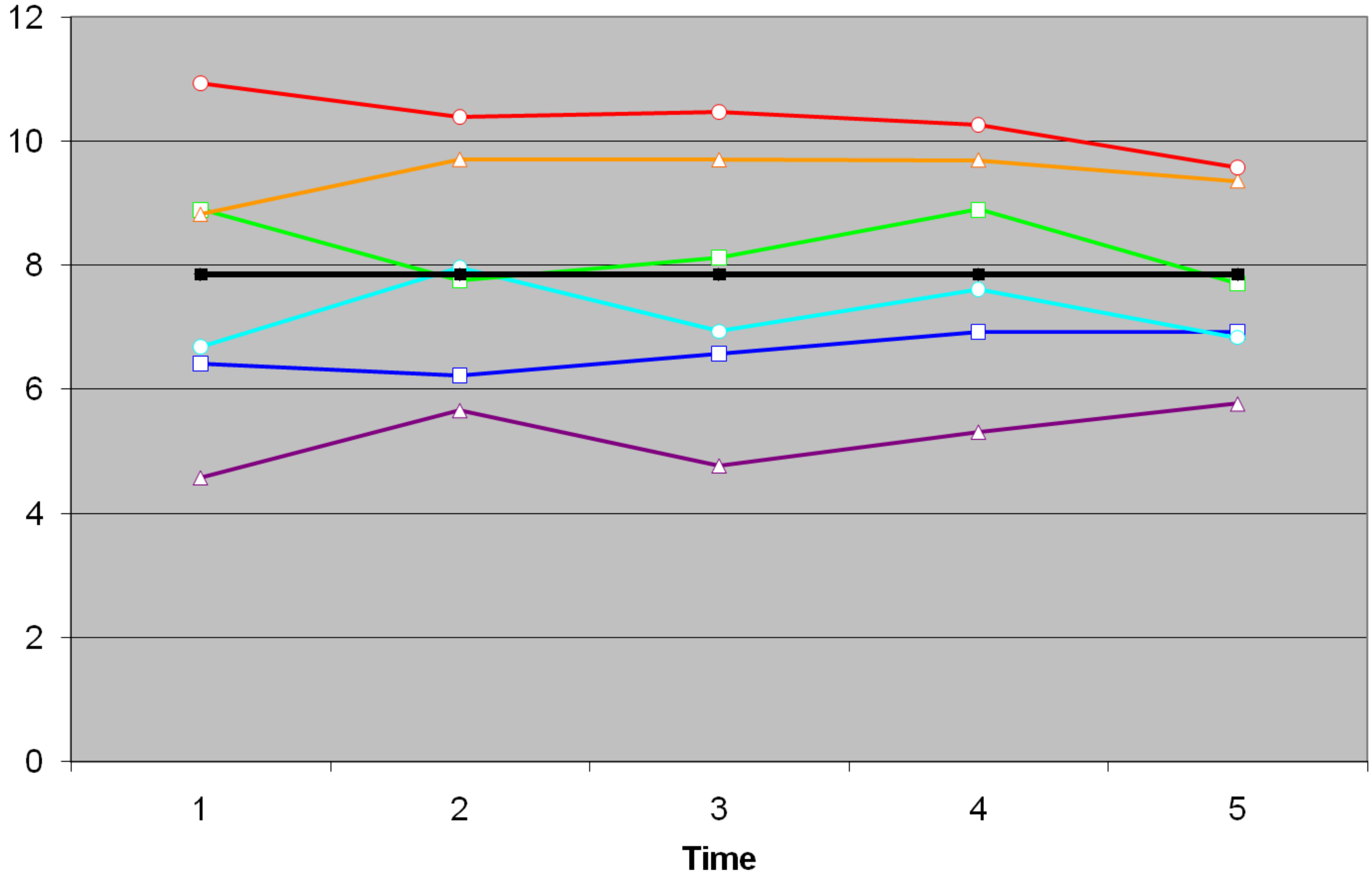
→ Differences from **OWN** mean

→ **INTRA**-Individual Differences

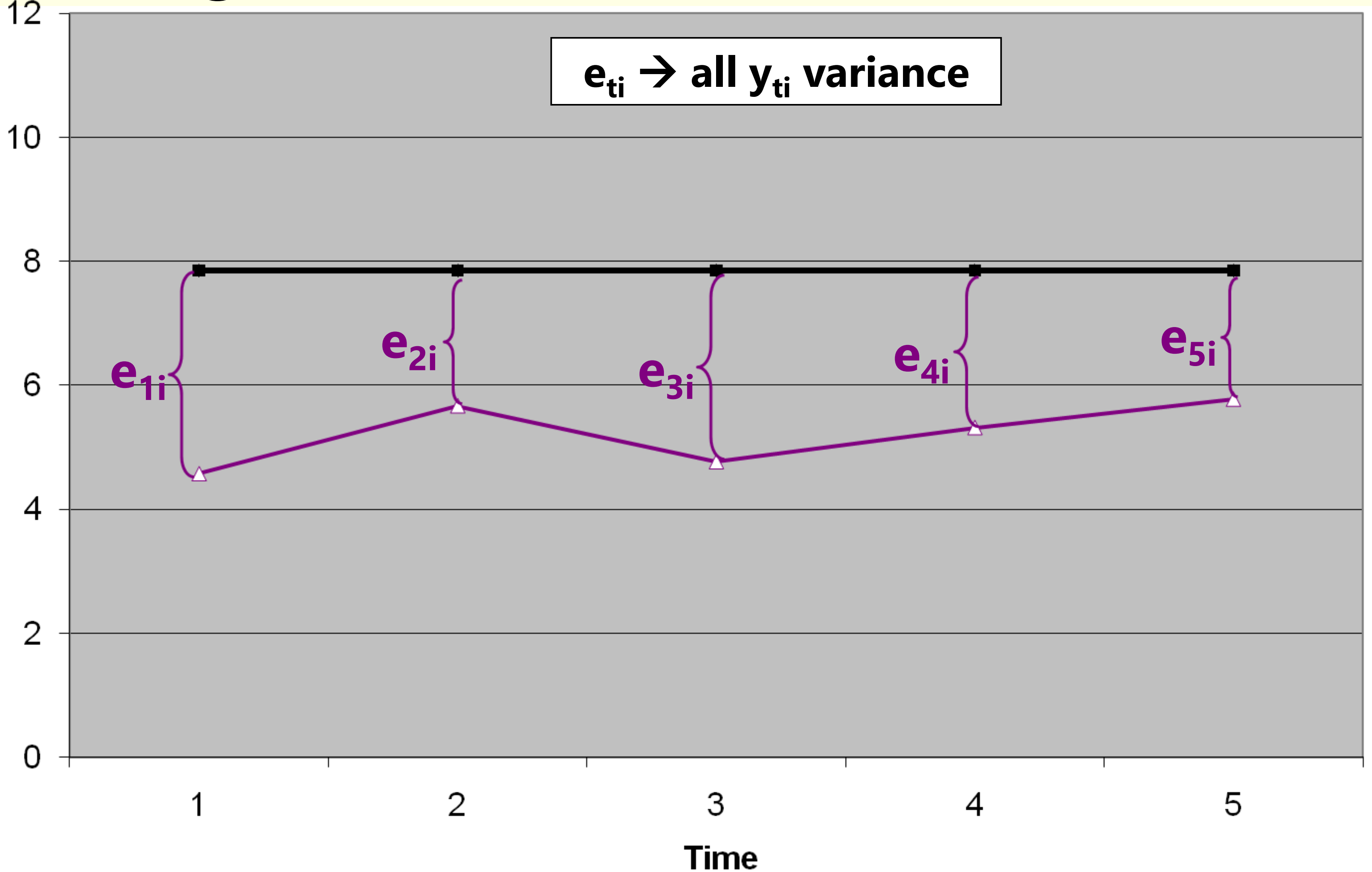
→ This part is only observable through longitudinal data.

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

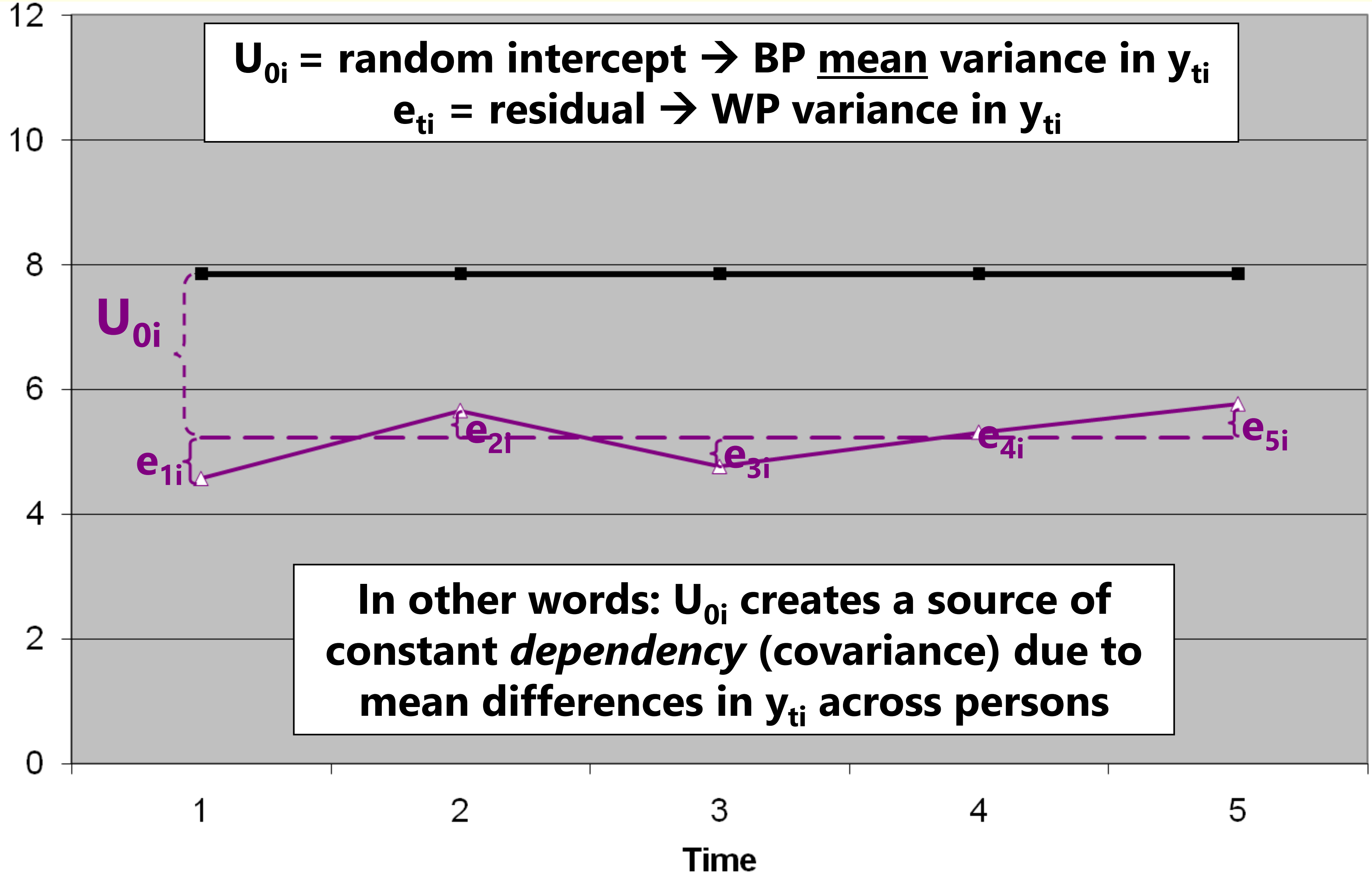
Hypothetical Longitudinal Data



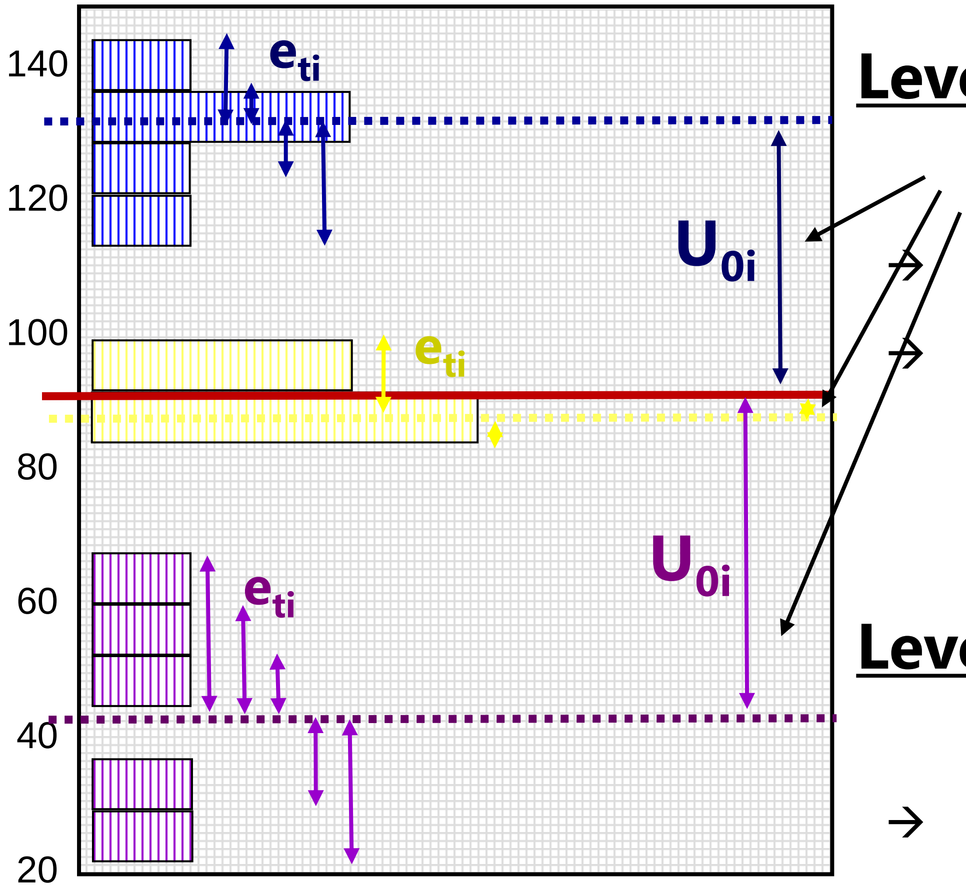
Only One Kind of "Error" in a Single-Level Model for the Variance



Two Distinct Kinds of "Error" in a Two-Level Model for the Variance



Empty Means, Two-Level Model



y_{ti} variance → 2 sources:

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

- → **Between**-Person variance in means
- → **INTER**-Individual differences from **GRAND** mean to be explained by time-invariant predictors

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

- → **Within**-Person variance
- → **INTRA**-Individual differences from **OWN** mean to be explained by time-varying predictors

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

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

$$y_i = \beta_0 + e_i$$

  - β_0 = fixed intercept = sample mean across all outcomes
  - e_i = residual deviation from sample mean

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

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

  - β_0 = fixed intercept = sample mean **of person means**
  - U_{0i} = random intercept = person mean deviation from sample mean
  - e_{ti} = time-specific residual deviation from person mean

Intraclass Correlation (ICC)

Intraclass Correlation (ICC; also known as "ICC1"):

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

ICC =

$$r(y_{1i}, y_{2i}) = \frac{Cov(y_{1i}, y_{2i})}{\sqrt{Var(y_{1i})}\sqrt{Var(y_{2i})}}$$

R matrix	**R CORR Matrix**
$\begin{bmatrix} \sigma^2_e + \tau^2_{u_0} & \tau^2_{u_0} & \tau^2_{u_0} \\ \tau^2_{u_0} & \sigma^2_e + \tau^2_{u_0} & \tau^2_{u_0} \\ \tau^2_{u_0} & \tau^2_{u_0} & \sigma^2_e + \tau^2_{u_0} \end{bmatrix}$	$\begin{bmatrix} 1 & \text{ICC} & \text{ICC} \\ \text{ICC} & 1 & \text{ICC} \\ \text{ICC} & \text{ICC} & 1 \end{bmatrix}$

- 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)**

Conditional Models: Single- vs Two-Level

- Single Outcome, **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

Example Data for BP and WP Models

- 50 kids per 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

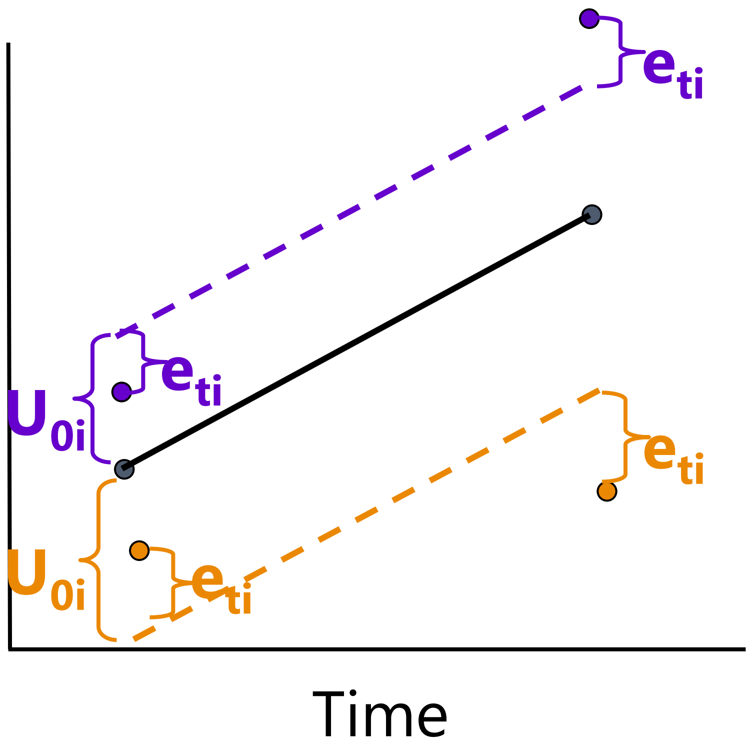
1.68 **2.04** **3.72** **2.71**

7.86

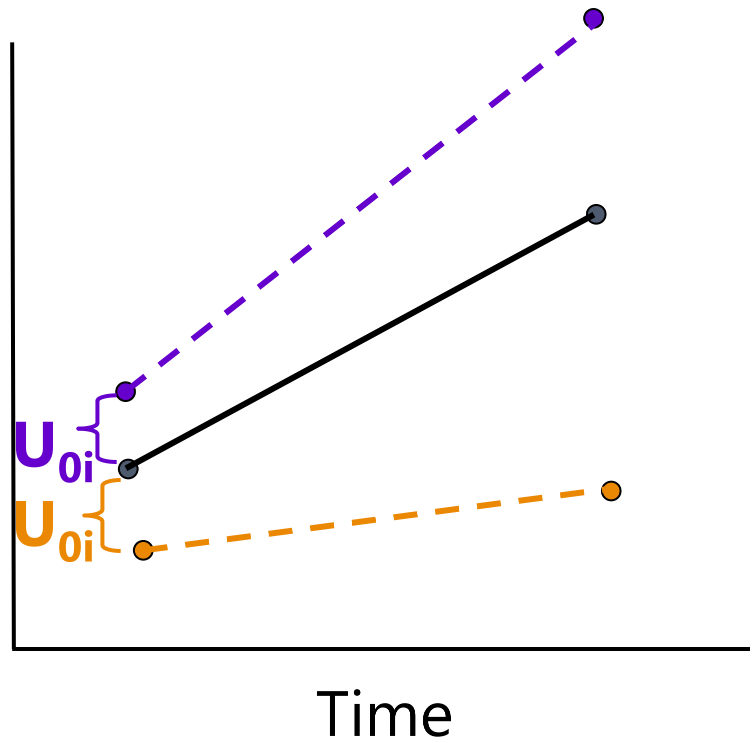
6.84

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

Same age slope, so error is leftover



Different time slopes, so no error is leftover



More About Fixed vs. Random Factors

- Technically speaking, in longitudinal samples, occasions are crossed with persons, both at level 2
  - Mean differences across persons (→ random intercept, usually)
  - Mean differences across occasions (→ fixed effects, usually)
  - Person*occasion interaction is level 1, not separable from residual error

- Because we almost always model mean differences across occasions using fixed effects, the occasion level-2 crossed dimension is then eliminated (~no variability should remain)
  - Level-2 model then includes only between-person differences (in intercepts at a minimum, likely in predictor slopes like time as well)
  - Level-1 model is actually person*occasion variance, which we abbreviate as "within-person" (i.e., "occasions nested in persons")

- Alternatively, intensive longitudinal designs may keep mean differences across occasions as a crossed random factor, like in this example study

ANOVA for longitudinal data?

- There are 3 possible "kinds" of ANOVA models we could use:
  - Between-Persons/Groups, Univariate RM, and Multivariate RM
- **THEY DO NOT ALLOW:**
  - **Missing occasions** (do listwise deletion when using least squares)
  - **Time-varying predictors** (covariates are BP predictors only)
- Each includes the same "**saturated**" model for the means → treats **time as categorical**: (# fixed effects = # occasions)
  - e.g., for four occasions: $\beta_0 + \beta_1(T_{1i}) + \beta_2(T_{2i}) + \beta_3(T_{3i})$
  - **The *time* predictor must be balanced and discrete in ANOVA!**
- ANOVA models differ by their **model for the variance**…
  - i.e., **how they "handle person dependency"** → what pattern they predict for the variance and covariance of the y_{ti} residuals across occasions…

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**" in SAS or "independent" in STATA (R varies), also called a "**diagonal**" matrix:

R matrix

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

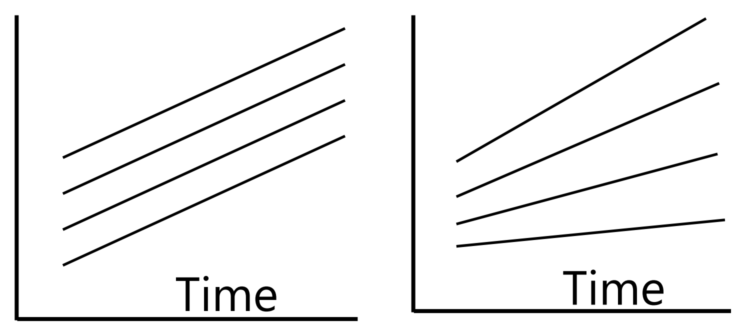
2a. Univariate Repeated Measures

- Separates total variance into **two** sources (**piles of variance**):
  - **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**" in SAS or "**exchangeable**" in STATA (R varies):
  - **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" should be the same across occasions, 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 whether data meet sphericity
  - **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 tweaked 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)
  - Time is categorical: Fixed effects for 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 (the way that 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, piecewise slopes)
  - 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, some R: **−2LL** gives BADNESS of fit, so **smaller** value = better model

- Given as just log-likelihood (LL) in STATA MIXED and Mplus, some R: **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, "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, use $-2\Delta LL = (-2LL_{fewer}) - (-2LL_{more})$
   if given LL, use $-2\Delta LL = -2 * (LL_{fewer} - LL_{more})$
2. Calculate **ΔDF** = $(\#\ Parms_{more}) - (\#\ Parms_{fewer})$
3. **Compare −2ΔLL to χ^2 distribution with numerator DF = ΔDF**
4. Get *p*-value (from CHIDIST in excel, LRTEST in STATA, ANOVA in R)

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(#size)*(#parameters)
  - What "#parameters" means depends on estimator in SAS: ML = ALL parameters; REML = variance model parameters only
    - In R and STATA, #parameters = ALL parameters regardless
  - What "#size" means differs by program: #size = level-2 N in SAS, but #size is total number of observations in R and STATA

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 (key: based on normally distributed residuals at all levels of analysis)
  - **REML = OLS** given complete outcomes, but it doesn't require them
  - Estimates variances the same way as in OLS (accurate) → $\dfrac{\sum(y_i - \hat{y}_i)^2}{N-k}$
- **"Maximum likelihood" (ML; also called FIML*)**
  - Is more general and is available for all of the above, as well as for non-normal outcomes and models with latent variables (CFA/SEM/IRT/DCM)
  - Is NOT the same as OLS: it under-estimates variances by not accounting for number of estimated fixed effects → $\dfrac{\sum(y_i - \hat{y}_i)^2}{N}$
- **FI = Full information → it uses all original data (they both do)*

Rules for LRTs by 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

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

- Maximum likelihood (**ML**; also used in MIXED)
  - Variances (and SEs) are too small in small N (< 100 units)
  - 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

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

LRTs using ML vs. REML in a nutshell

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** (matters less after N=50–100 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 Effects) Only**	**Variance Model (Random Effects) Only**	**Both Means and Variances Model (Fixed and Random)**
<u>Nested?</u> 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)
<u>Non-Nested?</u> 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)

<u>Nested</u> = one model is a <u>direct subset</u> of the other

<u>Non-Nested</u> = one model is <u>not a direct subset</u> 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 (AIC and BIC only)

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 (REML is better) −2ΔLL tests, no Wald *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