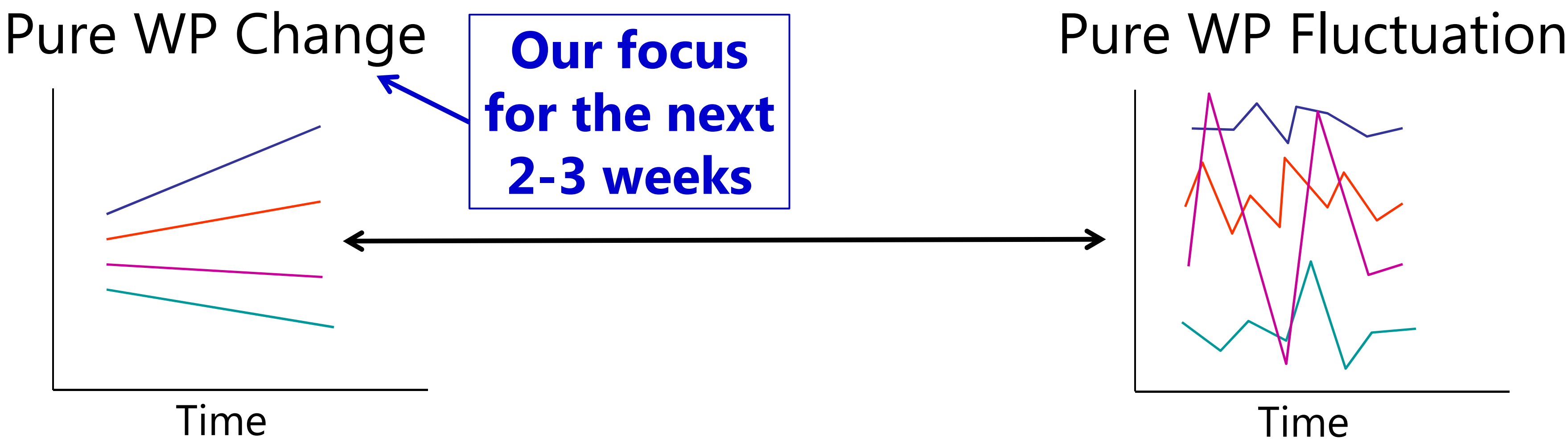
Introduction to Random Effects of Time and Model Estimation

- Today's Class:
  - The Big Picture
  - Multilevel model notation
  - Fixed vs. random effects of time
  - Random intercept vs. random slope models
  - Fun with likelihood estimation

Modeling Change vs. Fluctuation



Model for the Means:

- **WP Change** → describe pattern of *average* change (over "time")
- WP Fluctuation → *may* not need anything (if no systematic change)

Model for the Variances:

- **WP Change** → describe *individual differences* in change (random effects)
  → this allows variances and covariances to differ over time
- WP Fluctuation → describe pattern of variances and covariances over time

The Big Picture of Longitudinal Data: Models for the Means

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

Parsimony

Empty Model:
Predicts NO change over time

1 Fixed Effect

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

Saturated Means:
Reproduces mean at each occasion

 # Fixed Effects = # Occasions

Good fit

Name… that… Trajectory!

The Big Picture of Longitudinal Data: Models for the Variance

<u>Compound Symmetry (CS)</u>

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

Most useful model: likely somewhere in between!

<u>Unstructured (UN)</u>

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

Parsimony ⟵⟶ **Good fit**

Univariate RM ANOVA

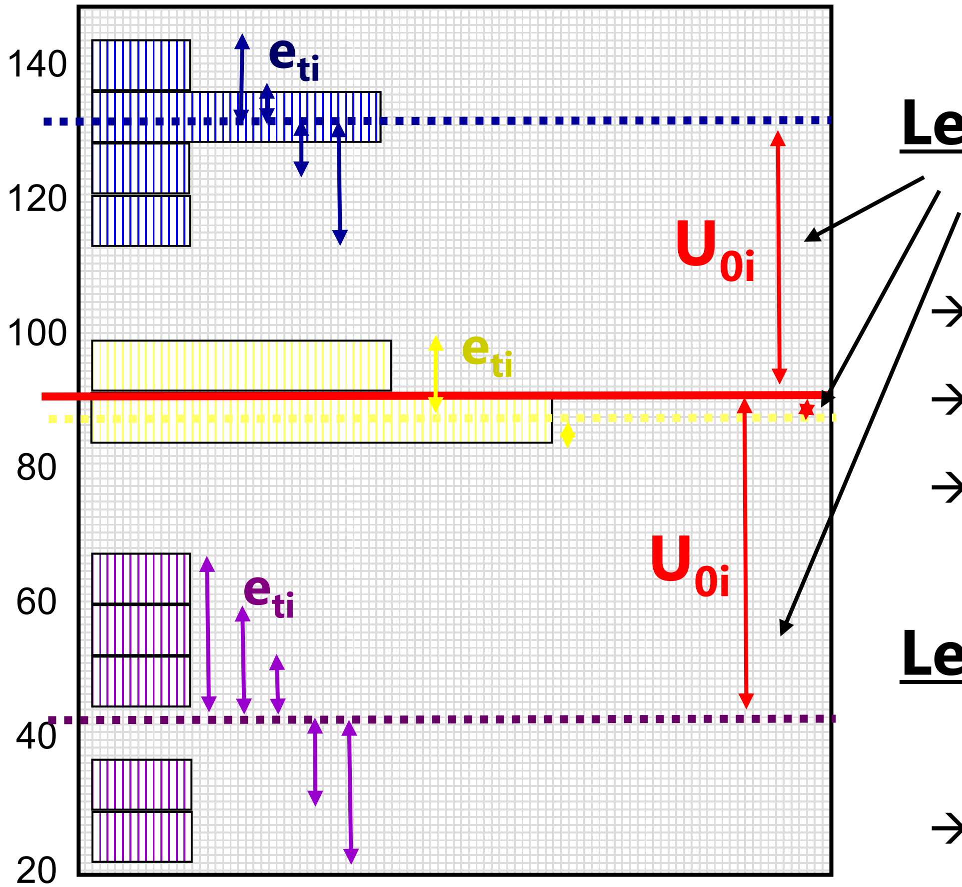
NAME ...THAT ... STRUCTURE!

Multivariate RM ANOVA

What is the pattern of variance and covariance over time?

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

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

Empty Means, Random Intercept Model

GLM Empty Model:

- $y_i = \beta_0 + e_i$

MLM Empty Model:

- Level 1:

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

- Level 2:

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

Fixed Intercept =grand mean (because no predictors yet)

Random Intercept = individual-specific deviation from predicted intercept

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

3 Total Parameters:

Model for the Means (1):

- Fixed Intercept γ_{00}

Model for the Variance (2):

- Level-1 Variance of e_{ti} → σ^2_e
- Level-2 Variance of U_{0i} → $\tau^2_{U_0}$

Composite equation:

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

Saturated Means, Random Intercept Model

- Although rarely shown this way, a saturated means, random intercept model would be represented as a multilevel model like this (for example *n* = 4 here, in which the time predictors are dummy codes to distinguish each occasion from time 0):
- Level 1:

$$y_{ti} = \beta_{0i} + \beta_{1i}(Time1_{ti}) + \beta_{2i}(Time2_{ti}) + \beta_{3i}(Time3_{ti}) + e_{ti}$$

- Level 2:

$$\beta_{0i} = \gamma_{00} + U_{0i}$$
$$\beta_{1i} = \gamma_{10}$$
$$\beta_{2i} = \gamma_{20}$$
$$\beta_{3i} = \gamma_{30}$$

Composite equation (6 parameters):

$$y_{ti} = \gamma_{00} + \gamma_{10}(Time1_{ti}) + \gamma_{20}(Time2_{ti}) + \gamma_{30}(Time3_{ti}) + U_{0i} + e_{ti}$$

Given the same random intercept model for the variance, the **G**, **R**, and **V** matrices would have the same form for the empty means model as for the saturated means model (but the latter would estimate remaining variance and covariance after controlling for all possible mean differences over time).

Matrices in a Random Intercept Model

RI and DIAG: Total predicted data matrix is called V matrix, created from the G [TYPE=UN] and R [TYPE=VC] matrices as follows:

$$\mathbf{V} = \mathbf{Z} * \mathbf{G} * \mathbf{Z}^T + \mathbf{R} = \mathbf{V}$$

$$\mathbf{V} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} \tau^2_{U_0} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix} + \begin{bmatrix} \sigma^2_e & 0 & 0 & 0 \\ 0 & \sigma^2_e & 0 & 0 \\ 0 & 0 & \sigma^2_e & 0 \\ 0 & 0 & 0 & \sigma^2_e \end{bmatrix} = \begin{bmatrix} \tau^2_{U_0} + \sigma^2_e & \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} \\ \tau^2_{U_0} & \tau^2_{U_0} + \sigma^2_e & \tau^2_{U_0} & \tau^2_{U_0} \\ \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} + \sigma^2_e & \tau^2_{U_0} \\ \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} + \sigma^2_e \end{bmatrix}$$

VCORR then provides the intraclass correlation, calculated as:
ICC = $\tau^2_{U_0} / (\tau^2_{U_0} + \sigma^2_e)$

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

assumes a constant correlation over time

For any random intercept model:
VCORR provides the "unconditional" ICC when requested from an **empty means** model. When paired with any other kind of means model (e.g., **saturated means** model), **VCORR** provides a "conditional" ICC instead (after controlling for fixed effects).

Augmenting the empty means, random intercept model with *time*

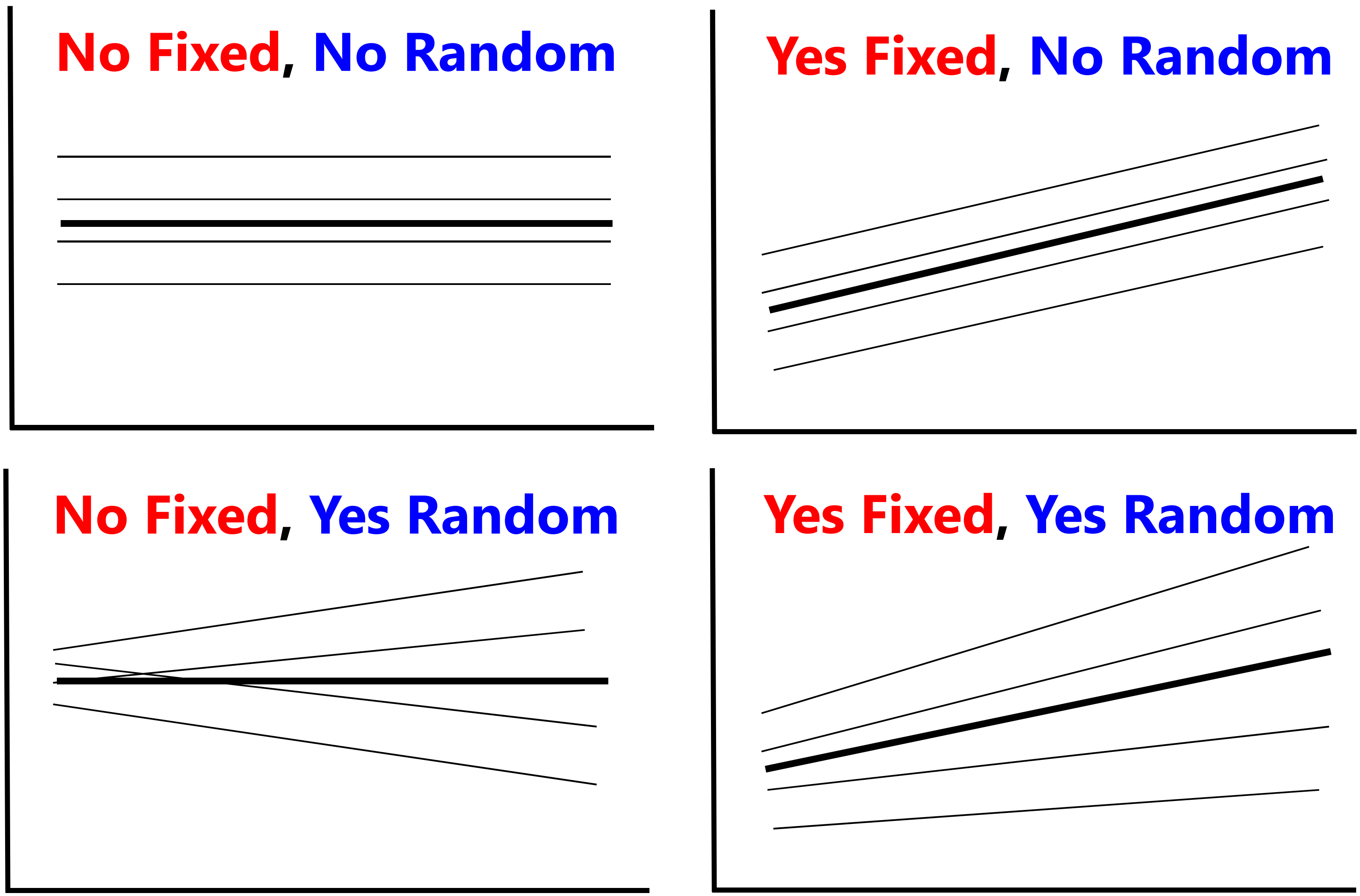
- 2 questions about the possible effects of *time*:

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

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

Fixed and Random Effects of Time

(Note: The intercept is random in every figure)



Fixed Linear Time, Random Intercept Model

(4 total parameters: effect of time is FIXED only)

Residual = time-specific deviation from individual's predicted outcome → estimated variance of σ_e^2

Multilevel Model

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

Fixed Intercept = predicted mean outcome at time 0

Fixed Linear Time Slope = predicted mean rate of change per unit time

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

Random Intercept = individual-specific deviation from fixed intercept → estimated variance of $\tau_{U_0}^2$

Composite Model

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

Because the effect of time is **fixed**, everyone is predicted to change at exactly the same rate.

Explained Variance from Fixed Linear Time

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

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

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

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

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

Increases in Random Intercept Variance

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

Random Intercept Models Imply…

- **People differ from each other systematically in only ONE way**—in intercept (U_{0i}), which implies **ONE kind of BP variance**, which translates to **ONE source of person dependency** (covariance or correlation in the outcomes from the same person)
- If so, after controlling for BP intercept differences (by estimating the variance of U_{0i} as $\tau^2_{U_0}$ in the **G** matrix), the $\mathbf{e_{ti}}$ **residuals** (whose variance and covariance are estimated in the **R** matrix) should be **uncorrelated with homogeneous variance across time**, as shown:

Level-2 **G** matrix: RANDOM TYPE=UN

$$\begin{bmatrix} \tau^2_{U_0} \end{bmatrix}$$

Level-1 **R** matrix: REPEATED TYPE=VC

$$\begin{bmatrix} \sigma^2_e & 0 & 0 & 0 \\ 0 & \sigma^2_e & 0 & 0 \\ 0 & 0 & \sigma^2_e & 0 \\ 0 & 0 & 0 & \sigma^2_e \end{bmatrix}$$

G and **R** matrices combine to create a total **V** matrix with CS pattern

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

Fixed Linear Time, Random Intercept Model (4 total parameters: effect of time is FIXED only)

- How the model predicts each element of the **V** matrix:

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

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

$\mathbf{\beta_{1i} = \gamma_{10}}$

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

Predicted Variance per *Time*:

$$\mathrm{Var}\left[y_{\mathrm{Time}}\right]$$

$$= \mathrm{Var}\left[\left(\gamma_{00} + U_{0i}\right) + \left(\gamma_{10}\right)\left(\mathrm{Time}\right) + e_{ti}\right]$$

$$= \mathrm{Var}\left[U_{0i} + e_{ti}\right]$$

$$= \tau^2_{U_0} + \sigma^2_e$$

Predicted Covariance:

$$\mathrm{Cov}\left[y_A, y_B\right]$$

$$= \mathrm{Cov}\left[\left(\gamma_{00} + U_{0i}\right) + \left(\gamma_{10}\right)\left(A\right) + e_{ti}\right], \left[\left(\gamma_{00} + U_{0i}\right) + \left(\gamma_{10}\right)\left(B\right) + e_{ti}\right]$$

$$= \mathrm{Cov}\left[U_{0i}\right], \left[U_{0i}\right]$$

$$= \tau^2_{U_0}$$

Fixed Linear Time, Random Intercept Model
(4 total parameters: effect of time is FIXED only)

- Scalar "mixed" model equation per person:

$$\mathbf{Y}_i = \mathbf{X}_i * \boldsymbol{\gamma} + \mathbf{Z}_i * \mathbf{U}_i + \mathbf{E}_i$$

$$\begin{bmatrix} y_{0i} \\ y_{1i} \\ y_{2i} \\ y_{3i} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} \gamma_{00} \\ \gamma_{10} \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} U_{0i} \end{bmatrix} + \begin{bmatrix} e_{0i} \\ e_{1i} \\ e_{2i} \\ e_{3i} \end{bmatrix}$$

$$\begin{bmatrix} y_{0i} \\ y_{1i} \\ y_{2i} \\ y_{3i} \end{bmatrix} = \begin{bmatrix} \gamma_{00} + \gamma_{10}(0) \\ \gamma_{00} + \gamma_{10}(1) \\ \gamma_{00} + \gamma_{10}(2) \\ \gamma_{00} + \gamma_{10}(3) \end{bmatrix} + \begin{bmatrix} U_{0i} \\ U_{0i} \\ U_{0i} \\ U_{0i} \end{bmatrix} + \begin{bmatrix} e_{0i} \\ e_{1i} \\ e_{2i} \\ e_{3i} \end{bmatrix}$$

$$\begin{bmatrix} y_{0i} \\ y_{1i} \\ y_{2i} \\ y_{3i} \end{bmatrix} = \begin{bmatrix} \gamma_{00} + \gamma_{10}(0) + U_{0i} + e_{0i} \\ \gamma_{00} + \gamma_{10}(1) + U_{0i} + e_{1i} \\ \gamma_{00} + \gamma_{10}(2) + U_{0i} + e_{2i} \\ \gamma_{00} + \gamma_{10}(3) + U_{0i} + e_{3i} \end{bmatrix}$$

$\mathbf{X}_i$ = *n* x *k* values of **predictors with fixed effects**, so can differ per person (*k* = 2: intercept, linear time)

$\boldsymbol{\gamma}$ = *k* x 1 estimated **fixed effects**, so will be the same for all persons (γ_{00} = intercept, γ_{10} = linear time)

$\mathbf{Z}_i$ = *n* x *u* values of **predictors with random effects**, so can differ per person (*u* = 1: intercept)

$\mathbf{U}_i$ = *u* x 1 estimated individual **random effects**, so can differ per person

$\mathbf{E}_i$ = *n* x *n* time-specific residuals, so can differ per person

Fixed Linear Time, Random Intercept Model
(4 total parameters: effect of time is FIXED only)

- Predicted total variances and covariances per person:

$$\mathbf{V}_i = \mathbf{Z}_i * \mathbf{G}_i * \mathbf{Z}_i^T + \mathbf{R}_i$$

$$\mathbf{V}_i = \begin{bmatrix}1\\1\\1\\1\end{bmatrix}\begin{bmatrix}\tau^2_{U_0}\end{bmatrix}\begin{bmatrix}1 & 1 & 1 & 1\end{bmatrix} + \begin{bmatrix}\sigma^2_e & 0 & 0 & 0\\ 0 & \sigma^2_e & 0 & 0\\ 0 & 0 & \sigma^2_e & 0\\ 0 & 0 & 0 & \sigma^2_e\end{bmatrix}$$

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

$\mathbf{V}_i$: Variance$[y_{time}] = \tau^2_{U_0} + \sigma^2_e$, Covariance$[y_A, y_B] = \tau^2_{U_0}$

$\mathbf{Z}_i$ = n x u values of **predictors with random effects**, so can differ per person (u = 1: intercept)

$\mathbf{Z}_i^T$ = u x n values of predictors with random effects (just $\mathbf{Z}_i$ transposed)

$\mathbf{G}_i$ = u x u estimated **random effects variances and covariances**, so will be the same for all persons ($\tau^2_{U_0}$ = intercept variance)

$\mathbf{R}_i$ = n x n **time-specific residual variances and covariances**, so will be same for all persons (here, just diagonal σ^2_e)

Summary so far…

- Regardless of what kind of model for the means you have…
  - Empty means = 1 fixed intercept that predicts no change
  - Saturated means = 1 fixed intercept + $n-1$ fixed effects for mean differences that perfectly predict the means over time
    - Is a description, not a model, and may not be possible with unbalanced time
  - Fixed linear time = 1 fixed intercept, 1 fixed linear time slope that predicts linear average change across time
    - Is a model that works with balanced or unbalanced time
    - May cause an increase in the random intercept variance by explaining residual variance
- A random intercept model…
  - Predicts constant total variance and covariance over time
    - Should be possible in balanced or unbalanced data
  - Still has residual variance (always there via default **R** matrix TYPE=VC)
- Now we'll see what happens when adding other kinds of random effects, such as a random linear effect of time…

Random Linear Time Model (6 total parameters)

Multilevel Model

Residual = time-specific deviation from individual's predicted outcome → estimated variance of σ_e^2

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

Fixed Intercept = predicted mean outcome at time 0

Fixed Linear Time Slope = predicted mean rate of change per unit time

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

Random Intercept = individual-specific deviation from fixed intercept at time 0 → estimated variance of $\tau_{U_0}^2$

Random Linear Time Slope = individual-specific deviation from fixed linear time slope → estimated variance of $\tau_{U_1}^2$

Also has an estimated **covariance** of random intercepts and slopes of $\tau_{U_{01}}$

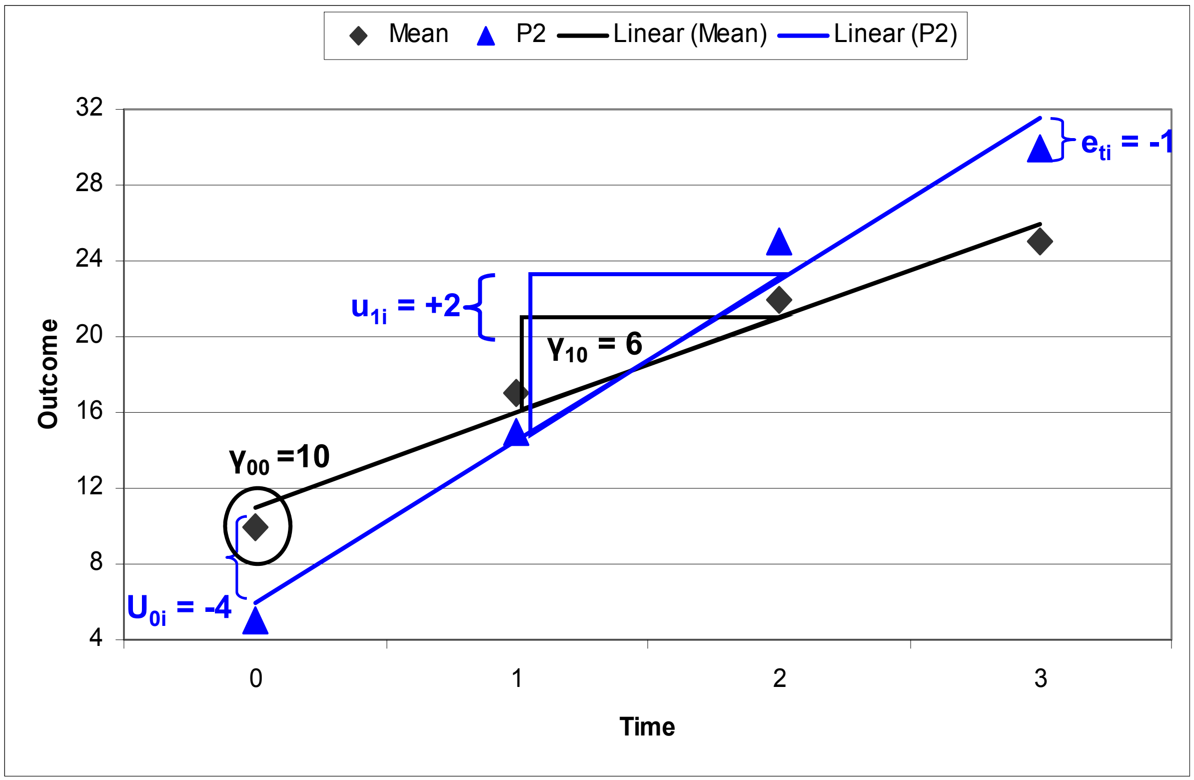
Composite Model

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

Random Linear Time Model

$$y_{ti} = (\gamma_{00} + U_{0i}) + (\gamma_{10} + U_{1i})(\text{Time}_{ti}) + e_{ti}$$

- γ_{00}: **Fixed Intercept**
- U_{0i}: Random **Intercept** Deviation
- γ_{10}: **Fixed Slope**
- U_{1i}: Random **Slope** Deviation
- e_{ti}: error for person i at time t



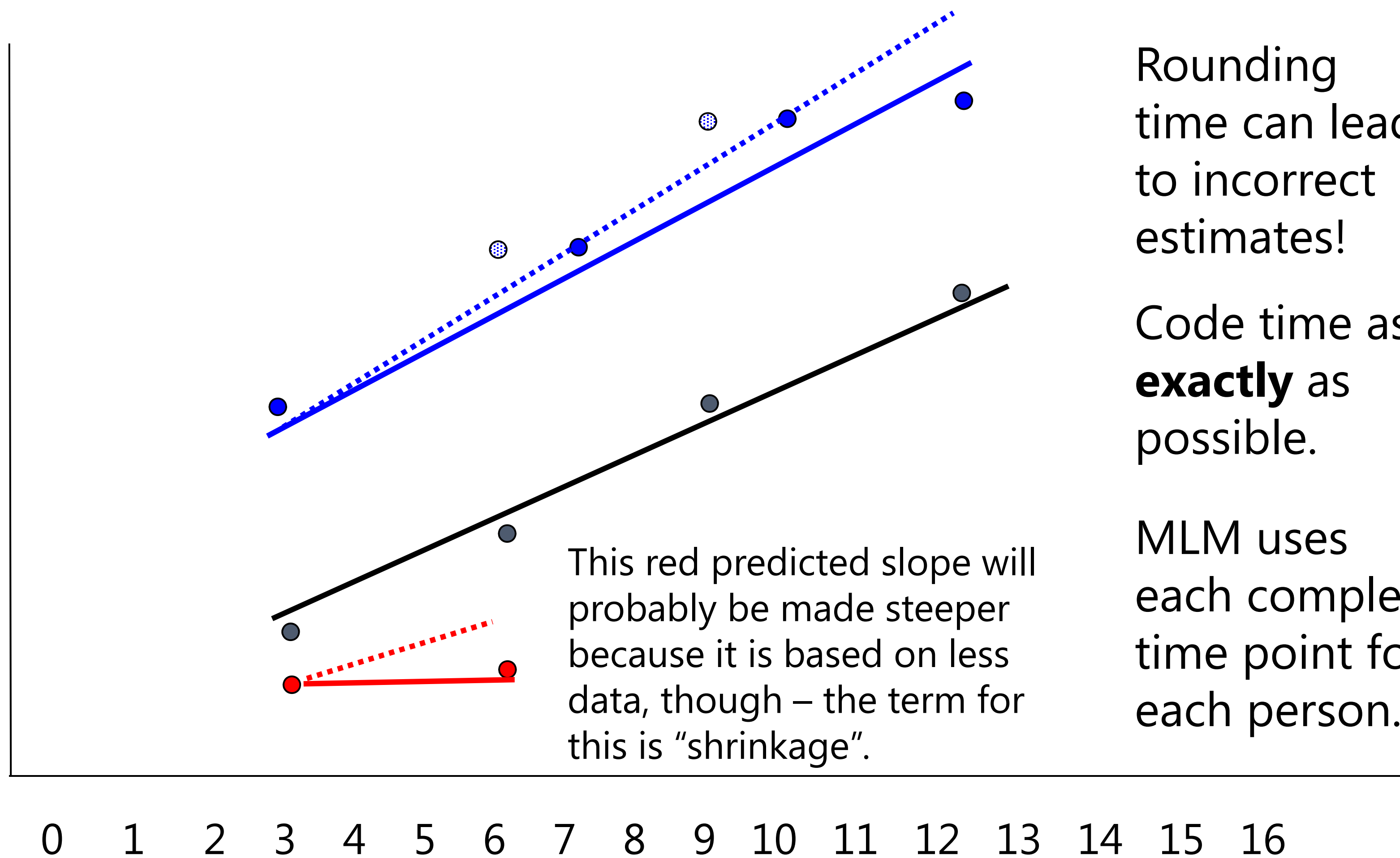
6 Parameters:

2 Fixed Effects:
γ_{00} Intercept, γ_{10} Slope

2 Random Effects Variances:
U_{0i} Intercept Variance $= \tau^2_{U_0}$
U_{1i} Slope Variance $= \tau^2_{U_1}$
Int-Slope Covariance $= \tau_{U_{01}}$

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

Unbalanced Time → Different time occasions across persons? No problem!



This red predicted slope will probably be made steeper because it is based on less data, though – the term for this is "shrinkage".

Rounding time can lead to incorrect estimates!

Code time as **exactly** as possible.

MLM uses each complete time point for each person.

Summary: Sequential Models for Effects of Time

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

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

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

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

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

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

$\beta_{1i} = \gamma_{10}$

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

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

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

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

$\beta_{1i} = \gamma_{10} + U_{1i}$

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

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

How MLM "Handles" Dependency

- Common description of the purpose of MLM is that it "addresses" or "handles" correlated (dependent) data…
- But where does this correlation come from?
  3 places (here, an example with health as an outcome):

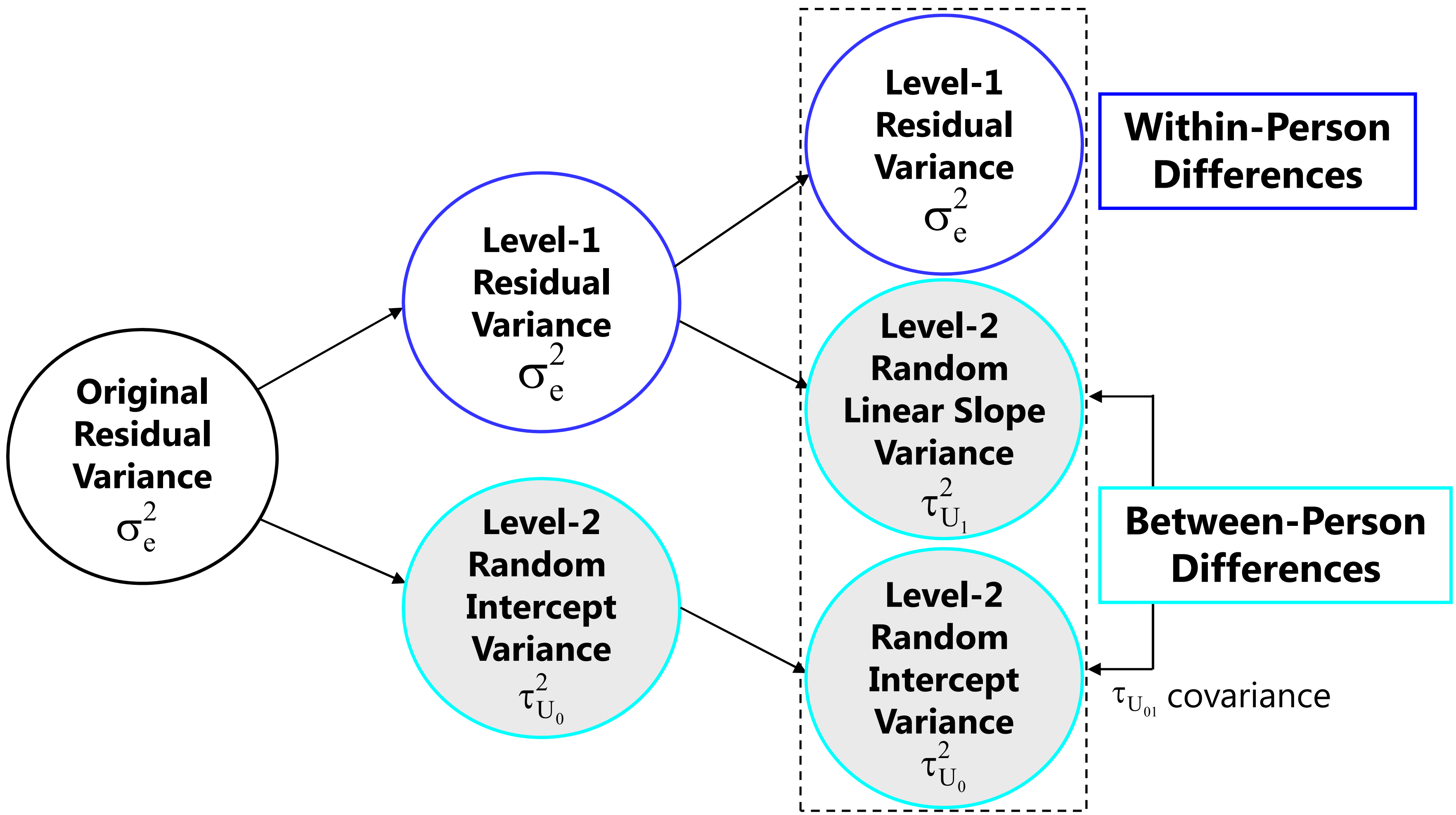
1. *Mean differences across persons*
   - Some people are just healthier than others (at every time point)
   - This is what a random intercept is for

2. *Differences in effects of predictors across persons*
   - Does *time* (or *stress*) affect health more in some persons than others?
   - This is what random slopes are for

3. Non-constant within-person correlation for unknown reasons
   - Occasions closer together may just be more related
   - This is what ACS models are for

MLM "Handles" Dependency

- Where does each kind of person dependency go? Into a new random effects variance component (or "pile" of variance):



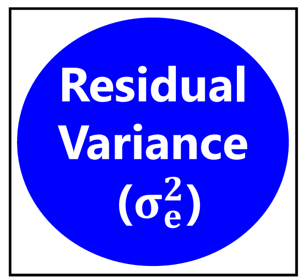
Piles of Variance

- By adding a random slope, we **carve up** our total variance into 3 piles:
  - BP (error) variance around intercept
  - BP (error) variance around slope
  - WP (error) residual variance

These 2 piles are 1 pile of "error variance" in Univ. RM ANOVA

- **But making piles does NOT make error variance go away…**

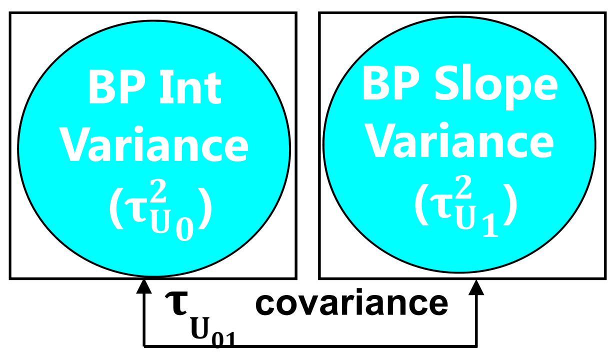
Level 1 (one source of) <u>Within-Person Variation</u>: gets accounted for by time-level predictors



FIXED effects make variance go away (explain variance).

RANDOM effects just make a new pile of variance.

Level 2 (two sources of) <u>Between-Person Variation</u>: gets accounted for by person-level predictors



Fixed vs. Random Effects of Persons

- Person dependency: via fixed effects in the model for the means or via random effects in the model for the variance?
  - Individual intercept differences can be included as:
    - **N-1 person dummy code fixed main effects OR 1 random U_{0i}**
  - Individual time slope differences can be included as:
    - **N-1*time person dummy code interactions OR 1 random U_{1i}*$time_{ti}$**
  - Either approach would appropriately control for dependency (fixed effects are used in some programs that 'control' SEs for sampling)
- Two important advantages of random effects:
  - Quantification: Direct measure of how much of the outcome variance is due to person differences (in intercept or in effects of predictors)
  - Prediction: Person differences (main effects and effects of time) then become predictable quantities – this can't happen using fixed effects
  - **Summary: Random effects give you *predictable* control of dependency**

Quantification of Random Effects Variances

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

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

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

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

Random Linear Time Models Imply:

- **People differ from each other systematically in TWO ways**—in intercept (U_{0i}) and slope (U_{1i}), which implies **TWO kinds of BP variance**, which translates to **TWO sources of person dependency** (covariance or correlation in the outcomes from the same person)
- If so, after controlling for both BP intercept and slope differences (by estimating the $\tau^2_{U_0}$ and $\tau^2_{U_1}$ variances in the **G** matrix), the $\mathbf{e_{ti}}$ **residuals** (whose variance and covariance are estimated in the **R** matrix) should be **uncorrelated with homogeneous variance across time**, as shown:

Level-2 **G** matrix: RANDOM TYPE=UN

$$\begin{bmatrix} \tau^2_{U_0} & \tau_{U_{10}} \\ \tau_{U_{01}} & \tau^2_{U_1} \end{bmatrix}$$

Level-1 **R** matrix: REPEATED TYPE=VC

$$\begin{bmatrix} \sigma^2_e & 0 & 0 & 0 \\ 0 & \sigma^2_e & 0 & 0 \\ 0 & 0 & \sigma^2_e & 0 \\ 0 & 0 & 0 & \sigma^2_e \end{bmatrix}$$

G and **R** combine to create a total **V** matrix whose per-person structure depends on the specific time occasions each person has (very flexible for unbalanced time)

Random Linear Time Model
(6 total parameters: effect of time is now RANDOM)

- How the model predicts each element of the **V** matrix:

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

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

$\mathbf{\beta_{1i} = \gamma_{10} + U_{0i}}$

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

Predicted *Time-Specific* Variance:

$$\mathrm{Var}\left[y_{ti}\right] = \mathrm{Var}\left[\left(\gamma_{00} + U_{0i}\right) + \left(\gamma_{10} + U_{1i}\right)\left(\mathrm{Time}_i\right) + e_{ti}\right]$$

$$= \mathrm{Var}\left[\left(U_{0i}\right) + \left(U_{1i} * \mathrm{Time}_i\right) + e_{ti}\right]$$

$$= \left\{\mathrm{Var}\left(U_{0i}\right)\right\} + \left\{\mathrm{Var}\left(U_{1i} * \mathrm{Time}_i\right)\right\} + \left\{2 * \mathrm{Cov}\left(U_{0i}, U_{1i} * \mathrm{Time}_i\right)\right\} + \left\{\mathrm{Var}\left(e_{ti}\right)\right\}$$

$$= \left\{\mathrm{Var}\left(U_{0i}\right)\right\} + \left\{\mathrm{Time}_i^2 * \mathrm{Var}\left(U_{1i}\right)\right\} + \left\{2 * \mathrm{Time}_i * \mathrm{Cov}\left(U_{0i}, U_{1i}\right)\right\} + \left\{\mathrm{Var}\left(e_{ti}\right)\right\}$$

$$= \left\{\tau_{U_0}^2\right\} + \boxed{\left\{\mathrm{Time}_i^2 * \tau_{U_1}^2\right\} + \left\{2 * \mathrm{Time}_i * \tau_{U_{01}}\right\}} + \left\{\sigma_e^2\right\}$$

Random Linear Time Model
(6 total parameters: effect of time is now RANDOM)

- How the model predicts each element of the **V** matrix:

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

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

$\mathbf{\beta_{1i} = \gamma_{10} + U_{0i}}$

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

Predicted *Time-Specific* Covariances (Time A with Time B):

$$\mathrm{Cov}\left[y_{Ai}, y_{Bi}\right] = \mathrm{Cov}\left[\left\{\left(\gamma_{00} + U_{0i}\right) + \left(\gamma_{10} + U_{1i}\right)\left(A_i\right) + e_{Ai}\right\}, \left\{\left(\gamma_{00} + U_{0i}\right) + \left(\gamma_{10} + U_{1i}\right)\left(B_i\right) + e_{Bi}\right\}\right]$$

$$= \mathrm{Cov}\left[\left\{U_{0i} + \left(U_{1i}A_i\right)\right\}, \left\{U_{0i} + \left(U_{1i}B_i\right)\right\}\right]$$

$$= \mathrm{Cov}\left[U_{0i}, U_{0i}\right] + \mathrm{Cov}\left[U_{0i}, U_{1i}B_i\right] + \mathrm{Cov}\left[U_{0i}, U_{1i}A_i\right] + \mathrm{Cov}\left[U_{1i}A_i, U_{1i}B_i\right]$$

$$= \left\{\mathrm{Var}\left(U_{0i}\right)\right\} + \left\{\left(A_i + B_i\right) * \mathrm{Cov}\left(U_{0i}, U_{1i}\right)\right\} + \left\{\left(A_i B_i\right)\mathrm{Var}\left(U_{1i}\right)\right\}$$

$$= \left\{\tau^2_{U_0}\right\} + \boxed{\left\{\left(A_i + B_i\right)\tau_{U_{01}}\right\} + \left\{\left(A_i B_i\right)\tau^2_{U_1}\right\}}$$

Random Linear Time Model
(6 total parameters: effect of time is now RANDOM)

- Scalar "mixed" model equation per person:

$$\mathbf{Y}_i = \mathbf{X}_i * \boldsymbol{\gamma} + \mathbf{Z}_i * \mathbf{U}_i + \mathbf{E}_i$$

$$\begin{bmatrix} y_{0i} \\ y_{1i} \\ y_{2i} \\ y_{3i} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} \gamma_{00} \\ \gamma_{10} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} U_{0i} \\ U_{1i} \end{bmatrix} + \begin{bmatrix} e_{0i} \\ e_{1i} \\ e_{2i} \\ e_{3i} \end{bmatrix}$$

$$\begin{bmatrix} y_{0i} \\ y_{1i} \\ y_{2i} \\ y_{3i} \end{bmatrix} = \begin{bmatrix} \gamma_{00} + \gamma_{10}(0) \\ \gamma_{00} + \gamma_{10}(1) \\ \gamma_{00} + \gamma_{10}(2) \\ \gamma_{00} + \gamma_{10}(3) \end{bmatrix} + \begin{bmatrix} U_{0i} + U_{1i}(0) \\ U_{0i} + U_{1i}(1) \\ U_{0i} + U_{1i}(2) \\ U_{0i} + U_{1i}(3) \end{bmatrix} + \begin{bmatrix} e_{0i} \\ e_{1i} \\ e_{2i} \\ e_{3i} \end{bmatrix}$$

$$\begin{bmatrix} y_{0i} \\ y_{1i} \\ y_{2i} \\ y_{3i} \end{bmatrix} = \begin{bmatrix} \gamma_{00} + \gamma_{10}(0) + U_{0i} + U_{1i}(0) + e_{0i} \\ \gamma_{00} + \gamma_{10}(1) + U_{0i} + U_{1i}(1) + e_{1i} \\ \gamma_{00} + \gamma_{10}(2) + U_{0i} + U_{1i}(2) + e_{2i} \\ \gamma_{00} + \gamma_{10}(3) + U_{0i} + U_{1i}(3) + e_{3i} \end{bmatrix}$$

$\mathbf{X}_i$ = *n* x *k* values of **predictors with fixed effects**, so can differ per person (*k* = 2: intercept, linear time)

$\boldsymbol{\gamma}$ = *k* x 1 estimated **fixed effects**, so will be the same for all persons (γ_{00} = intercept, γ_{10} = linear time)

$\mathbf{Z}_i$ = *n* x *u* values of **predictors with random effects**, so can differ per person (*u* = 2: intercept, linear time)

$\mathbf{U}_i$ = *u* x 2 estimated individual **random effects**, so can differ per person

$\mathbf{E}_i$ = *n* x *n* time-specific residuals, so can differ per person

Random Linear Time Model
(6 total parameters: effect of time is now RANDOM)

- Predicted total variances and covariances per person:

$$\mathbf{V}_i = \mathbf{Z}_i * \mathbf{G}_i * \mathbf{Z}_i^T + \mathbf{R}_i$$

$$\mathbf{V}_i = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} \tau^2_{U_0} & \tau_{U_{01}} \\ \tau_{U_{01}} & \tau^2_{U_1} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{bmatrix} + \begin{bmatrix} \sigma^2_e & 0 & 0 & 0 \\ 0 & \sigma^2_e & 0 & 0 \\ 0 & 0 & \sigma^2_e & 0 \\ 0 & 0 & 0 & \sigma^2_e \end{bmatrix}$$

$\mathbf{V}_i$ matrix: $\text{Variance}\left[y_{time}\right]$

$$= \tau^2_{U_0} + \left[(\text{time})^2 \tau^2_{U_1}\right] + \left[2(\text{time})\tau_{U_{01}}\right] + \sigma^2_e$$

$\mathbf{V}_i$ matrix: $\text{Covariance}\left[y_A, y_B\right]$

$$= \tau^2_{U_0} + \left[(A+B)\tau_{U_{01}}\right] + \left[(AB)\tau^2_{U_1}\right]$$

$\mathbf{V}_i$ matrix = complicated ☺

$\mathbf{Z}_i$ = n x u values of **predictors with random effects**, so can differ per person (u = 2: int., time slope)

$\mathbf{Z}_i^T$ = u x n values of predictors with random effects (just $\mathbf{Z}_i$ transposed)

$\mathbf{G}_i$ = u x u estimated **random effects variances and covariances**, so will be the same for all persons ($\tau^2_{U_0}$ = int. var., $\tau^2_{U_1}$ = slope var.)

$\mathbf{R}_i$ = n x n **time-specific residual variances and covariances**, so will be same for all persons (here, just diagonal σ^2_e)

Building **V** across persons: Random Linear Time Model

- **V** for two persons with **unbalanced time** observations:

$$\mathbf{V} = \mathbf{Z} * \mathbf{G} * \mathbf{Z}^T + \mathbf{R}$$

$$\mathbf{V} = \begin{bmatrix} 1 & 0.0 & 0 & 0 \\ 1 & 1.0 & 0 & 0 \\ 1 & 2.0 & 0 & 0 \\ 1 & 3.0 & 0 & 0 \\ 0 & 0 & 1 & 0.2 \\ 0 & 0 & 1 & 1.4 \\ 0 & 0 & 1 & 2.3 \\ 0 & 0 & 1 & 3.5 \end{bmatrix} \begin{bmatrix} \tau_{U_0}^2 & \tau_{U_{01}} & 0 & 0 \\ \tau_{U_{01}} & \tau_{U_1}^2 & 0 & 0 \\ 0 & 0 & \tau_{U_0}^2 & \tau_{U_{01}} \\ 0 & 0 & \tau_{U_{01}} & \tau_{U_1}^2 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0.0 & 1.0 & 2.0 & 3.0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0.2 & 1.4 & 2.3 & 3.5 \end{bmatrix} + \begin{bmatrix} \sigma_e^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_e^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_e^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_e^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_e^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_e^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_e^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_e^2 \end{bmatrix}$$

- The giant combined **V** matrix across persons is how the multilevel or mixed model is actually estimated in SAS
- Known as "**block diagonal**" structure → predictions are given for each person, but 0's are given for the elements that describe relationships between persons (because persons are supposed to be independent here!)

Building **V** across persons: Random Linear Time Model

- **V** for two persons also with ***different n*** per person:

$$\mathbf{V} = \mathbf{Z} * \mathbf{G} * \mathbf{Z}^T + \mathbf{R}$$

$$\mathbf{V} = \begin{bmatrix} 1 & 0.0 & 0 & 0 \\ 1 & 1.0 & 0 & 0 \\ 1 & 2.0 & 0 & 0 \\ 1 & 3.0 & 0 & 0 \\ 0 & 0 & 1 & 0.2 \\ 0 & 0 & 1 & 1.4 \\ 0 & 0 & 1 & 3.5 \end{bmatrix} \begin{bmatrix} \tau^2_{U_0} & \tau_{U_{01}} & 0 & 0 \\ \tau_{U_{01}} & \tau^2_{U_1} & 0 & 0 \\ 0 & 0 & \tau^2_{U_0} & \tau_{U_{01}} \\ 0 & 0 & \tau_{U_{01}} & \tau^2_{U_1} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0.0 & 1.0 & 2.0 & 3.0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0.2 & 1.4 & 3.5 \end{bmatrix} + \begin{bmatrix} \sigma^2_e & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma^2_e & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2_e & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2_e & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2_e & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma^2_e & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2_e \end{bmatrix}$$

- The "block diagonal" does not need to be the same size or contain the same time observations per person…
- **R** matrix can also include non-0 covariance or differential residual variance across time (as in ACS models), although the models based on the idea of a "lag" won't work for unbalanced or unequal-interval time

G, R, and V: The Take-Home Point

- The partitioning of variance into piles…
  - **Level 2 = BP → G** matrix of random effects variances/covariances
  - **Level 1 = WP → R** matrix of residual variances/covariances
  - **G** and **R** combine via **Z** to create **V** matrix of total variances/covariances
  - Many flexible options that allows the variances and covariances to vary in a time-dependent way that better matches the actual data
    - Can allow variance and covariance due to other predictors, too

Compound Symmetry (CS)

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

Random effects models use **G** and **R** to predict something in-between!

Unstructured (UN)

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

Parsimony ← Univariate RM ANOVA — Multivariate RM ANOVA → **Good fit**

Two Sides of Any Model: Estimation

- **Fixed Effects in the Model for the Means:**
  - How the expected outcome for a given observation varies as a function of values on *known* predictor variables
  - Fixed effects predict the Y values per se *but are not parameters that are solved for iteratively in maximum likelihood estimation*

- **Random Effects in the Model for the Variances:**
  - How model residuals are related across observations (persons, groups, time, etc) – *unknown* things due to sampling
  - Random effects variances and covariances are a mechanism by which complex patterns of variance and covariance among the Y residuals can be predicted (not the Y values, but their dispersion)
  - Anything besides level-1 residual variance σ_e^2 must be solved for iteratively – increases the dimensionality of estimation process
  - Estimation utilizes the predicted **V** matrix for each person
  - In the material that follows, **V** will be based on a random linear model

End Goals of Maximum Likelihood Estimation

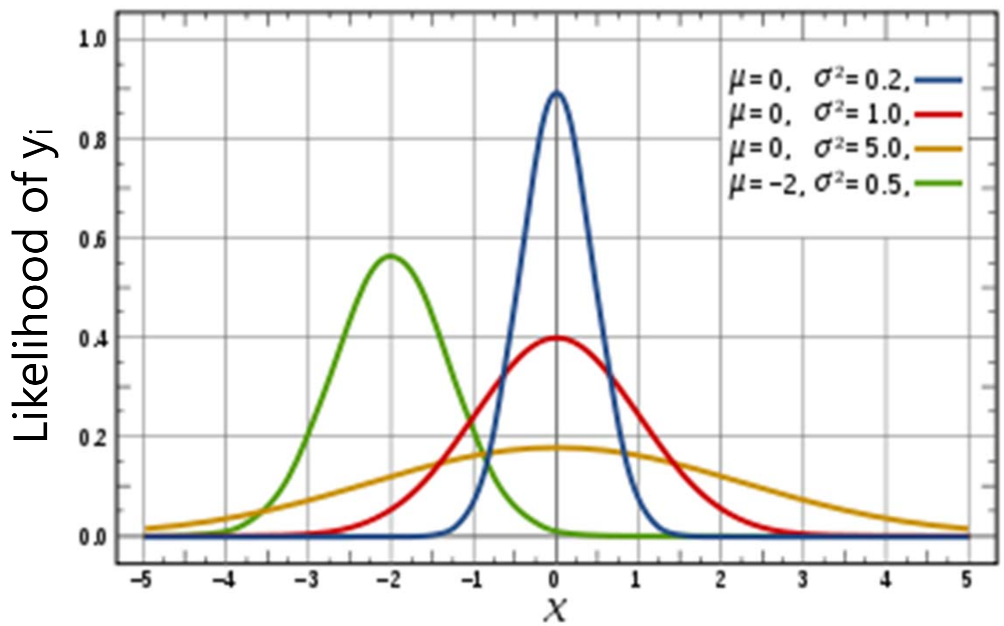
1. Obtain "most likely" values for each unknown model parameter (random effects variances and covariances, residual variances and covariances, which then are used to calculate the fixed effects) → **the estimates**

2. Obtain an index as to how likely each parameter value actually is (i.e., "really likely" or pretty much just a guess?) → **the standard error (SE) of the estimates**

3. Obtain an index as to how well the model we've specified actually describes the data → **the model fit indices**

How does all this happen? The magic of multivariate normal…(but let's start with univariate normal first)

Univariate Normal



- This function tells us how **likely** any value of y_i is given two pieces of info:
  - predicted value $\hat{y}_i$
  - residual variance σ_e^2

Univariate Normal PDF (two ways):

$$f(y_i) = \frac{1}{\sqrt{2\pi\sigma_e^2}} * \exp\left[-\frac{1}{2} * \frac{\left(y_i - \hat{y}_i\right)^2}{\sigma_e^2}\right]$$

$$f(y_i) = \left(2\pi\sigma_e^2\right)^{-1/2} * \exp\left[-\frac{1}{2} * \left(y_i - \hat{y}_i\right)\left(\sigma_e^2\right)^{-1}\left(y_i - \hat{y}_i\right)\right]$$

- Example: regression

$$y_i = \beta_0 + \beta_1 X_i + e_i$$
$$\hat{y}_i = \beta_0 + \beta_1 X_i$$
$$e_i = y_i - \hat{y}_i \qquad \sigma_e^2 = \frac{\sum_{i=1}^{N} e_i^2}{N-2}$$

Multivariate Normal for Y_i (height for all *n* outcomes for person *i*)

$$\text{Univariate Normal PDF: } f(y_i) = \left(2\pi\sigma_e^2\right)^{-1/2} * \exp\left[-\frac{1}{2} * \left(y_i - \hat{y}_i\right)\left(\sigma_e^2\right)^{-1}\left(y_i - \hat{y}_i\right)\right]$$

$$\text{Multivariate Normal PDF: } f(\mathbf{Y}_i) = \left(2\pi\right)^{-n/2} * \left|\mathbf{V}_i\right|^{-1/2} * \exp\left[-\frac{1}{2} * \left(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\gamma}\right)^T \left(\mathbf{V}_i\right)^{-1}\left(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\gamma}\right)\right]$$

- In a random linear time model, the only fixed effects (in $\boldsymbol{\gamma}$) that predict the $\mathbf{Y}_i$ outcome values are the fixed intercept and fixed linear time slope
- The model also gives us $\mathbf{V}_i$ → the model-predicted total variance and covariance matrix across the occasions, taking into account the time values
- Uses $|\mathbf{V}_i|$ = determinant of $\mathbf{V}_i$ = summary of *non-redundant* info
  - Reflects sum of variances across occasions controlling for covariances
- $(\mathbf{V}_i)^{-1}$ → matrix inverse → like dividing (so can't be 0 or negative)
  - $(\mathbf{V}_i)^{-1}$ must be "positive definite", which in practice means no 0 random variances and no out-of-bound correlations between random effects
  - Otherwise, SAS uses "generalized inverse" → questionable results

Now Try Some Possible Answers...
(e.g., for the 4 **V** parameters in this random linear model example)

- Plug $\mathbf{V}_i$ predictions into log-likelihood function, sum over persons:

$$L = \prod_{i=1}^{N} \left\{ (2\pi)^{-n/2} * \left|\mathbf{V}_i\right|^{-1/2} * \exp\left[-\frac{1}{2} (\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\gamma})^T (\mathbf{V}_i)^{-1} (\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\gamma}) \right] \right\}$$

$$LL = \sum_{i=1}^{N} \left\{ \left[-\frac{n}{2} \log(2\pi) \right] + \left[-\frac{1}{2} \log\left|\mathbf{V}_i\right| \right] + \left[-\frac{1}{2} (\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\gamma})^T (\mathbf{V}_i)^{-1} (\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\gamma}) \right] \right\}$$

- Try one set of possible parameter values for $\mathbf{V}_i$, compute LL
- Try another possible set for $\mathbf{V}_i$, compute LL….
  - Different algorithms are used to decide which values to try given that each parameter has its own distribution → like an uncharted mountain
  - Calculus helps the program scale this multidimensional mountain
    - At the top, all first partial derivatives (linear slopes at that point) ≈ 0
    - Positive first partial derivative? Too *low*, try again. Negative? Too *high*, try again.
    - Matrix of partial first derivatives = "score function" = "gradient" (as in NLMIXED output for models with truly nonlinear effects)

End Goals 1 and 2: Model Estimates and SEs

- Process terminates (the model "converges") when the next set of tried values for Vi don't improve the LL very much…
  - e.g., SAS default convergence criteria = .00000001
  - Those are the values for the parameters that, relative to the other possible values tried, are "most likely" → the variance estimates

- But we need to know how trustworthy those estimates are…
  - Precision is indexed by the steepness of the multidimensional mountain, where steepness → more negative partial second derivatives
  - Matrix of partial second derivatives = "Hessian matrix"
  - Hessian matrix * -1 = "information matrix"
  - So steeper function = more information = more precision = smaller SE

$$\text{Each parameter SE} = \frac{1}{\sqrt{\text{information}}}$$

What about the Fixed Effects?

- Likelihood mountain does NOT include fixed effects as additional search dimensions (only variances and covariances that make $\mathbf{V}_i$)
- **Fixed effects are determined** given the parameters for $\mathbf{V}_i$:

$$\boldsymbol{\gamma} = \left\{\sum_{i=1}^{N}\left(\mathbf{X}_i^T\mathbf{V}_i^{-1}\mathbf{X}_i\right)\right\}^{-1}\sum_{i=1}^{N}\left(\mathbf{X}_i^T\mathbf{V}_i^{-1}\mathbf{Y}_i\right), \qquad \mathrm{Cov}\left(\boldsymbol{\gamma}\right) = \left\{\sum_{i=1}^{N}\left(\mathbf{X}_i^T\mathbf{V}_i^{-1}\mathbf{X}_i\right)\right\}^{-1}$$

All we need is $\mathbf{V}_i$ and the data: **X**, **Y**

γ = fixed effect estimates

$\mathrm{Cov}(\gamma) = \gamma$ sampling variance (SQRT of diagonal = SE)

- This is actually what happens in regular regression (GLM), too:

$$\text{GLM matrix solution: } \boldsymbol{\beta} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\left(\mathbf{X}^T\mathbf{Y}\right), \qquad \mathrm{Cov}\left(\boldsymbol{\beta}\right) = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\sigma_e^2$$

$$\text{GLM scalar solution: } \beta = \frac{\sum_{i=1}^{N}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{N}(x_i-\bar{x})^2}, \quad \mathrm{Cov}\left(\beta\right) = \frac{\sigma_e^2}{\sum_{i=1}^{N}(x_i-\bar{x})^2}$$

- **Implication: fixed effects don't cause estimation problems**…

What about ML vs. REML?

- **REML** estimates of random effects variances and covariances are **unbiased** because they account for the uncertainty that results from simultaneously also estimating fixed effects (whereas ML estimates do not, so they are too small)

- What does this mean? Remember "population" vs. "sample" formulas for computing variance?

$$\text{Population: } \sigma_e^2 = \frac{\sum_{i=1}^{N}(y_i - \mu)^2}{N} \qquad \text{Sample: } \sigma_e^2 = \frac{\sum_{i=1}^{N}(y_i - \bar{y})^2}{N-1}$$

  - N-1 is used because the mean had to be estimated from the data (i.e., the mean is the fixed intercept)…

- Same idea: ML estimates of random effects variances will be downwardly biased by a factor of (N – k) / N, where N = # persons and k = #fixed effects… it just looks way more complicated

What about ML vs. REML?

$$\text{ML:}\quad LL = \left[-\frac{T-0}{2}\log(2\pi)\right] + \left[-\frac{1}{2}\sum_{i=1}^{N}\log\left|\mathbf{V}_i\right|\right] + \left[-\frac{1}{2}\sum_{i=1}^{N}\left(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\gamma}\right)^T \mathbf{V}_i^{-1}\left(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\gamma}\right)\right]$$

$$\text{REML:}\ LL = \left[-\frac{T-k}{2}\log(2\pi)\right] + \left[-\frac{1}{2}\sum_{i=1}^{N}\log\left|\mathbf{V}_i\right|\right] + \left[-\frac{1}{2}\sum_{i=1}^{N}\left(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\gamma}\right)^T \mathbf{V}_i^{-1}\left(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\gamma}\right)\right]$$

$$+\left[-\frac{1}{2}\log\left|\sum_{i=1}^{N}\mathbf{X}_i^T\mathbf{V}_i^{-1}\mathbf{X}_i\right|\right]$$

$$\text{where:}\quad \left[-\frac{1}{2}\log\left|\sum_{i=1}^{N}\mathbf{X}_i^T\mathbf{V}_i^{-1}\mathbf{X}_i\right|\right] = \left[\frac{1}{2}\log\left|\left(\sum_{i=1}^{N}\mathbf{X}_i^T\mathbf{V}_i^{-1}\mathbf{X}_i\right)^{-1}\right|\right] = \underbrace{\left[\frac{1}{2}\log\left|\mathrm{Cov}(\boldsymbol{\gamma})\right|\right]}$$

- Extra part in REML is the sampling variance of the fixed effects… it is added back in to account for uncertainty in estimating fixed effects
- REML maximizes the likelihood of the residuals specifically, so models with different fixed effects are not on the same scale and are not comparable
  - This is why you can't do -2ΔLL tests in REML when the models to be compared have different fixed effects → the model residuals are defined differently

End Goal #3: How well do the model predictions match the data?

- End up with ML or REML LL from predicting $\mathbf{V}_i$ → so how good is it?
- Absolute model fit assessment is only possible when the $\mathbf{V}_i$ matrix is organized the same for everyone – in other words, balanced data
  - Items are usually fixed, so can get absolute fit in CFA and SEM → χ^2 test is based on match between actual and predicted data matrix
  - Time is often a continuous variable, so no absolute fit provided in MLM (or in SEM when using random slopes or T-scores for unbalanced time)
    - Can compute absolute fit when the saturated means, unstructured variance model is estimable in ML → is -2ΔLL versus "perfect" model for time

- Relative model fit is given as **−2LL** in SAS, in which smaller is better
  - -2* needed to conduct "likelihood ratio" or "deviance difference" tests
  - Also information criteria:
    - **AIC**: -2LL + 2*(#parms)
    - **BIC**: -2LL + log(N)*(#parms)
    - #parms = all parameters in ML; #parms = variance model parameters only in REML

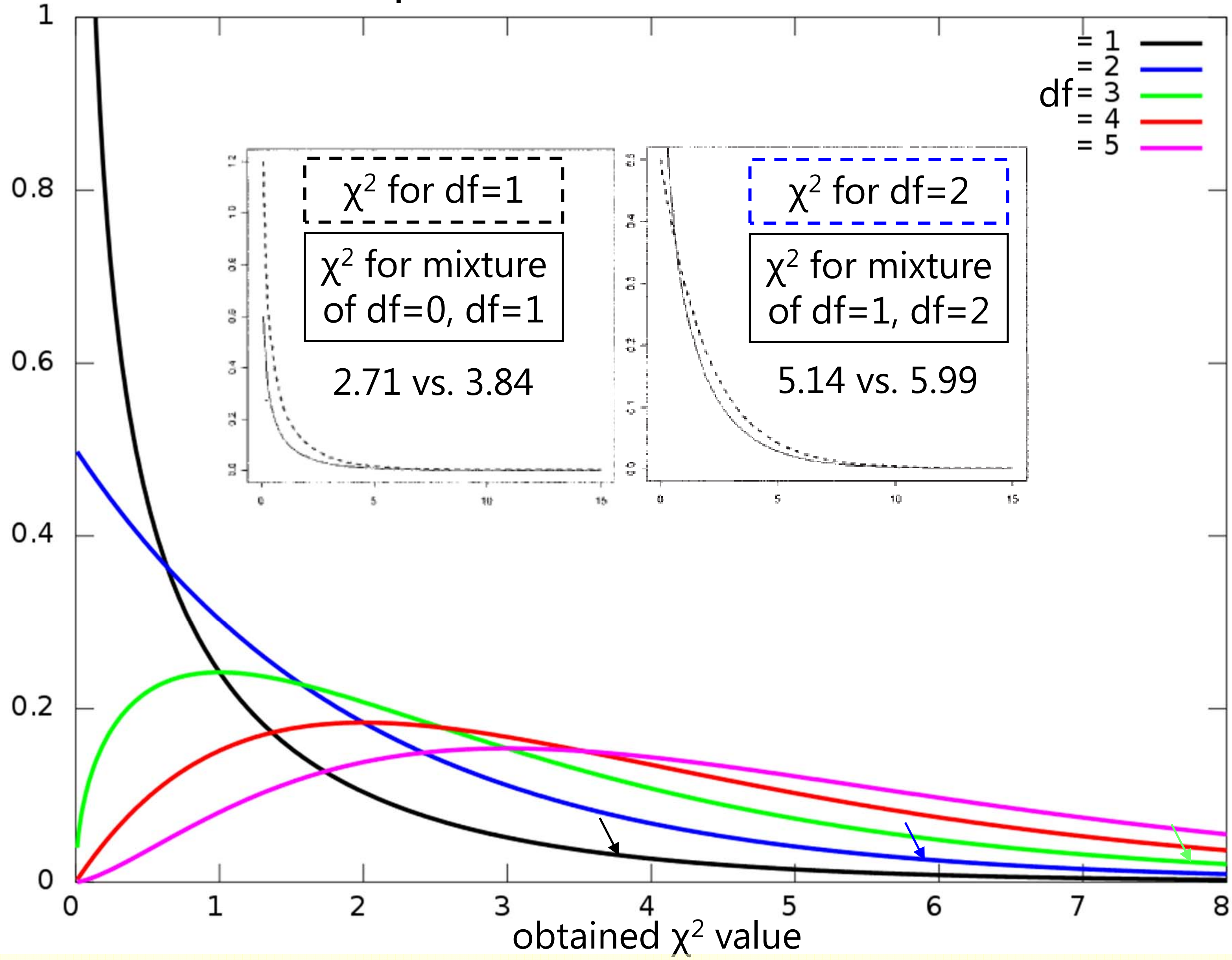
What about testing variances > 0?

- $-2\Delta LL$ between two nested models is χ^2-distributed only when the added parameters do not have a boundary (like 0 or 1)
  - Ok for fixed effects (could be any positive or negative value)
  - NOT ok for tests of random variances (must be > 0)
  - Ok for tests of heterogeneous variances and covariances (extra parameters can be phrased as unbounded deviations)

- When testing addition of parameters that have a boundary, $-2\Delta LL$ will follow a **mixture** of χ^2 distributions instead
  - e.g., when adding random intercept variance (test > 0?)
    - When estimated as positive, will follow χ^2 with df=1
    - When estimated as negative… can't happen, will follow χ^2 with df=0
  - End result: **$-2\Delta LL$ will be too conservative in boundary cases**

χ^2 Distributions

small pictures from Stoel et al., 2006



Critical Values for 50:50 Mixture of Chi-Square Distributions

	Significance Level				
df (q)	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**
0 vs. 1	1.64	2.71	3.84	5.41	6.63
1 vs. 2	3.81	5.14	6.48	8.27	9.63
2 vs. 3	5.53	7.05	8.54	10.50	11.97
3 vs. 4	7.09	8.76	10.38	12.48	14.04
4 vs. 5	8.57	10.37	12.10	14.32	15.97
5 vs. 6	10.00	11.91	13.74	16.07	17.79
6 vs. 7	11.38	13.40	15.32	17.76	19.54
7 vs. 8	12.74	14.85	16.86	19.38	21.23
8 vs. 9	14.07	16.27	18.35	20.97	22.88
9 vs. 10	15.38	17.67	19.82	22.52	24.49
10 vs. 11	16.67	19.04	21.27	24.05	26.07

This may work ok if only one new parameter is bounded … for example:

+ Random Intercept df=1: 2.71 vs. 3.84

+ Random Linear df=2: 5.14 vs. 5.99

+ Random Quad df=3: 7.05 vs. 7.82

Critical values such that the right-hand tail probability =

$0.5 \times \Pr(\chi^2_q > c) + 0.5 \times \Pr(\chi^2_{q+1} > c)$

Source: Appendix C (p. 484) from Fitzmaurice, Laird, & Ware (2004). *Applied Longitudinal Analysis*. Hoboken, NJ: Wiley

Solutions for Boundary Problems when using −2ΔLL tests

- If adding random intercept variance only, use p < .10; $\chi^2(1) > 2.71$
  - Because χ^2 (0) = 0, can just cut p-value in half to get correct p-value

- If adding ONE random slope variance (and covariance with random intercept), can use mixture p-value from $\chi^2(1)$ and $\chi^2(2)$

$$\text{Mixture } p\text{-value} = 0.5 * \text{prob}(\chi^2_1 > -2\Delta\text{LL}) + 0.5 * \text{prob}(\chi^2_2 > -2\Delta\text{LL})$$

so critical χ^2 = 5.14, not 5.99

- However – using a 50/50 mixture assumes a diagonal information matrix for the random effects variances (assumes the values for each are arrived at independently, which pry isn't the case)

- Two options for more complex cases:
  - Simulate data to determine actual mixture for calculating p-value
  - Accept that −2ΔLL is conservative in these cases, and use it anyway → I'm using ~ to acknowledge this: e.g., −2ΔLL(~2) > 5.99, p < .05

Predicted Level-2 $\mathbf{U}_i$ Random Effects *(aka Empirical Bayes or BLUP Estimates)*

- Level-2 $\mathbf{U}_i$ random effects require further explanation…
  - Empty two-level model: $\mathbf{y_{ti} = \gamma_{00} + U_{0i} + e_{ti}}$
  - U_{0i}'s are deviated person means, right? Well, not exactly…
- 3 ways of representing size of individual differences in individual intercepts and slopes across people:
  - Get individual OLS intercepts and slopes; calculate their variance
  - Estimate variance of the $\mathbf{U}_i$'s (what we do in MLM)
  - Predict individual $\mathbf{U}_i$'s; calculate their variance (2-stage MLM)
- Expected order of magnitude of variance estimates:
  - OLS variance > MLM variance > Predicted $\mathbf{U}_i$'s variance
  - Why are these different? **Shrinkage**.

What about the U's?

- Individual $\mathbf{U}_i$ values are NOT estimated in the ML process
  - **G** matrix variances and covariances are sufficient statistics for the estimation process assuming multivariate normality of $\mathbf{U}_i$ values
  - Individual $\mathbf{U}_i$ random effects are **predicted** by asking for the SOLUTION on the RANDOM statement as: $\mathbf{U}_i = \mathbf{G}_i \mathbf{Z}_i^T \mathbf{V}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\gamma})$
    - Which then create individual estimates as $\boldsymbol{\beta}_{0i} = \boldsymbol{\gamma}_{00} + \mathbf{U}_{0i}$ and $\boldsymbol{\beta}_{1i} = \boldsymbol{\gamma}_{10} + \mathbf{U}_{1i}$

- What isn't obvious: the composite $\boldsymbol{\beta}_i$ values are weighted combinations of the fixed effects ($\boldsymbol{\gamma}$) and individual OLS estimates (β_{OLSi}) :

$$\text{Random Effects: } \boldsymbol{\beta}_i = \mathbf{W}_i \boldsymbol{\beta}_{OLSi} + \left(\mathbf{I} - \mathbf{W}_i\right)\boldsymbol{\gamma} \qquad \text{where: } \mathbf{W}_i = \mathbf{G}_i \left[\mathbf{G}_i + \sigma_e^2 \left(\mathbf{Z}_i^T \mathbf{Z}_i\right)^{-1}\right]^{-1}$$

  - The more "true" variation in intercepts and slopes there is in the data (in **G**), the more the $\boldsymbol{\beta}_i$ estimates are based on individual OLS estimates
  - But the more "unexplained" residual variation there is around the individual trajectories (in **R**), the more the fixed effects are heavily weighted instead
    - = **SHRINKAGE** (more so for people with fewer occasions, too)

What about the U's?

- Point of the story – $\mathbf{U}_i$ values are NOT single scores:
  - They are the mean of a distribution of possible values for each person (i.e., as given by the SE for each $\mathbf{U}_i$, which is also provided)
  - These "best estimates" of the $\mathbf{U}_i$ values are shrunken anyway

- Good news: you don't need those $\mathbf{U}_i$ values in the first place!
  - Goal of MLM is to estimate and predict the variance of the $\mathbf{U}_i$ values (in **G**) with person-level characteristics directly within the same model
  - If you want your $\mathbf{U}_i$ values to be predictors instead, then you need to buy your growth curve model at the SEM store instead of the MLM store
  - We can use the predicted $\mathbf{U}_i$ values to examine potential violations of model assumptions, though…
    - Get $\mathbf{U}_i$ values by adding: ODS OUTPUT SolutionR=dataset;
    - Get e_{ti} residuals by adding OUTP=dataset after / on MODEL statement
    - Add RESIDUAL option after / on MODEL statement to make plots

Estimation: The Grand Finale

- Estimation in MLM is all about the random effects variances and covariances
  - The more there are, the harder it is to find them (the more dimensions of the likelihood mountain there are to scale)
  - "Non-positive-definite" **G** matrix means "broken model"
  - Fixed effects are solved for after-the-fact, so they rarely cause estimation problems
  - Individual random effects are not model parameters, but can be predicted after-the-fact (with some problems in doing so)

- Estimation comes in two flavors:
  - ML → maximize the data; compare any nested models
  - REML → maximize the residuals; compare models that differ in their model for the variance only