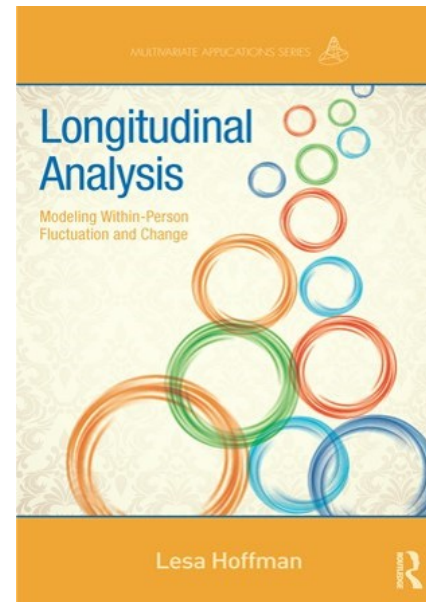
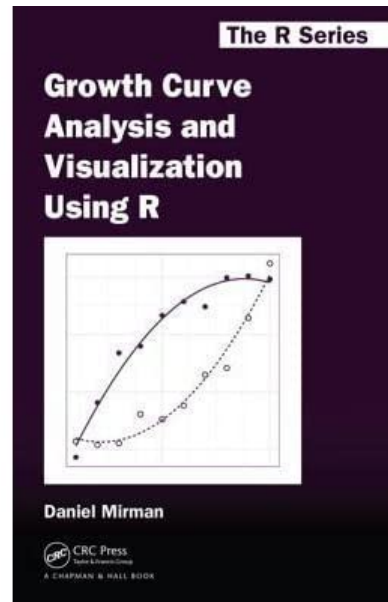




PSY9185 - MULTILEVEL MODELS

Longitudinal data analysis using multilevel models
Nikolai Czajkowski

- Introduction to longitudinal data and LMM.
- Plotting longitudinal data.
- Basic two-level longitudinal
 - Fixed and random coefficients
- Non-linear and piecewise continuous change
- Including predictor in the model
 - Time invariant predictors
 - Time varying predictors



- Several examples borrowed from Mirman (2014).
- Lesa Hoffman has links to several full video courses in longitudinal modelling on her homepage.
 - <http://www.lesahoffman.com/>

Fitting mixed models in R:

- <https://m-clark.github.io/mixed-models-with-R/introduction.html>
- <https://rpsychologist.com/r-guide-longitudinal-lme-lmer>

Books on R:

- Cookbook for R: <http://www.cookbook-r.com/>
- R for data science: <https://r4ds.had.co.nz/>

Online resources

- R-bloggers; R news and tutorials contributed by hundreds of R bloggers.
 - (<http://www.r-bloggers.com/>)
- Stackoverflow: question and answer site for programmers.
 - (<http://stackoverflow.com/>)

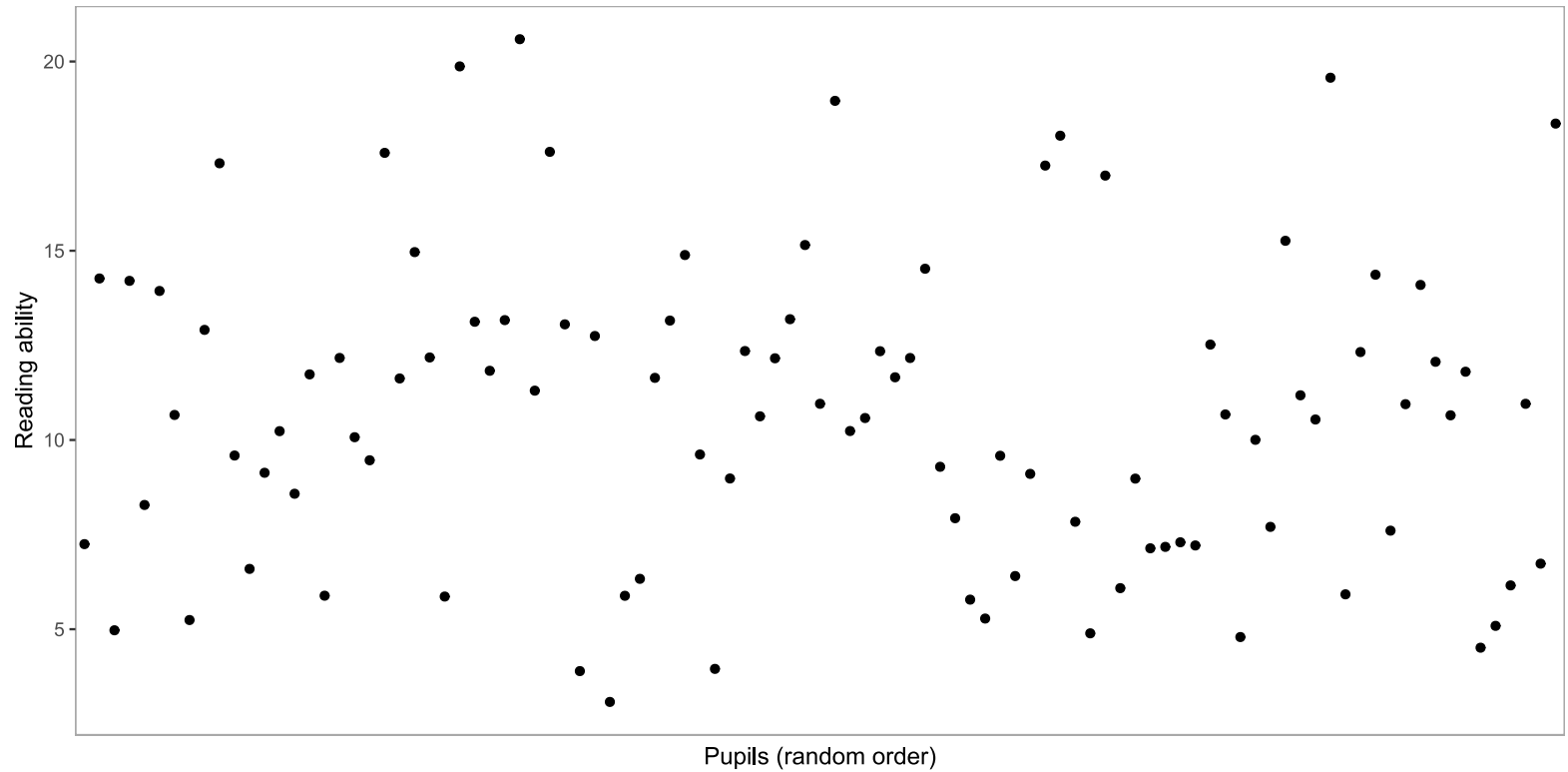
To tutorials available here:

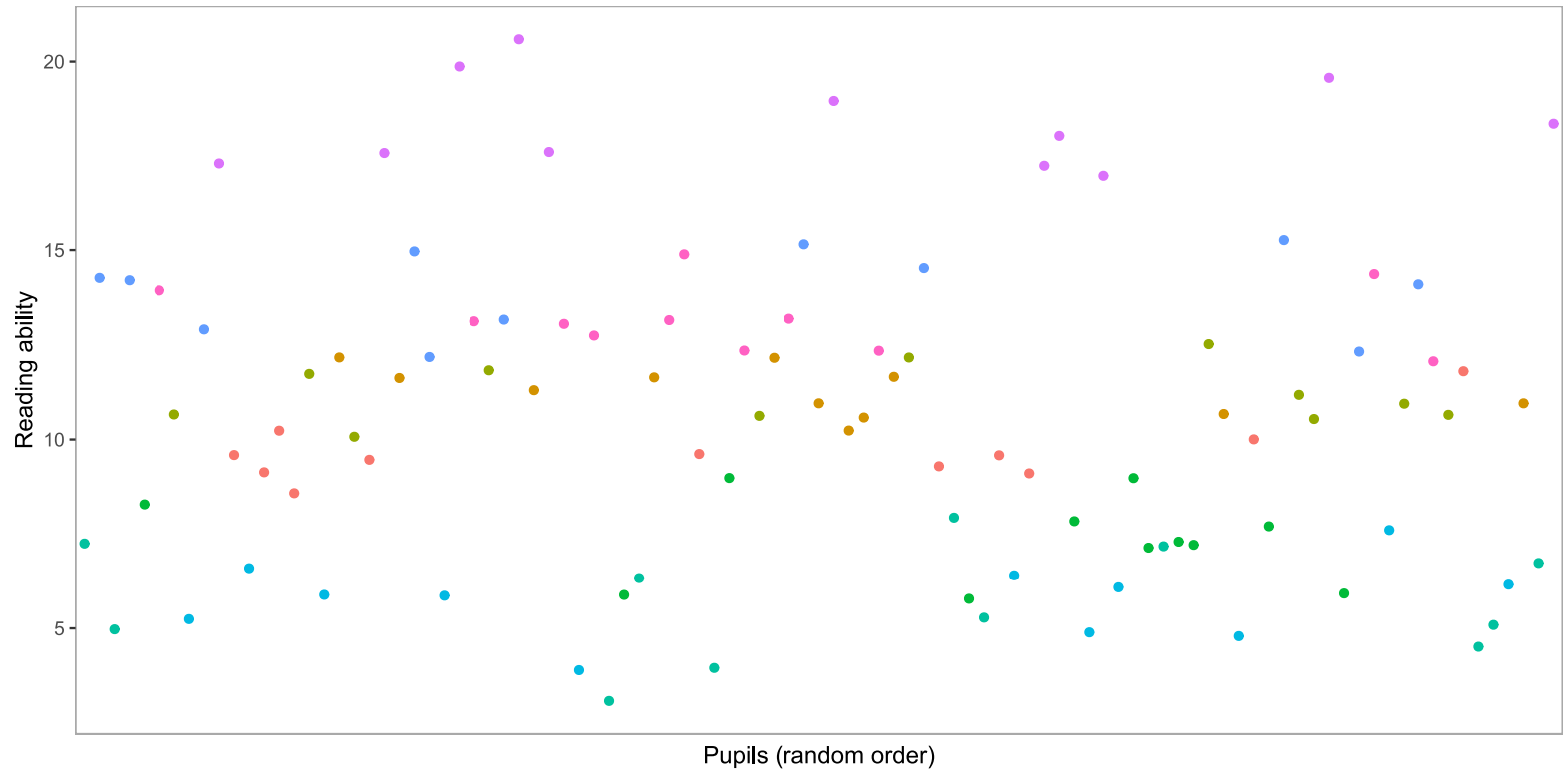
Clustered data: <https://shinyibv02.uio.no/connect/#/apps/433caf55-e0da-477a-920e-bc044ef5030f/access>

Longitudinal analysis: <https://shinyibv02.uio.no/connect/#/apps/bf3cac08-66ca-45d0-a45c-9b9402aa6c84/access>

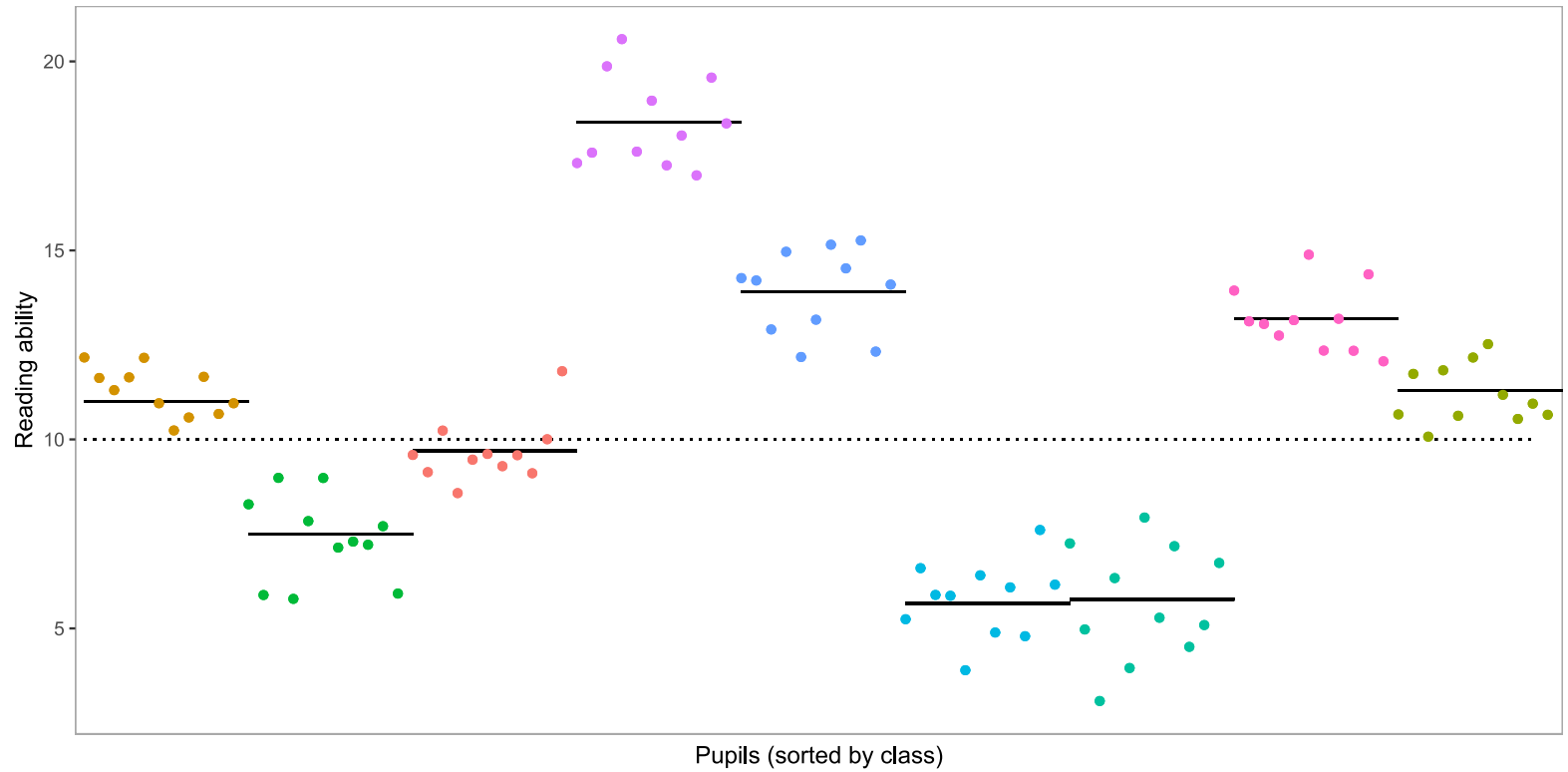
SECTION: INTRODUCING LONGITUDINAL ANALYSIS

MULTILEVEL MODELS FOR CLUSTERED DATA (1)

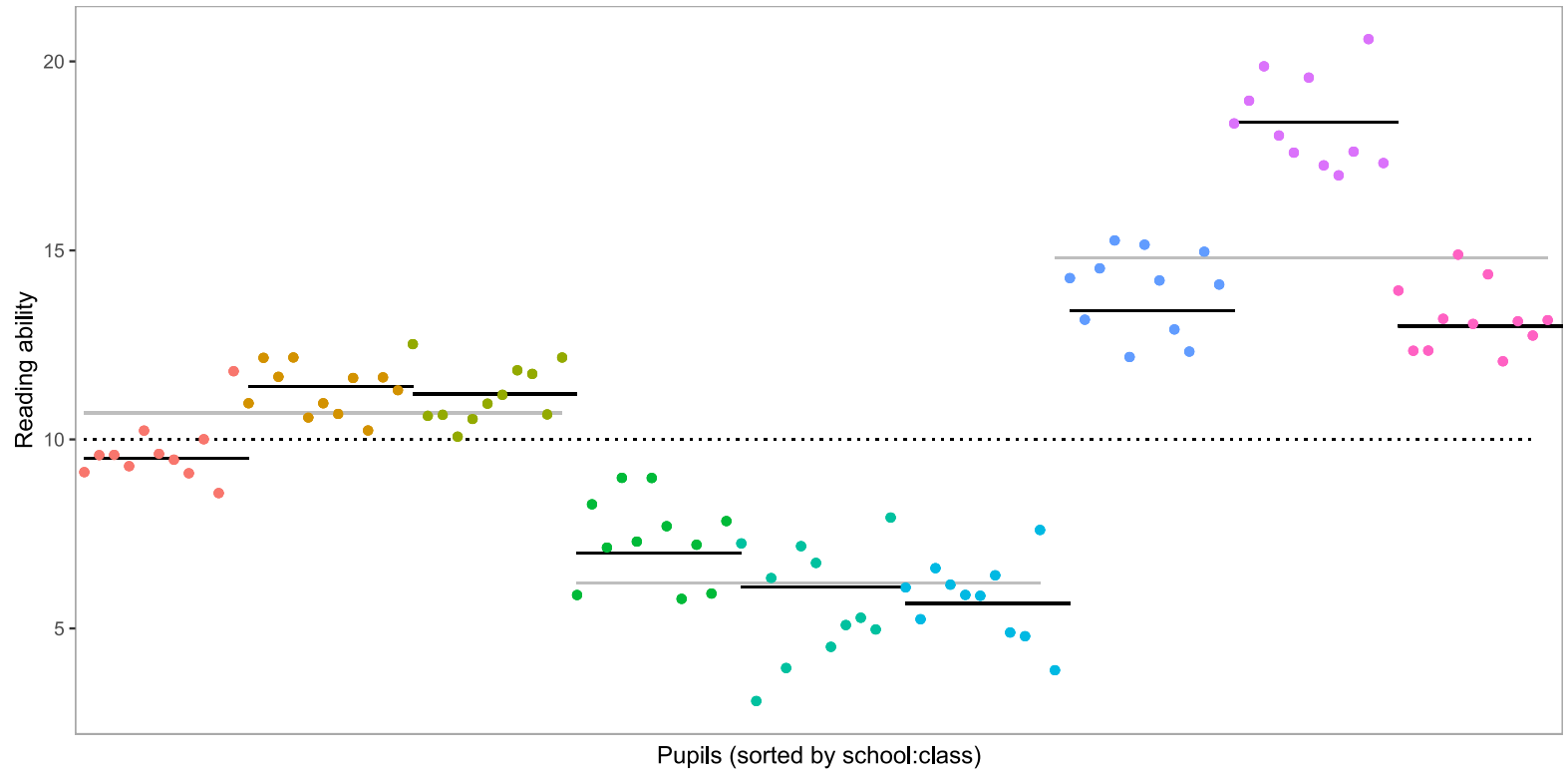




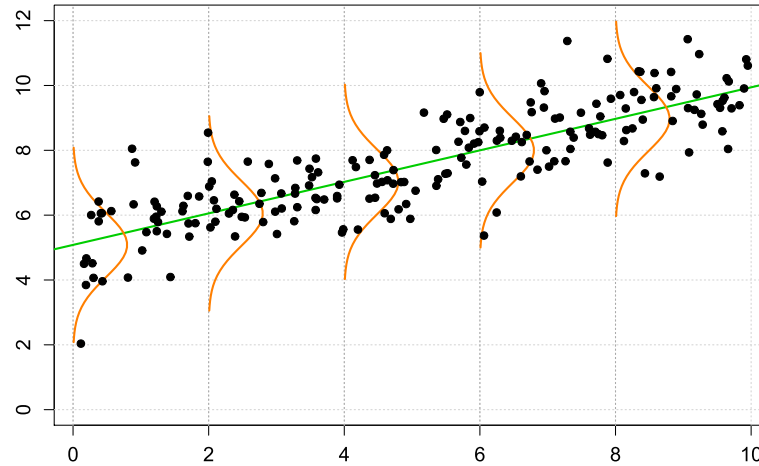
CLASSES 3



CLASSES 4

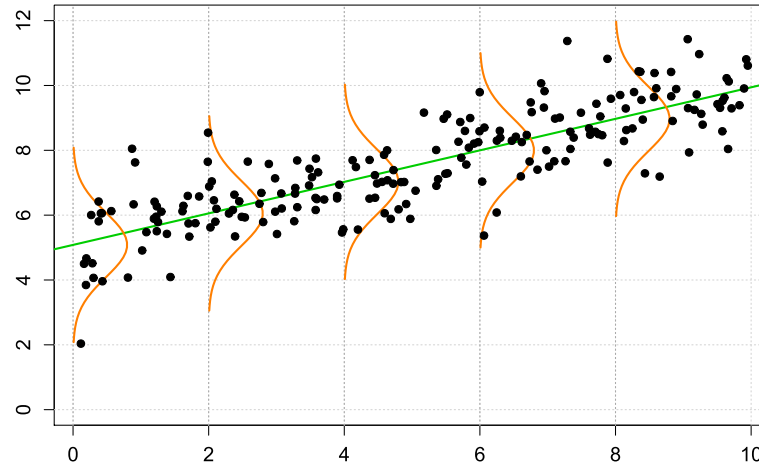


TWO SIDES OF ALL MODELS



- Models are an idealized account of:
 1. How the expected value is related to an independent variable (*mean*).
 2. How people vary (*variance*).
- Linear regression only models the mean value, not the variance.

ASSUMPTIONS IN LINEAR MODELS



$$Y_i = b_0 + b_1 \cdot X_i + \epsilon_i \quad \epsilon_i \sim N(0, \sigma^2)$$

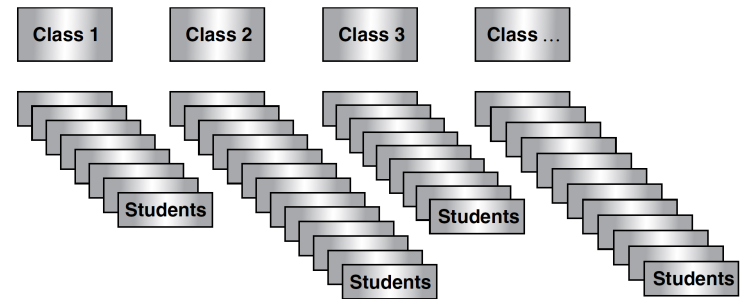
1. The relationship between X and Y is linear.
2. The error variance (around the regression line) is *normally distributed* and is constant (i.e. the sd is the same) at every level of X.
3. The observations are independent.
4. The independent variables are measured without error.

WHICH ASSUMPTION IS REALLY IMPORTANT?

- We often focus on the extent to which the residuals are normally distributed, but in practise this rarely has a major impact on the standard errors (or p-values).
- *Dependency* in the data can render all p-values completely biased.

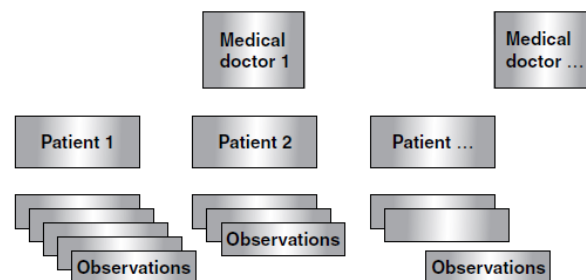
MULTILEVEL DATA STRUCTURES

- Data can be *nested* within higher order units.
- Different variables are measured at different levels.
 - Traditionally handled by *aggregation* or *disaggregation*.



LONGITUDINAL DATA STRUCTURES

- In longitudinal multilevel models, observations are nested within persons.
 - Level 1 units are individual observations
 - Level 2 units are usually individuals
 - People may be nested under higher order units (therapists, treatment centers, etc).
- *Time* will be included as a level 1 predictor.



WHAT ARE LONGITUDINAL MODELS?

- **Time course data** are the result of repeated measurements at multiple (at least two) time points.
 - These sorts of data are also called *longitudinal*.
- Two key properties distinguish time course data from other kinds of data:
 1. Groups of observations all come from one source (nested data).
 2. The repeated measurements are *related by a continuous variable*, usually that variable is time, but any continuous variable will do.
 - We will also look at other variables, like *set-size*.
- For analyses we will use multilevel (mixed) models that include time course data as a *level-1 predictor*.

BENEFITS OF MULTILEVEL MODELS IN GENERAL

- **Flexibility in modelling dependency across observations**
 - LMMs provide a flexible framework to model and account for various sources of non-independence, reducing biased estimates and increasing the accuracy of inferences.
- **Avoid choosing between individuals or groups as the unit of analysis**
 - Cross-level interactions are uniquely possible in MLM.
- **Increased statistical power**
 - By including random effects, LMMs can improve the precision of estimates and increase statistical power compared to traditional regression models.
- **Generalizability**
 - LMMs allow researchers to investigate both the average effect of predictors and their variations across different levels, helping to generalize findings across various contexts and populations.
- **Study variation between and within people**
 - Ask more complex and interesting questions

BENEFITS OF MLM FOR LONGITUDINAL DATA

- **Handle dependency**
 - Traditional analyses only designed to handle dependency due to constant mean differences.
- **Handles missing**
 - Number of observations can vary between individuals.
- **Flexible treatment of time**
 - Measures can be taken at fixed or at varying occasions.
 - No assumption that an equal amount of time has elapsed.
- **Investigate within-person relationships**
 - Between and within effects can differ, even in direction.

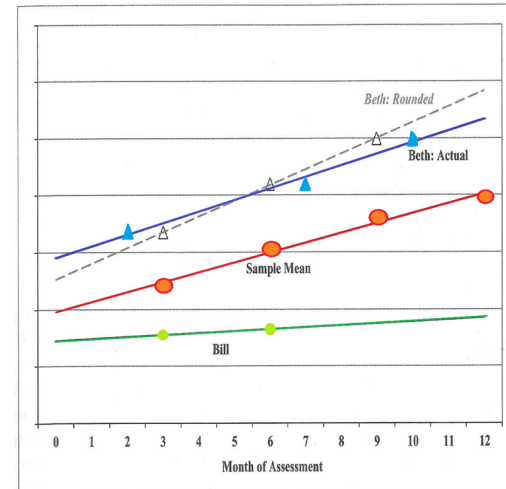


Figure 1.1 Example of incomplete and unbalanced time.

- <https://www.bayesrulesbook.com/chapter-15.html>

SECTION: PLOTTING LONGITUDINAL DATA

Plotting is an even more critical first step longitudinal analysis than cross-sectional studies.

We will make the following classes of figures:

- Individual change curves.
- Plots of group mean levels and CI/confidence regions.
- Stratified plots (assess interactions).

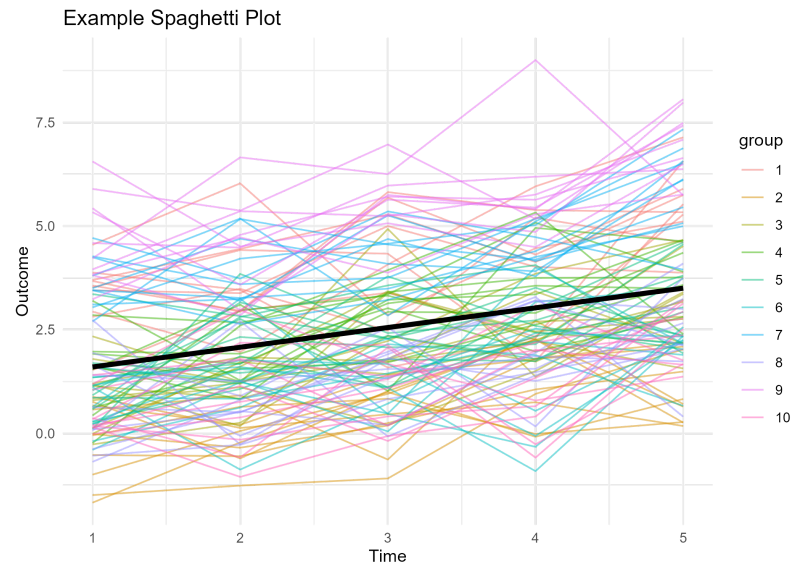
Why ggplot?

1. Plots are publication grade quality.
2. The syntax is flexible and powerful framework for visualizing data.
3. Summary statistics like means and standard errors can be computed while plotting.

INTRODUCTION TO GGPLOT2

- ggplot2 is a powerful R package for data visualization
- Built on the principles of the "Grammar of Graphics"
 - first assign variables in your data to properties of the graph. These assignments or mappings are called the aesthetics of your graph. Then you select “geometries,” or geoms — points, lines, bars, etc. — for those aesthetics.
- Create complex, multi-layered plots using a consistent syntax
- Easily customizable and extensible
- Widely used in data analysis and research

SPAGHETTI PLOTS FOR LONGITUDINAL DATA



- Visualize individual trajectories over time
- Identify patterns and trends in the data
- Assess variability between and within groups
- Detect potential outliers or influential cases
- Compare individual trajectories to the overall fixed effect of time

GENERATING A SPAGHETTI PLOT WITH GGPLOT2

Below is the R-code used to generate the plot.

```
library(ggplot2)

# Create spaghetti plot
ggplot(data, aes(x = time, y = y, group = id, color = group)) +
  geom_line(alpha = 0.5) +
  geom_line(data = fixed_effect_data, aes(x = time, y = y, group = 1), color = "black",
           size = 1.5, linetype = "solid") +
  labs(title = "Example Spaghetti Plot", x = "Time", y = "Outcome") +
  theme_minimal()

# Save plot to a file
ggsave("example_spaghetti_plot.png", width = 7, height = 5)
```


SECTION: TIME-ONLY (UNCONDITIONAL) MODELS

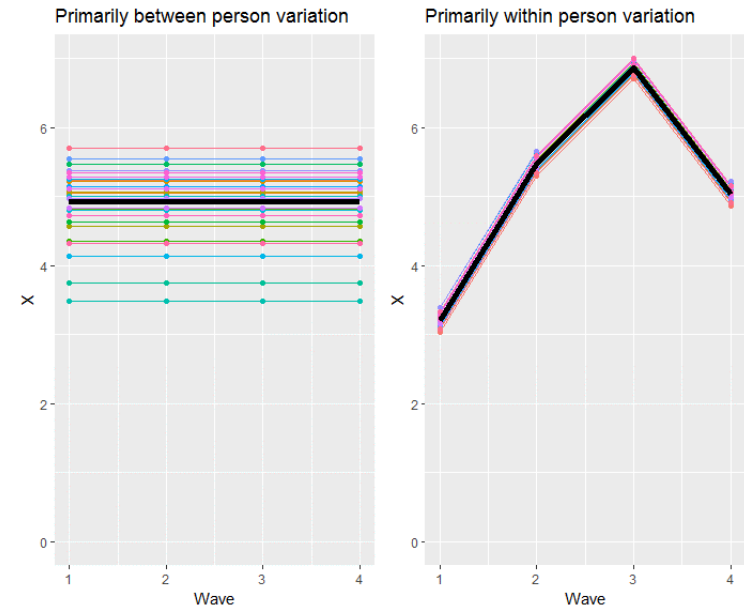
FEATURES OF LONGITUDINAL DATA

Multilevel longitudinal models permit analyses at different levels:

- **Between-person variance:**
 - Inter-individual variation.
 - Ex. Biological sex, ethnicity.
- **Within-person variance:**
 - Intraindividual variation.
 - Ex. Sleep the previous night.

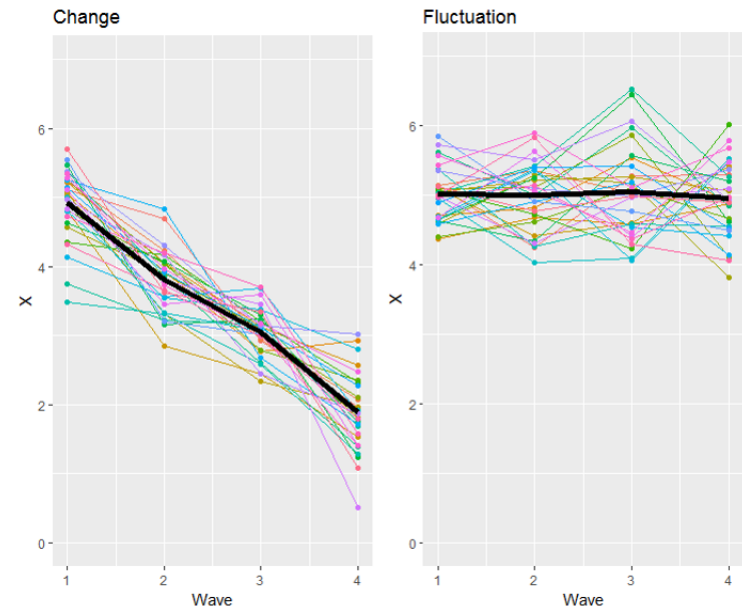
Relationships observed at the within-person level need not (and often will not) mirror those at the between-person level of analysis.

- Eg. resting heart rate vs. exercise.



FEATURES OF LONGITUDINAL DATA (2)

- **Within person change:** Specific type of within person variation, that refers to any systematic change that is expected as a result of meaningful passage of time.
- **Within person fluctuation:** Undirected variation over repeated assessments seen in contexts in which one would not expect systematic change.



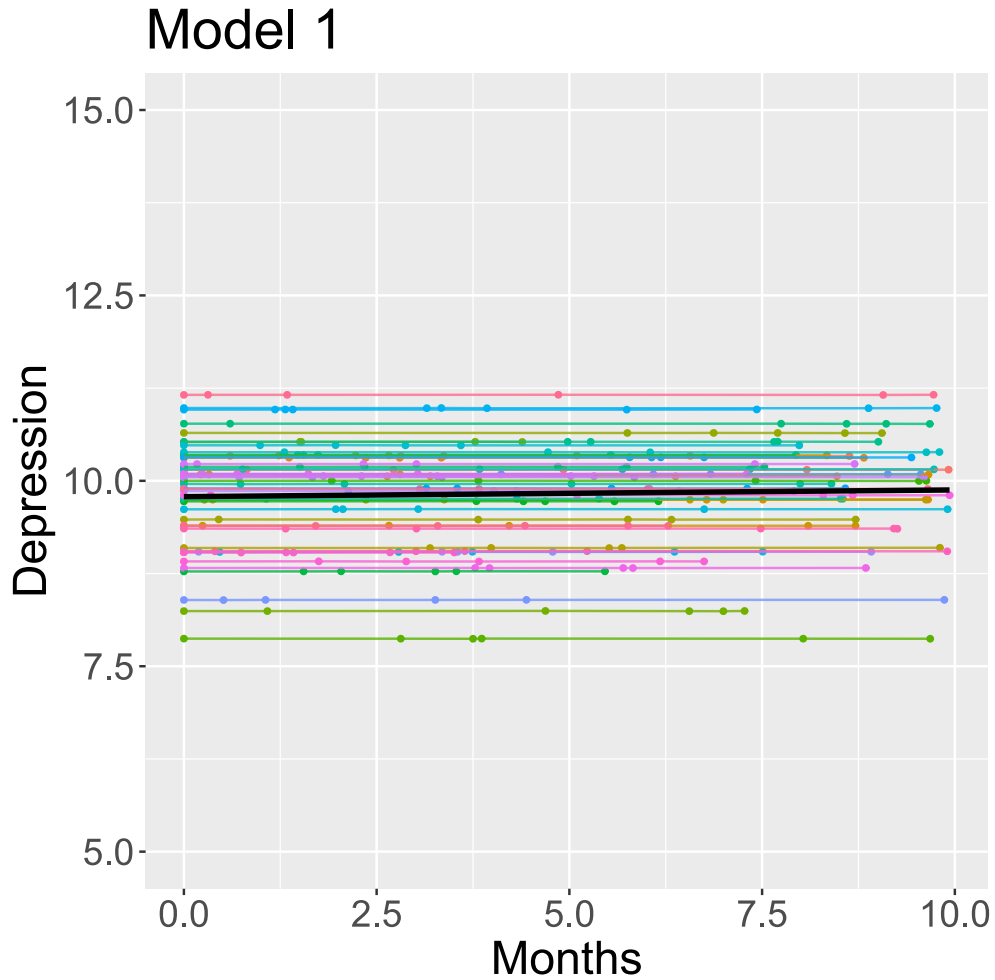
In **unconditional** longitudinal models, *time* is the only predictor.

- Included as a level-1 variable.

Critical first questions:

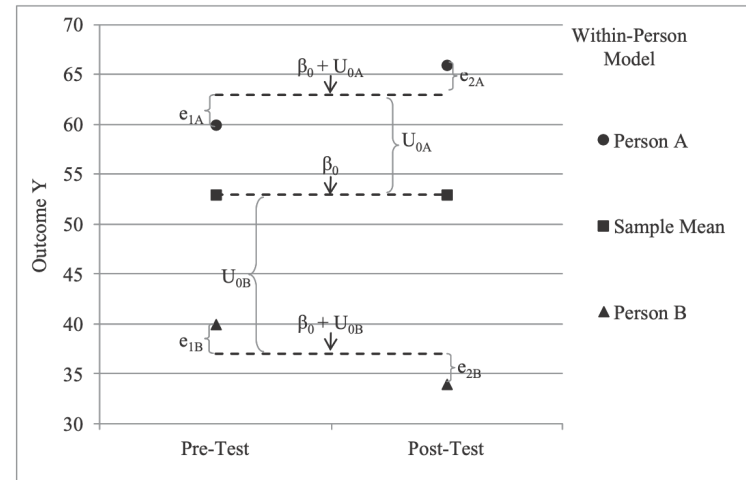
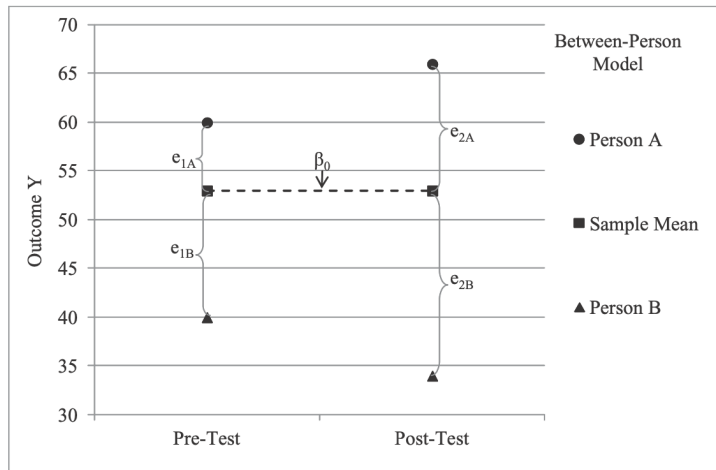
- What units should time be measured in?
- What constitutes "time=0"?
 - E.g. If *age*, it is best to center age on another value than "0".
- How do you expect development over time to be?
 - Linear, non.linear, abrupt/discontinuous?

INTRODUCING IDEALIZED GROWTH CURVES



BETWEEN-PERSON AND WITHIN-PERSON EMPTY MODELS

Empty models partition the variance, but don't account for it.



Between person empty model

$$y_{ti} = \underbrace{\beta_0}_{\text{Fixed Intercept}} + \underbrace{\epsilon_{ti}}_{\text{Error}}$$

Within person empty model

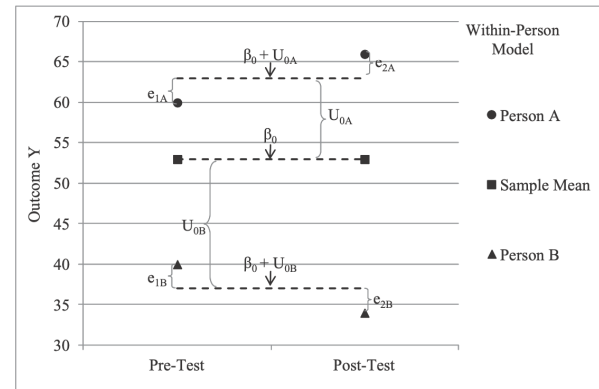
$$y_{ti} = \underbrace{\beta_0}_{\text{Fixed Intercept}} + \underbrace{U_{0i}}_{\text{Random Intercept}} + \underbrace{\epsilon_{ti}}_{\text{Error}}$$

FIXED AND RANDOM EFFECTS

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

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

$$\text{Composite: } y_{ti} = (\gamma_{00} + U_{0i}) + \epsilon_{ti}$$

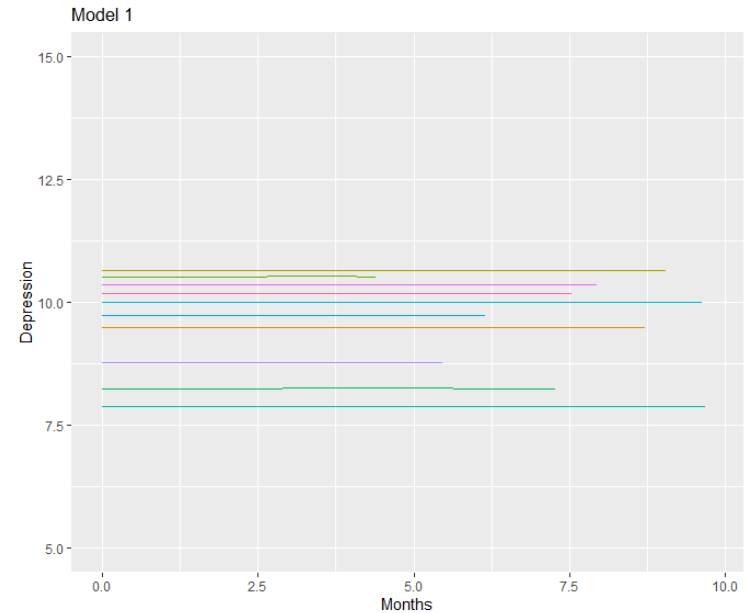


$$y_{ti} = \beta_0 + u_{0i} + \epsilon_{ti}$$

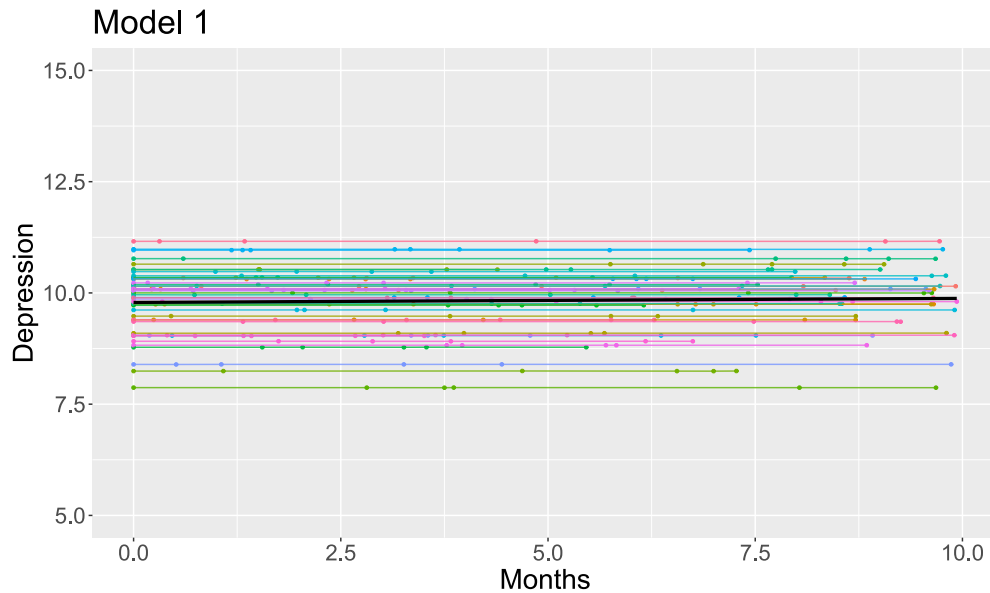
- We instead argue that deviation from the grand mean is attributable to a normally distributed residual at level 2. Now we can account for this variance by estimating a parameter (variance of U_{0i}) instead of estimating the fixed effects of 49 dummy variables. Random effects are not designed to make inferences between the specific variants included in the study (can't compare participant 1 to participant 50). Could we include a random effect of time? This would require at least three measurements per person.

WHY RANDOM INTERCEPT?

- Differences between individuals *could* be handled by N-1 dummy variables.
 - Many estimated parameters and loss of power.
- With MLM, clustering is accounted for by only one parameter, the *variance* of a normal distribution.



MODEL 1: RANDOM INTERCEPT



$$\text{Level 2: } \beta_{0i} = \gamma_{00} + U_{0i} \quad U_{0i} \sim N(0, \sigma_{U_{0i}})$$

$$\text{Level 1: } y_{ti} = \beta_{0i} + \epsilon_{ti} \quad \epsilon_{ti} \sim N(0, \sigma_{\epsilon})$$

Composite:

$$y_{ti} = \left(\underbrace{\gamma_{00}}_{\text{Fixed Intercept}} + \underbrace{U_{0i}}_{\text{Random Intercept}} \right) + \underbrace{\epsilon_{ti}}_{\text{Error}}$$

Multilevel (mixed) models contain two classes of coefficients.

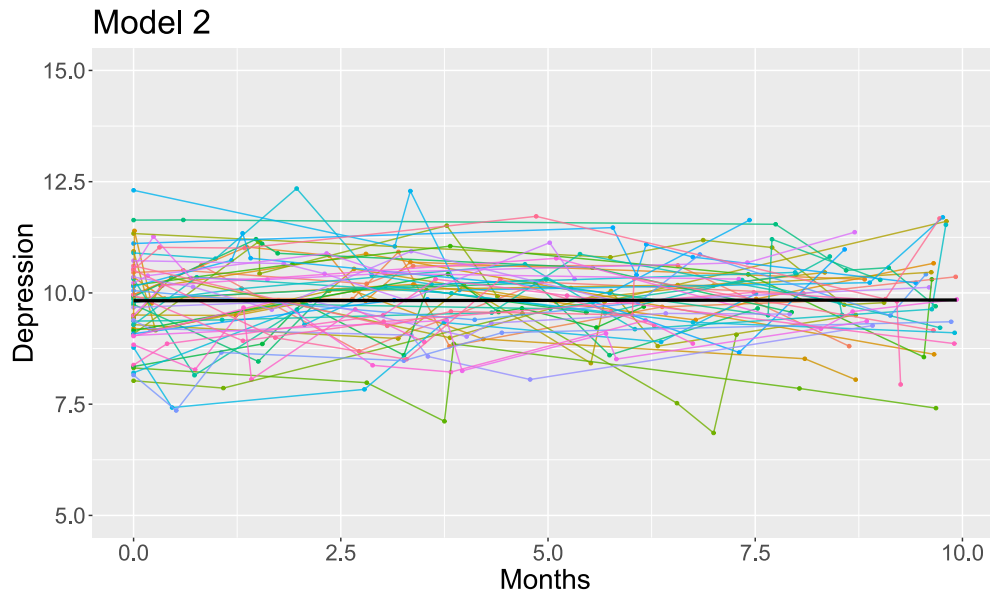
Fixed effects: The «structural» part of the model, specifying the expected conditional mean.

- Empty model: Only intercepts are included.
- Unconditional models: Intercepts + passage of time.
- Conditional models: Intercepts + variables accounting for differences in the passage of time.

Random effects: Coefficients that specify the stochastic (error) part of the model.

- Describes how the residuals of the Y outcome are distributed and related across the observations.

MODEL 2: RANDOM INTERCEPT



$$\text{Level 2: } \beta_{0i} = \gamma_{00} + U_{0i} \quad U_{0i} \sim N(0, \sigma_{U_{0i}})$$

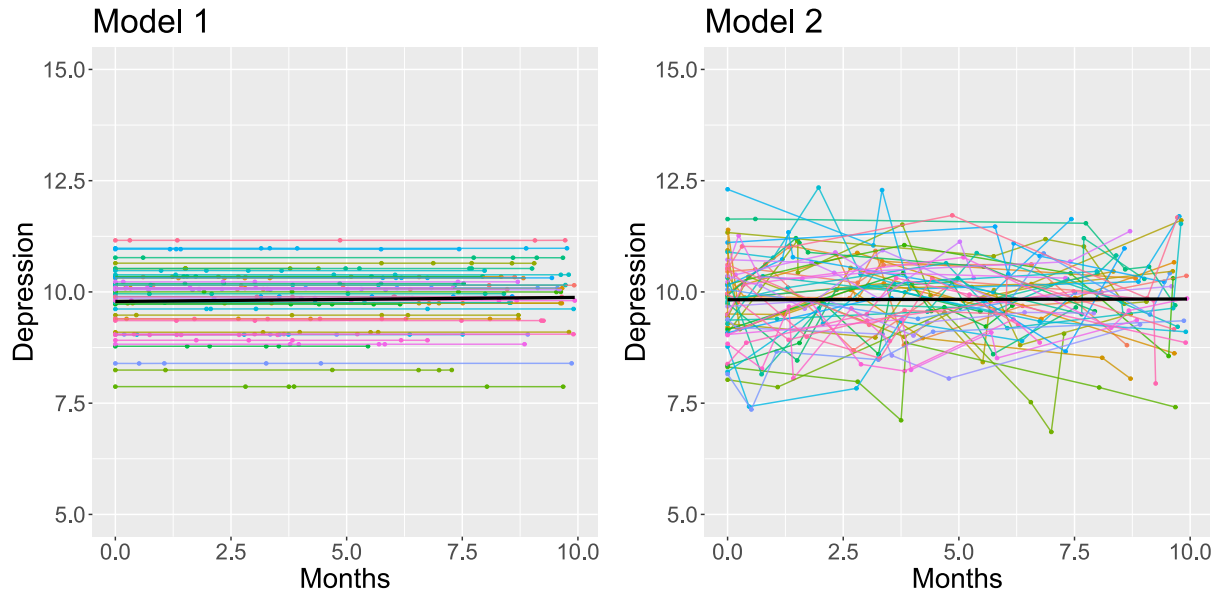
$$\text{Level 1: } y_{ti} = \beta_{0i} + \epsilon_{ti} \quad \epsilon_{ti} \sim N(0, \sigma_{\epsilon})$$

$$\text{Composite: } y_{ti} = (\gamma_{00} + U_{0i}) + \epsilon_{ti}$$

INTRACLASS CORRELATION COEFFICIENT (ICC)

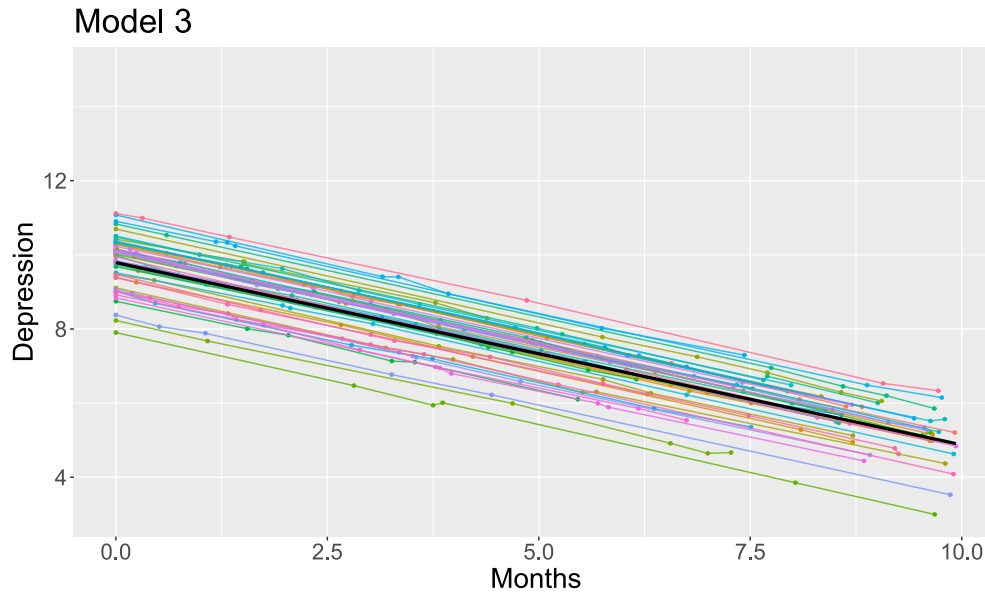
- Measures the **proportion of total variance** in the outcome that is attributable to the grouping structure (e.g., individuals within groups)
- Ranges from 0 to 1:
 - 0: No correlation between observations within the same group
 - 1: Observations within the same group are identical
- Important for assessing **group-level effects** in mixed models
- A high ICC indicates a strong grouping effect and the need for a multilevel or mixed model to account for the clustered data structure

INTRACLASS CORRELATION COEFFICIENT (ICC) 2



$$ICC = \frac{BP \text{ variation}}{BP + WP \text{ variation}} = \frac{Var(U_{0i})}{Var(U_{0i}) + Var(\epsilon_{ti})}$$

MODEL 3: FIXED SLOPE, RANDOM INTERCEPT



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

$$\text{Level 2: } \beta_{1i} = \gamma_{10}$$

$$\text{Level 1: } y_{ti} = \beta_{0i} + \beta_{1i} \cdot \text{Time}_{ti} + \epsilon_{ti}$$

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

$$y_{ti} = \left(\underbrace{\gamma_{00}}_{\text{Fixed Intercept}} + \underbrace{U_{0i}}_{\text{Random Intercept}} \right) + \underbrace{\gamma_{10}}_{\text{Fixed Slope}} \cdot \text{Time}_{ti} + \underbrace{\epsilon_{ti}}_{\text{Error}}$$

- **Likelihood Ratio Tests:**

- Performed with the `anova()` function.
- Only valid under maximum likelihood (ML) estimation, not restricted maximum likelihood (REML).

- **Information Criteria:**

- Akaike Information Criterion (AIC) can be calculated using the `AIC()` function.

- **Pseudo R^2 :**

- **Purpose:** Assess the proportion of variance explained by the predictors in the model, similar to R^2 in linear regression.
- Calculated separately for Level 1 and Level 2 effects.
- **Range:** 0 to 1; higher values indicate better model fit.
- **Limitation:** Not directly comparable to R^2 in linear regression. Interpretation should be cautious.

OVER VS. UNDERFITTING (SCYLLA OG CHARYBDIS)



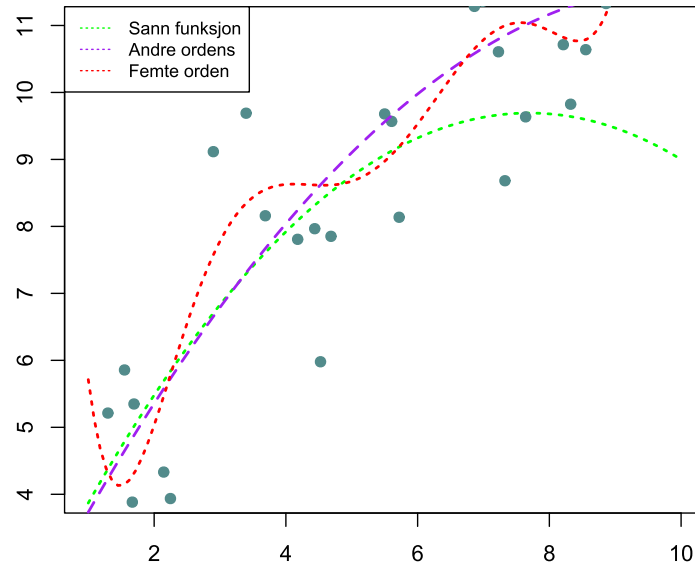
Overfitting: Poor prediction as a result of learning too much from your data.

Underfitting: Poor prediction as a result of learning too little from your data.

- You can always get better fit to your data by adding more predictors.
- Every dataset contains both systematic and unsystematic variance (noise), so overly complex models may fit well to the current data, but predict less variance in *new* data.

<https://xcelab.net/rm/statistical-rethinking/>

WHAT HAPPENS AS THE MODEL BECOMES TOO COMPLEX?



Over: The green curve is the *true* (generating) function. The red fits better, but adapts to a lot of the noise in the data.

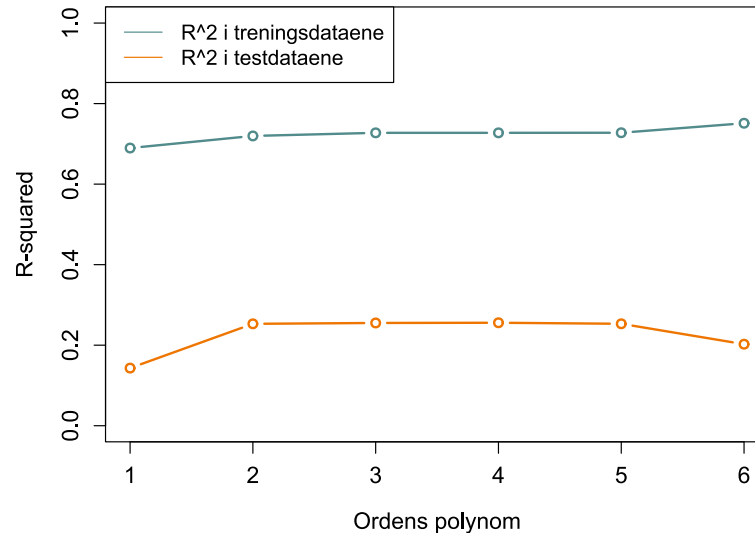
K-FOLD CROSS VALIDATION



- **Cross-validation** involves partitioning the data into *training* and *test* subsets.
- The model is built based on the training sets, and evaluated on the test sets.
 - Cornerstone in machine learning approaches.

(Introduction to statistical learning)

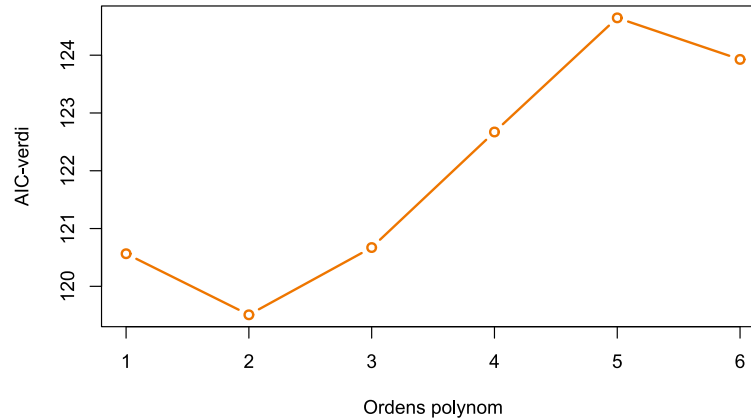
K-FOLD KRYSSVALIDERING (2)



Notice:

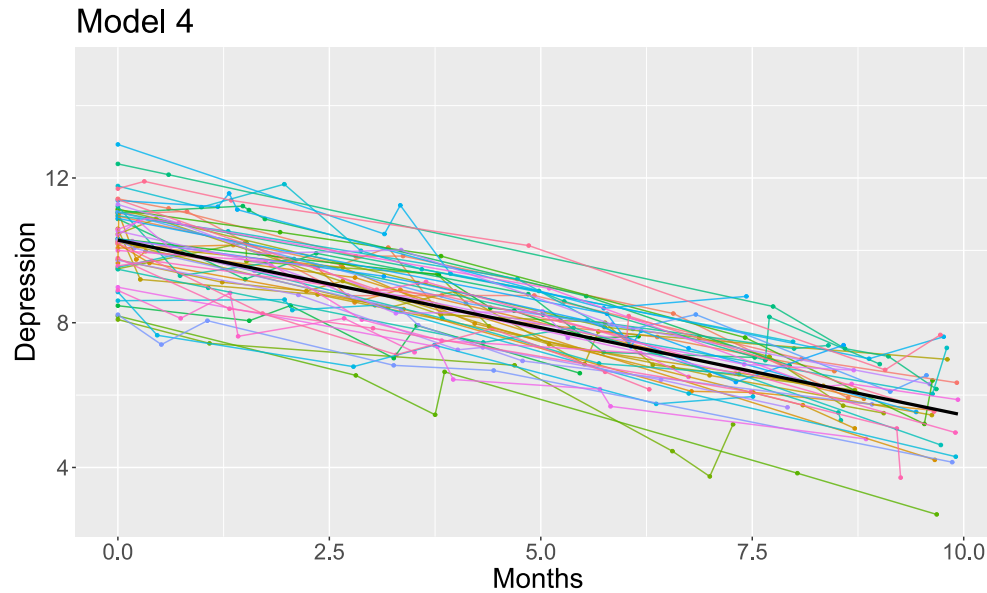
1. It is easy to explain variance in the training data, but far more difficult to explain it in (new) test data.
2. Initially, as the model grows in complexity, explained variance increases. However, as it gets overly complex, explained variance in test data starts to *decrease*.

METODE 3: INFORMATION CRITERIA (AIC / BIC)



- **Information criteria** (IC) is a class of statistics devised to strike the optimal between over and underfitting.
- The model with the *lowest* AIC/BIC value is chosen as the best one.
 - Non-nested models can be compared.

MODEL 4: FIXED SLOPE, RANDOM INTERCEPT



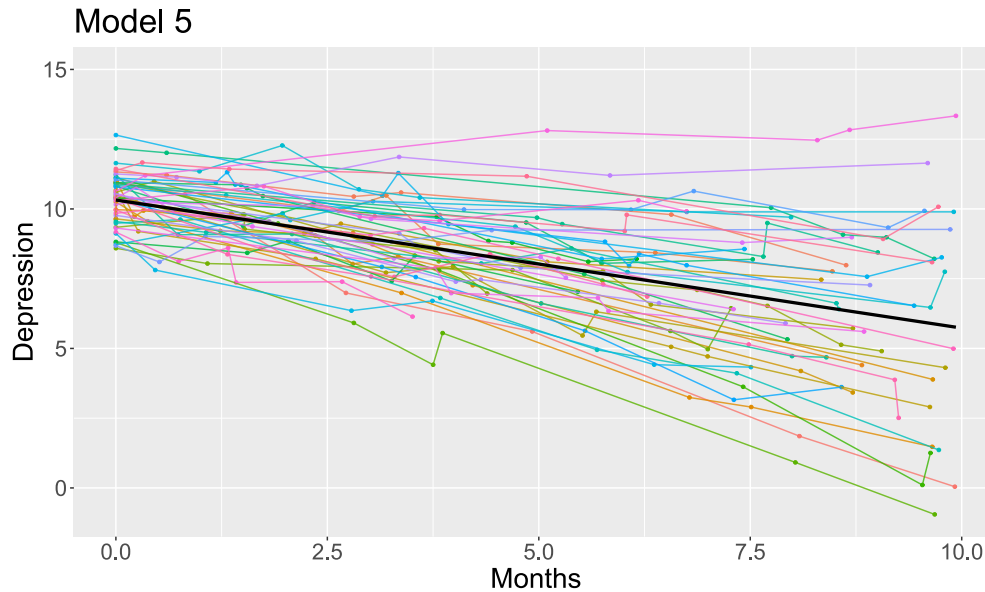
Level 2: $\beta_{0i} = \gamma_{00} + U_{0i}$

Level 2: $\beta_{1i} = \gamma_{10}$

Level 1: $y_{ti} = \beta_{0i} + \beta_{1i} \cdot Time_{ti} + \epsilon_{ti}$

Composite: $y_{ti} = (\gamma_{00} + U_{0i}) + \gamma_{10} \cdot Time_{ti} + \epsilon_{ti}$

MODEL 5: RANDOM SLOPE, RANDOM INTERCEPT



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

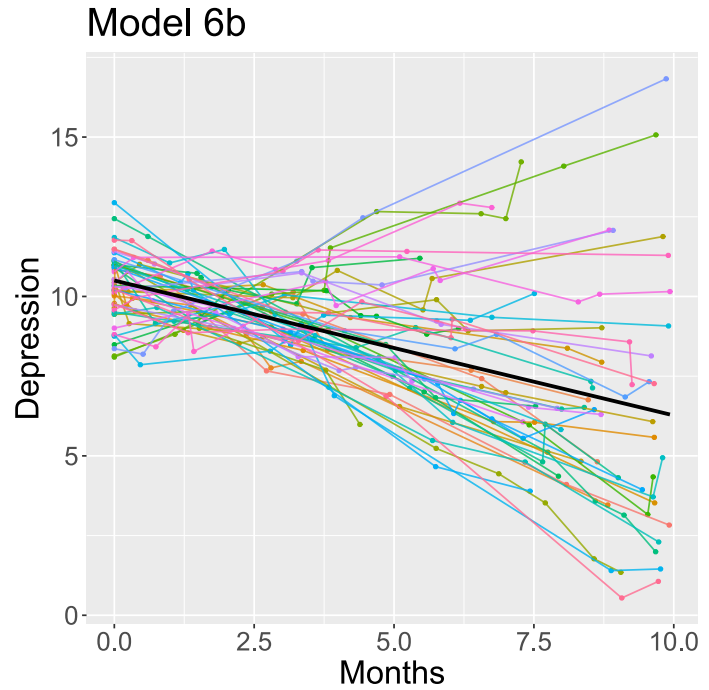
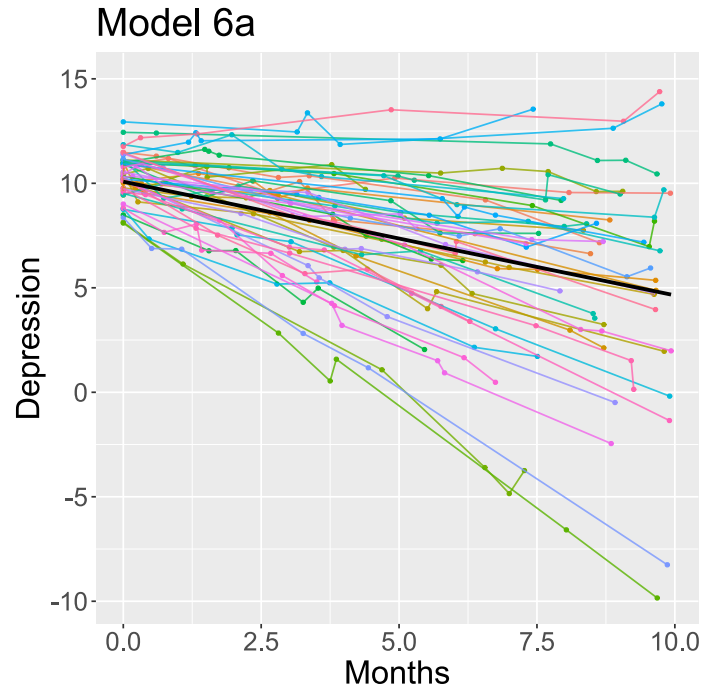
$$\text{Level 2: } \beta_{1i} = \gamma_{10} + U_{1i}$$

$$\text{Level 1: } y_{ti} = \beta_{0i} + \beta_{1i} \cdot \text{Time}_{ti} + \epsilon_{ti}$$

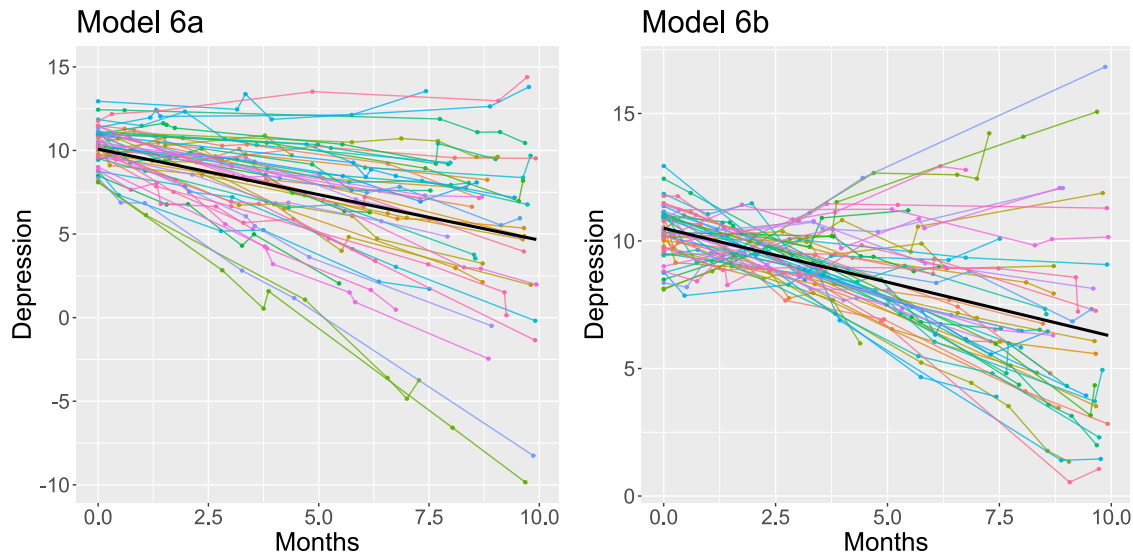
Composite:

$$y_{ti} = \underbrace{(\gamma_{00} + U_{0i})}_{\text{Fixed Intercept} + \text{Random Intercept}} + \underbrace{(\gamma_{10} + U_{1i})}_{\text{Fixed Slope} + \text{Random Slope}} \cdot \text{Time}_{ti} + \underbrace{\epsilon_{ti}}_{\text{Error}}$$

HOW DO THESE DIFFER?



RANDOM EFFECTS CAN BE CORRELATED



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

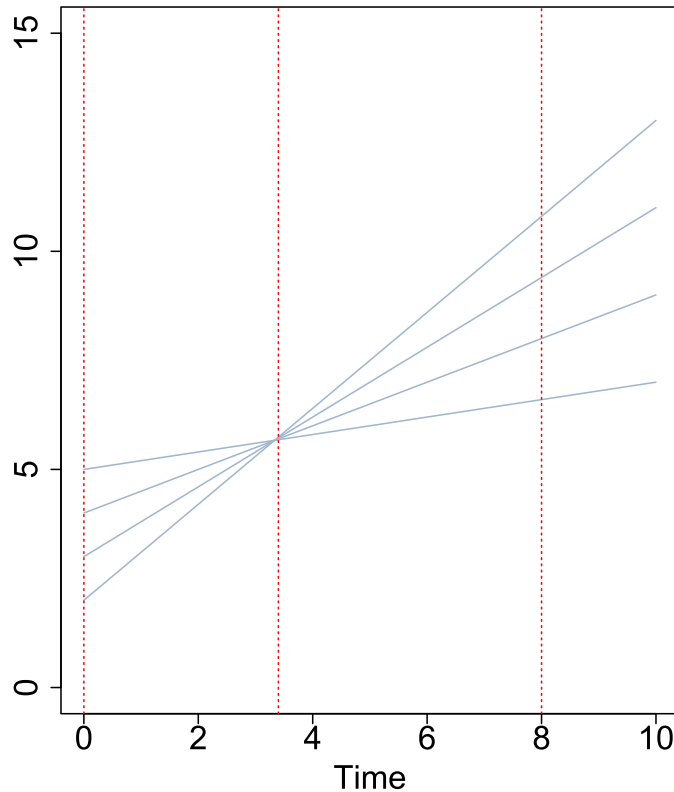
Level 2: $\beta_{1i} = \gamma_{10} + U_{1i}$

Level 1: $y_{ti} = \beta_{0i} + \beta_{1i} \cdot Time_{ti} + \epsilon_{ti}$

$$G = \begin{bmatrix} Var(U_{0i}) & Cov(U_{0i}, U_{1i}) \\ Cov(U_{0i}, U_{1i}) & Var(U_{1i}) \end{bmatrix}$$

Composite: $y_{ti} = (\gamma_{00} + U_{0i}) + (\gamma_{10} + U_{1i}) \cdot Time_{ti} + \epsilon_{ti}$

SENSITIVITY OF THE INTERCEPT-SLOPE CORRELATION TO THE CENTERING OF TIME



- **Impact on Interpretability:**

- Centering time can affect the interpretability of the (fixed) intercept.

- **Intercept-Slope Correlation:**

- The interpretation of any (random) intercept-slope correlation is conditional on the location of the intercept.

- **Example**

- Centering at $t=0$ will result in an estimated correlation less than 0 (< 0).
- Centering at $t=3.5$ will result in an estimated correlation equal to 0 ($= 0$).

$$\text{Level - 2 } \mathbf{G} = \begin{bmatrix} \tau_{U_0}^2 & \tau_{U_{01}} \\ \tau_{U_{01}} & \tau_{U_1}^2 \end{bmatrix}, \text{ Level - 1 } \mathbf{R} = \begin{bmatrix} \sigma_e^2 & 0 & 0 & 0 \\ 0 & \sigma_e^2 & 0 & 0 \\ 0 & 0 & \sigma_e^2 & 0 \\ 0 & 0 & 0 & \sigma_e^2 \end{bmatrix}$$

The model implied covariance matrix is a function of two matrices:

- **G** is a covariance matrix for level-2 random coefficients.
- **R** is a covariance matrix for level-1 random coefficients.
 - Usually the R matrix is diagonal, but some programs allow you to specify a different structure for this matrix.

CALCULATING THE IMPLIED COVARIANCE (1)

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

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

The implied covariance matrix is a function of both the G and R matrices.

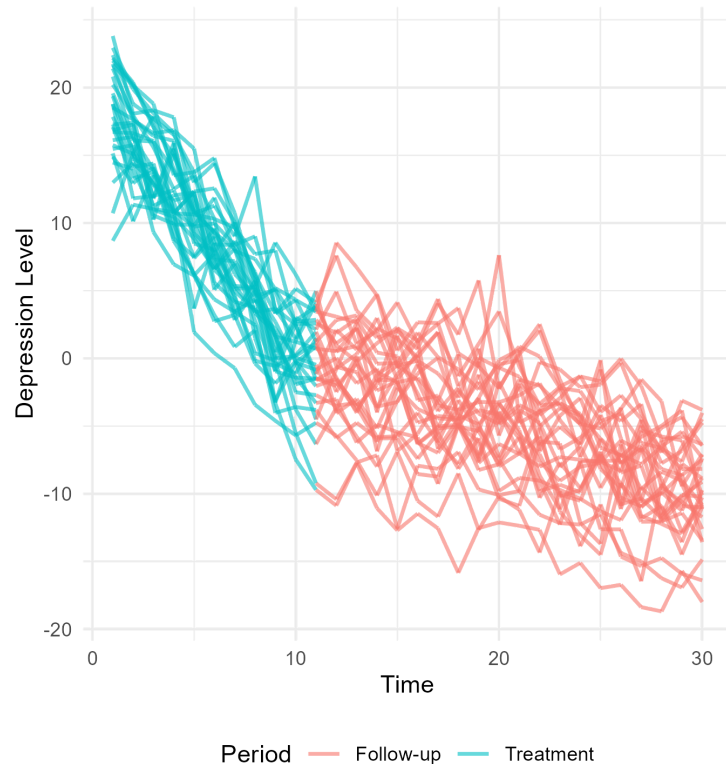
TECHNICAL DETAIL: CALCULATING THE IMPLIED COVARIANCE

$$\mathbf{V} = \begin{bmatrix} 1 & 0.0 & 0 & 0 \\ 1 & 1.0 & 0 & 0 \\ 1 & 2.0 & 0 & 0 \\ 1 & 3.0 & 0 & 0 \\ 0 & 0 & 1 & 0.2 \\ 0 & 0 & 1 & 1.4 \\ 0 & 0 & 1 & 3.5 \end{bmatrix} \begin{bmatrix} 2.26 & 0.05 & 0 & 0 \\ 0.05 & 0.91 & 0 & 0 \\ 0 & 0 & 2.26 & 0.05 \\ 0 & 0 & 0.05 & 0.91 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0.0 & 1.0 & 2.0 & 3.0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0.2 & 1.4 & 3.5 \end{bmatrix} + \begin{bmatrix} 0.70 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.70 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.70 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.70 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.70 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.70 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.70 \end{bmatrix}$$

$$\begin{bmatrix} 2.26 & 2.32 & 2.37 & 2.43 & 0 & 0 & 0 \\ 2.32 & 3.28 & 4.25 & 5.21 & 0 & 0 & 0 \\ 2.37 & 4.25 & 6.12 & 7.99 & 0 & 0 & 0 \\ 2.43 & 5.21 & 7.99 & 10.77 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2.32 & 2.61 & 3.10 \\ 0 & 0 & 0 & 0 & 2.61 & 4.20 & 6.99 \\ 0 & 0 & 0 & 0 & 3.10 & 6.99 & 13.78 \end{bmatrix} \begin{bmatrix} 0.70 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.70 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.70 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.70 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.70 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.70 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.70 \end{bmatrix} = \begin{bmatrix} 2.96 & 2.32 & 2.37 & 2.43 & 0 & 0 & 0 \\ 2.32 & 3.98 & 4.25 & 5.21 & 0 & 0 & 0 \\ 2.37 & 4.25 & 6.82 & 7.99 & 0 & 0 & 0 \\ 2.43 & 5.21 & 7.99 & 11.47 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3.02 & 2.61 & 3.10 \\ 0 & 0 & 0 & 0 & 2.61 & 4.90 & 6.99 \\ 0 & 0 & 0 & 0 & 3.10 & 6.99 & 14.48 \end{bmatrix}$$

MODELLING NON-LINEAR CHANGE OVER TIME

MODELLING NON-LINEAR CHANGE OVER TIME

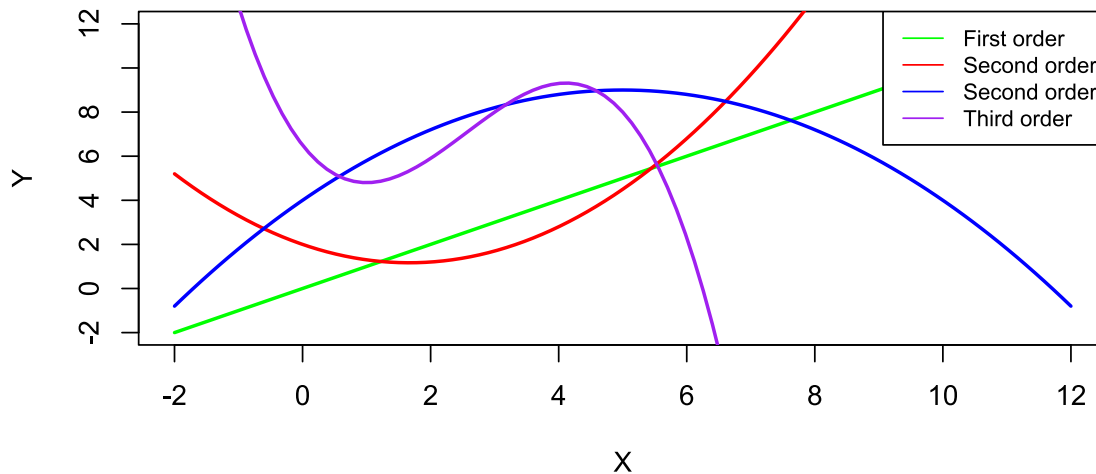


- Change considered until now has been purely *linear*.
- General Approaches to Dealing with Non-linearity
 - Polynomial models
 - Piecewise-discontinuous models
 - Splines

USING POLYNOMIALS TO MODEL NON-LINEAR CHANGE

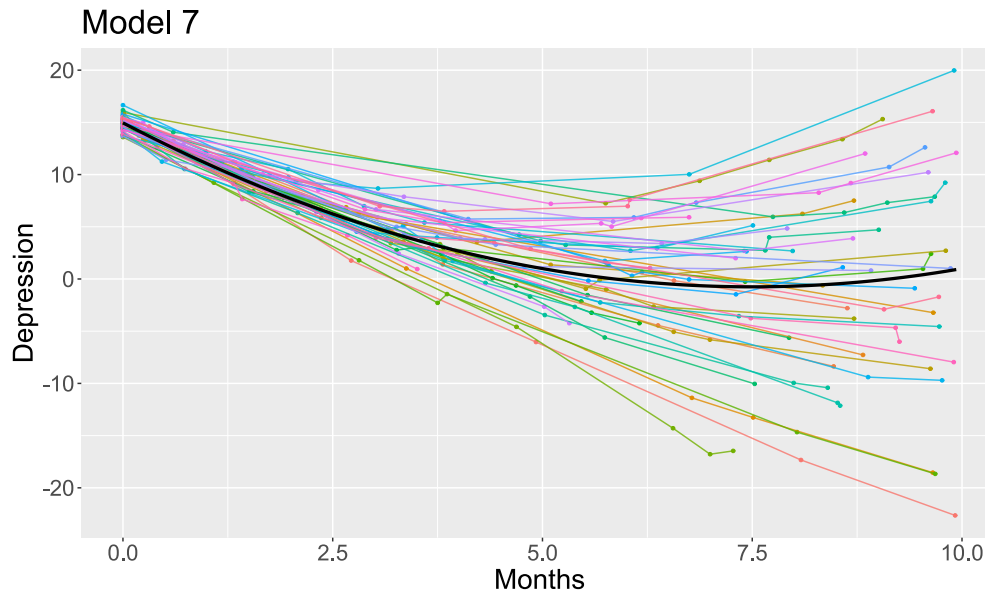
- Linear regression is more flexible than the name implies, and can be used to model (some) non-linear relationships.
- Traditionally **polynomials** used to model non-linear relationships.

$$E(Y|X) = b_0 + b_1 \cdot X + b_2 \cdot X^2 + \dots + b_p \cdot X^p$$



- Polynomial functions are only valid in a restricted range.
- Intercept and linear terms must be included in the model for it to be meaningful (regardless of significance).

MODEL 7: POLYNOMIAL CHANGE



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

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

Level 2: $\beta_{2i} = \gamma_{20} + U_{2i}$

Level 1: $y_{ti} = \beta_{0i} + \beta_{1i} \cdot Time_{ti} + \beta_{2i} \cdot Time_{ti}^2 + \epsilon_{ti}$

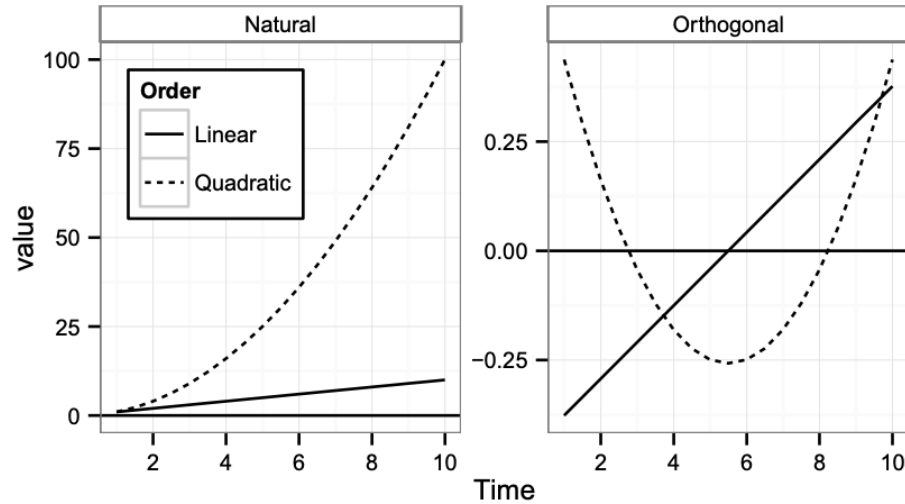
Composite:

$$y_{ti} = \underbrace{(\gamma_{00})}_{\text{Fixed Intercept}} + \underbrace{(U_{0i})}_{\text{Random Intercept}} + \underbrace{(\gamma_{10})}_{\text{Fixed Lin Slope}} + \underbrace{(U_{1i})}_{\text{Random Lin Slope}} \cdot Time_{ti} + \underbrace{(\gamma_{20})}_{\text{Fixed Quad Slope}} + \underbrace{(U_{2i})}_{\text{Random Quad Slope}} \cdot Time_{ti}^2 + \underbrace{\epsilon_{ti}}_{\text{Error}}$$

CHOOSING POLYNOMIAL DEGREE IN LONGITUDINAL MODELS

- **Consider Data:**
 - Think about the data in terms of the number of times the curve changes direction (corresponding to the number of inflection points).
- **Statistical Approach:**
 - Include only and all of the polynomial orders that improve model fit.
- **Theoretical Approach:**
 - Include only those terms for which the experimenter predicted an effect.

ORTHOGONAL POLYNOMIALS



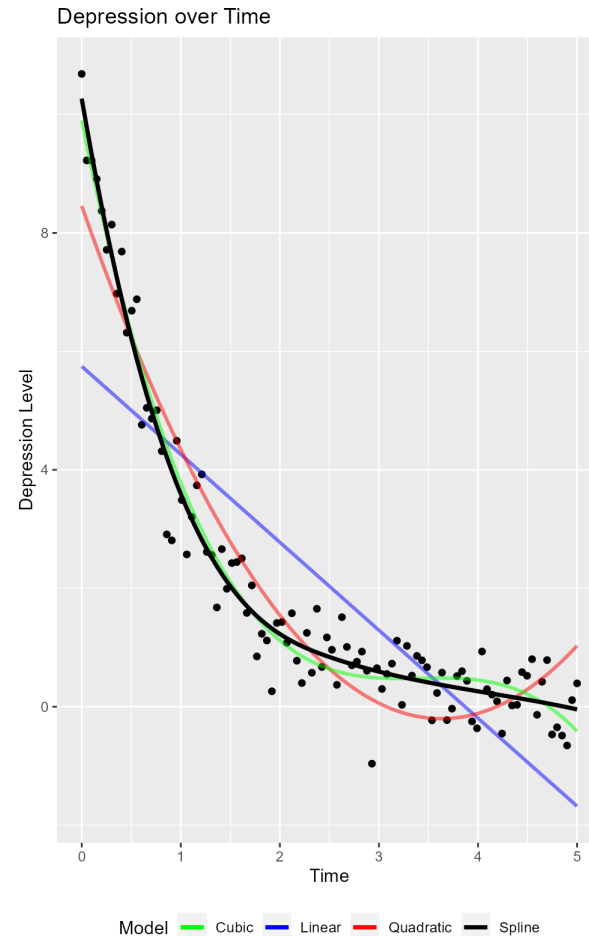
- **Natural Polynomials:**

- Allow testing for differences at "Time 0."
- Useful when such differences need to be tested.

- **Orthogonal Polynomials:**

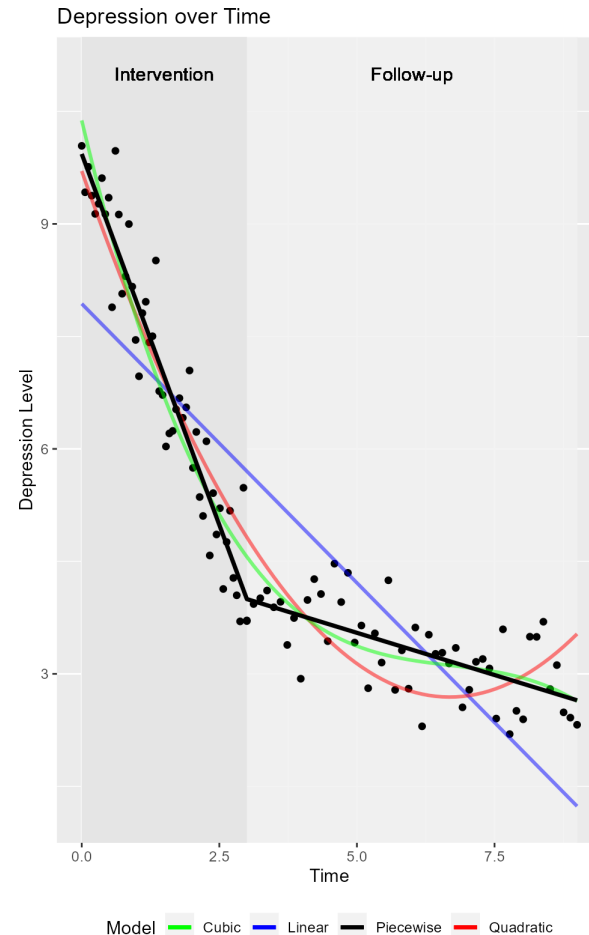
- Provide the same estimates as natural polynomials.
- Uncorrelated, so p-values may differ (sometimes considerably).

- Splines allow for flexibility in modeling complex, non-linear relationships
- They can capture local patterns and smooth changes over time
- Reduce overfitting compared to high-order polynomials
- Cubic splines are commonly used for their smoothness and continuity
- Natural splines can impose constraints to reduce extreme behavior at the endpoints



PIECEWISE LINEAR MODELS

- Better fit for discontinuous data:
 - More accurately capture abrupt changes
- Interpretability:
 - Coefficients of piecewise continuous models have simple interpretations
- Parsimonious representation:
 - May require fewer parameters than high-degree polynomial
- Avoid overfitting:
 - Less likely to overfit the data compared to high-degree polynomial



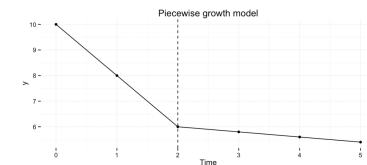
PIECEWISE LINEAR MODELS: CODING

Coding scheme 1: separate slopes

Time	0	1	2	3	4	5
Time 1	0	1	2	2	2	2
Time 2	0	0	0	1	2	3

Coding scheme 2: incremental/decremental slope

Time	0	1	2	3	4	5
Time 1	0	1	2	3	4	5
Time 2	0	0	0	1	2	3



These two coding schemes only differ in the interpretation of the regression coefficients.

- In scheme 1 the two slope coefficients represent the actual slope in the respective time period.
- In scheme 2 the coefficient for time 2 represents the *deviation* from the slope in period 1, + i.e. if the estimate is 0 then the rate of change is the same in both periods.

<https://rpsychologist.com/r-guide-longitudinal-lme-lmer#piecewise-growth-curve>

GETTING P-VALUES

- P-values are not reported by default, as the number of degrees of freedom is hard to calculate.
- Some approximations can be made, for example using the `lmerTest` package.

```
# P-values for individual coefficients
require(lmerTest)
m1 ← lmer(y ~ x | id), data=dt)
coef(summary(m1))

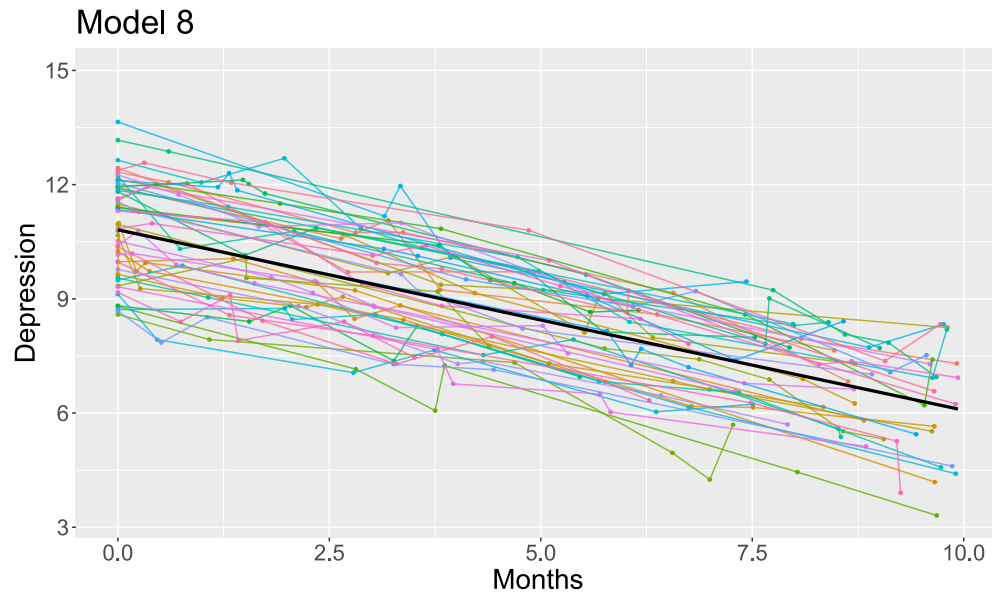
# Alternatively, confidence intervals
confint(vs_3) # Profile CI
confint(vs_3, method="boot", nsim=100) # Bootstrapped CI

# Alternatively, compare models with anova()
anova(m1,m2)
```

REPORTING GROWTH CURVE RESULTS

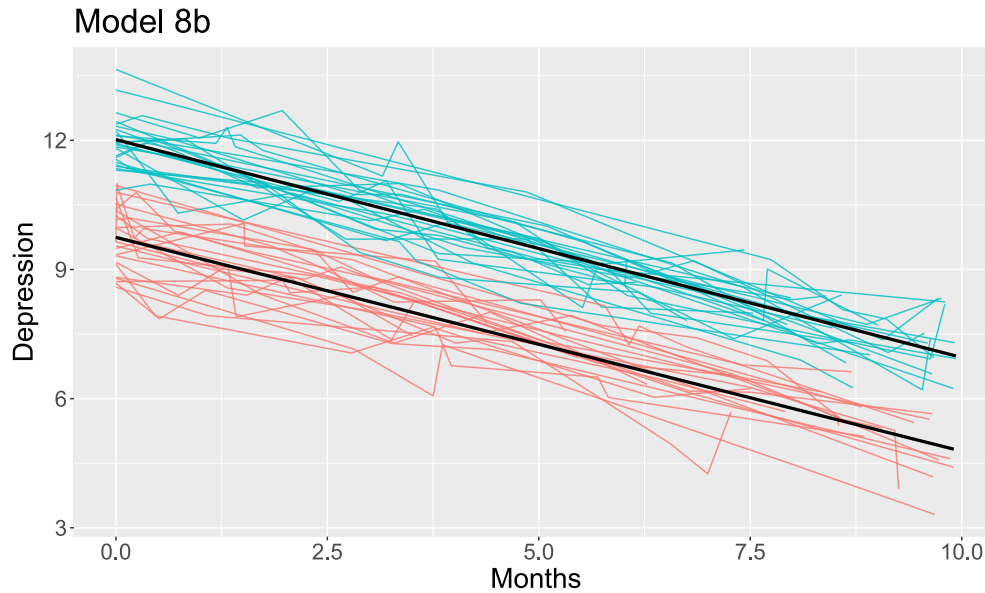
- Model selection:
 - Report the criteria used to select the best-fitting model, e.g., AIC, BIC, or likelihood ratio tests.
- Fixed effects:
 - Report the estimated coefficients, standard errors, t-values, and p-values for each fixed effect predictor.
 - Interpret the direction and magnitude of the relationships between the predictors and the outcome variable.
- Random effects:
 - Report the estimated variances and standard deviations for random intercepts and slopes.
 - Interpret the amount of variability in the intercepts and slopes across the different levels of the model.
- Model fit:
 - Report goodness-of-fit statistics, such as pseudo R-squared or deviance explained

MODEL 8



Why is there random intercept variance?

MODEL 8



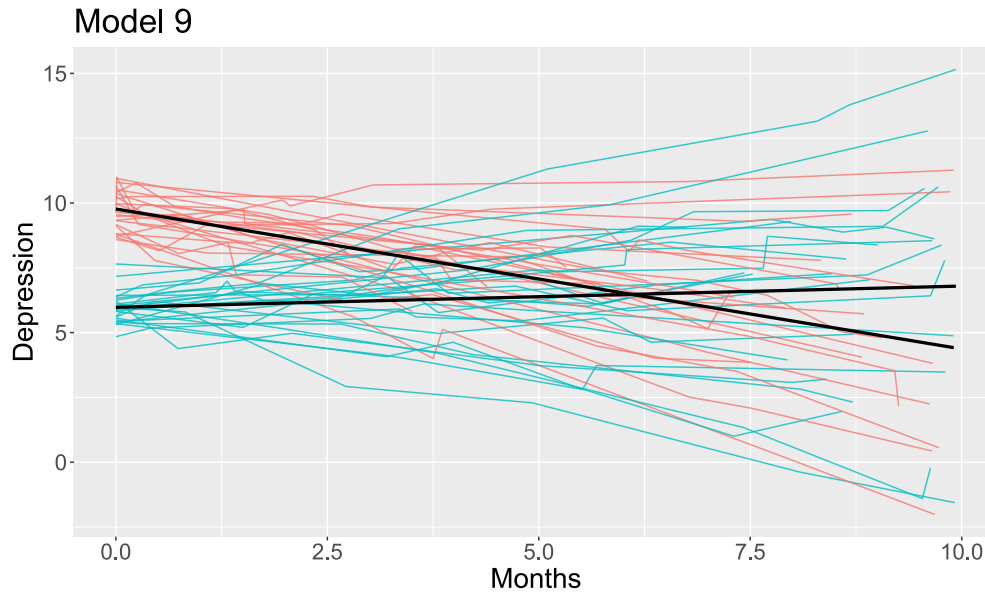
Level 2: $\beta_{0i} = \gamma_{00} + \gamma_{01} \cdot \text{Group}_i + U_{0i}$

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

Level 1: $y_{ti} = \beta_{0i} + \beta_{1i} \cdot \text{Time}_{ti} + \epsilon_{ti}$

Composite: $y_{ti} = (\gamma_{00} + \gamma_{01} \cdot \text{Group}_i + U_{0i}) + (\gamma_{10} + U_{1i}) \cdot \text{Time}_{ti} + \epsilon_{ti}$

MODEL 9



Level 2: $\beta_{0i} = \gamma_{00} + \gamma_{01} \cdot \text{Group}_i + U_{0i}$

Level 2: $\beta_{1i} = \gamma_{10} + \gamma_{11} \cdot \text{Group}_i + U_{1i}$

Level 1: $y_{ti} = \beta_{0i} + \beta_{1i} \cdot \text{Time}_{ti} + \epsilon_{ti}$

Composite: $y_{ti} = (\gamma_{00} + \gamma_{01} \cdot \text{Group}_i + U_{0i}) + (\gamma_{10} + \gamma_{11} \cdot \text{Group}_i + U_{1i}) \cdot \text{Time}_{ti} + \epsilon_{ti}$

- **Effect on t/p-values:**

- The impact on t/p-values of estimating random components depends on the level.
- Moving variance from Level 1 to Level 2 can provide more power to detect Level 1 predictors.

- **Improved Model Fit:**

- Including random effects can help account for unobserved heterogeneity in the data, leading to better model fit.

- **Generalizability:**

- Including random effects allows for more accurate generalizations to the larger population.

- **Cautions:**

- Adding random effects can increase model complexity.
- It is important to carefully consider the theoretical justifications for including random effects in the model.

- **ML (Maximum Likelihood)**

- Estimates fixed and random effects simultaneously
- Can result in biased estimates of random effect variances
- Suitable for model selection and comparison

- **REML (Restricted Maximum Likelihood)**

- Estimates fixed effects and random effects separately
- Provides unbiased estimates of random effect variances
- Less suitable for model selection and comparison

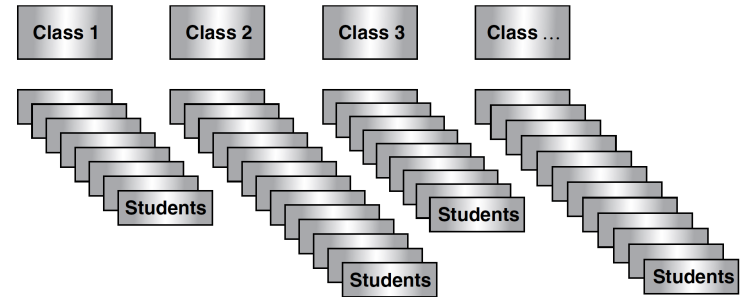
- **Which to use?**

- For unbiased variance estimates, use REML
- For model selection or comparison, use ML
- Sometimes, both are used in a two-step process: REML to estimate variance components and ML to compare models

$$\text{standardized coefficient} = \frac{\text{unstandardized coefficient} \times \text{stand. dev. explanatory var.}}{\text{stand. dev. outcome var.}}$$

- **Purpose:** To compare the relative importance of predictors and facilitate interpretation.
- **Procedure:**
 1. Standardize continuous predictor variables (center and scale)
 2. Refit the linear mixed model with standardized predictors
 3. Interpret the standardized coefficients as effect sizes
- **Interpretation:**
 - Each standardized coefficient represents the change in outcome variable (in standard deviations) for a one standard deviation change in the predictor
 - Larger absolute values indicate stronger relationships between predictors and the outcome variable
- **Considerations:**
 - Only applicable to continuous predictor variables

- There no MLM statistic entirely equivalent to R^2 in ordinary regression.
- While R^2 increases monothically as independent variables are added in linear regression, this may not hold in multilevel models.
- in MLM we can calculate an of explained variance at each level.
 - This is referred to as *pseudo* R^2 , and can behave in unusual ways.
 - Can decrease as predictor on another level is included.



EXAMPLE: MODELLING POPULARITY IN PUPILS

Model	Intercept only	Level 1 predictors	Level 2 predictors	Random coefficient	Cross-level interaction
Fixed part					
Intercept	5.08	2.14	0.81	0.74	-1.21
Extraversion		0.44	0.45	0.45	0.80
Gender		1.25	1.25	1.25	1.24
T. exp			0.09	0.09	0.23
Extra × T.exp					-0.02
Random part					
σ_e^2	1.22	0.59	0.59	0.55	0.55
σ_{u0}^2	0.69	0.62	0.29	1.28	0.45
σ_{u2}^2				0.03	0.004
$\sigma_{u_{02}}$				-0.18	-0.03
Deviance	6327.5	4934.0	4862.3	4812.8	4747.6

$$R_1^2 = \left(\frac{\sigma_{elb}^2 - \sigma_{elm}^2}{\sigma_{elb}^2} \right), \quad R_1^2 = \left(\frac{1.22 - 0.59}{1.22} \right) = .52.$$

$$R_2^2 = \left(\frac{\sigma_{u0|b}^2 - \sigma_{u0|m}^2}{\sigma_{u0|b}^2} \right), \quad R_2^2 = \left(\frac{0.69 - 0.62}{0.69} \right) = .10.$$

$$R_{\beta_2}^2 = \left(\frac{\sigma_{u2|b}^2 - \sigma_{u2|m}^2}{\sigma_{u2|b}^2} \right), \quad R_{extrav}^2 = \left(\frac{0.034 - 0.0047}{0.034} \right) = .86.$$

SECTION: A CLOSER LOOK AT RANDOM EFFECTS

KEEPING IT MAXIMAL

- A full or **maximal random effect structure** is the case where all of the factors that could hypothetically vary across individual observational units are allowed to do so.
- The general principle is keep it maximal (Barr et al., 2013): the random effects should include as much of the structure of the data as possible.

```
m_1 ← lmer(Accuracy ~ (ot1+ot2)*TP + (ot1 | Subject), data=WordLearnEx, REML=FALSE)  
m_2 ← lmer(Accuracy ~ (ot1+ot2)*TP + (ot1+ot2 | Subject), data=WordLearnEx, REML=FALSE)
```

Removing a time term from the random effects primarily reduces the standard error of the corresponding fixed effect estimate, making more significant.

EXAMPLE: COGNITIVE PERFORMANCE AND AGE

- **Research question:** Does cognitive performance decline with age and does this decline rate vary between individuals?
- **Data structure:**
 - Repeated measures of cognitive performance for each individual
 - Age as a time-varying predictor
- **Linear mixed model:**
 - Random intercepts and slopes for age
 - Correlation between intercepts and slopes

Reason to Drop the Intercept-Slope Correlation

- **Hypothesis:** Initial cognitive performance and rate of decline are unrelated.
 - Example: Individuals with high initial performance decline at the same rate as those with low initial performance.
- **Model modification:** Set the correlation between intercepts and slopes to zero.

HOW TO SPECIFY NON-CORRELATED INTERCEPT AND SLOPE

```
m_corr ← lmer(y ~ time + (1 | id), data=dt)
m_noncorr ← lmer(y ~ time + (1 | id) + (0 + time | id), data=dt)
```

Reason to Drop the Intercept-Slope Correlation

- **Hypothesis:** Initial cognitive performance and rate of decline are unrelated.
 - Example: Individuals with high initial performance decline at the same rate as those with low initial performance.
- **Model modification:** Set the correlation between intercepts and slopes to zero.
 - This enforces the hypothesis that initial performance and decline rate are independent.
- **Interpretation:** If the modified model fits the data well, it suggests that the rate of cognitive decline is not dependent on initial performance levels.

EFFECT OF OMITTING RANDOM COEFFICIENTS

	Estimate	Std. Error	<i>t</i>	<i>p</i>
Intercept	0.779	0.022	35.831	0.000
Linear	0.286	0.038	7.577	0.000
Quadratic	-0.051	0.033	-1.532	0.125
High TP: Intercept	0.053	0.031	1.724	0.085
High TP: Linear	0.001	0.053	0.020	0.984
High TP: Quadratic	-0.116	0.047	-2.481	0.013

	Estimate	Std. Error	<i>t</i>	<i>p</i>
Intercept	0.779	0.022	35.847	0.000
Linear	0.286	0.037	7.789	0.000
Quadratic	-0.051	0.030	-1.681	0.093
High TP: Intercept	0.053	0.031	1.724	0.085
High TP: Linear	0.001	0.052	0.021	0.983
High TP: Quadratic	-0.116	0.043	-2.722	0.006

	Estimate	Std. Error	<i>t</i>	<i>p</i>
Intercept	0.779	0.022	35.847	0.000
Linear	0.286	0.037	7.789	0.000
Quadratic	-0.051	0.030	-1.678	0.093
High TP: Intercept	0.053	0.031	1.724	0.085
High TP: Linear	0.001	0.052	0.021	0.983
High TP: Quadratic	-0.116	0.043	-2.717	0.007

```
m_2<-lmer(Accuracy ~ (ot1+ot2)*TP + (ot1+ot2 | Subject))
```

$$G = \begin{bmatrix} \text{Var}(U_0) & \text{Cov}(U_0, U_1) & \text{Cov}(U_0, U_2) \\ \text{Cov}(U_0, U_1) & \text{Var}(U_1) & \text{Cov}(U_1, U_2) \\ \text{Cov}(U_0, U_2) & \text{Cov}(U_1, U_2) & \text{Var}(U_2) \end{bmatrix}$$

```
m_1<-lmer(Accuracy ~ (ot1+ot2)*TP + (ot1 | Subject))
```

$$G = \begin{bmatrix} \text{Var}(U_0) & \text{Cov}(U_0, U_1) & 0 \\ \text{Cov}(U_0, U_1) & \text{Var}(U_1) & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

```
m_3<-lmer(Accuracy ~ (ot1+ot2)*TP +
           (1 | Subject) +
           (0+ot1 | Subject) +
           (0+ot2 | Subject))
```

$$G = \begin{bmatrix} \text{Var}(U_0) & 0 & 0 \\ 0 & \text{Var}(U_1) & 0 \\ 0 & 0 & \text{Var}(U_2) \end{bmatrix}$$

Depending upon the variation among clusters, which is learned from the data as well, the model pools information across clusters. This pooling tends to improve estimates about each cluster.

Complete pooling: First, suppose you ignore the varying intercepts and just use the overall mean across all clusters.

No pooling Estimate separate fixed effect in each cluster.

Partial pooling Multilevel approach. Extreme values are pulled towards an overall average, and this is stronger for smaller groups.

1. COMPLETE POOLING

This approach assumes the treatment has the same effect on everyone, ignoring individual-specific variation. It estimates an average treatment effect across all individuals.

Consequences:

- *Biased estimates*: Complete pooling may yield biased treatment effect estimates due to ignored individual variation.
- *Inaccurate inferences*: Hypothesis tests and confidence intervals may be misleading, leading to incorrect conclusions.
- *Lack of personalization*: No information about the treatment's effectiveness for specific individuals, hindering tailored interventions.

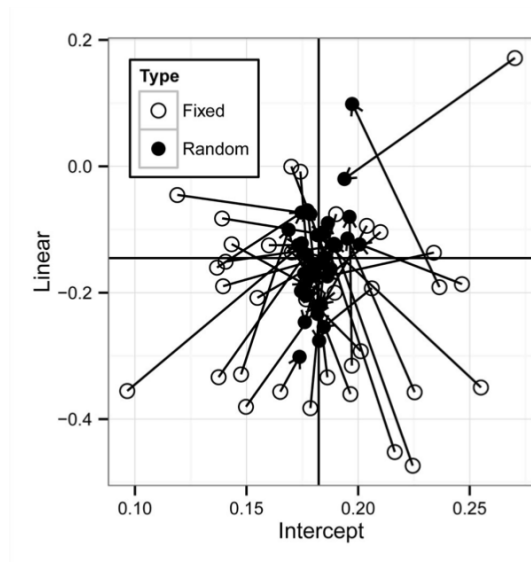
2. PARTIAL POOLING

This approach estimates separate effects for each individual while "borrowing strength" from the group, typically using random effects or mixed models. It allows for individual-specific treatment effects and considers the overall population effect.

Consequences:

- *More accurate estimates:* Partial pooling accounts for individual variation, providing more accurate treatment effect estimates.
- *Improved inferences:* Hypothesis tests and confidence intervals are more accurate, leading to reliable conclusions.
- *Personalized insights:* Provides information on the treatment's effectiveness for specific individuals, aiding personalized interventions.

SHRINKAGE IN PARTIAL POOLING



- Pooling data across clusters tends to *shrink* their deviation from the overall mean levels.
- **Shrinkage** effect on individual participant intercept and linear term parameter estimates. For each participant, the arrow shows the change in the parameter estimate from a model that treats participants as fixed effects (open circles) to a model that treats participants as random effects (filled circles). The black vertical and horizontal lines indicate the population-level fixed effect.

3. No POOLING

This approach estimates separate effects for each individual without sharing information between them, fitting separate models and ignoring potential similarities.

Consequences:

- *Overfitting*: No pooling may overfit the data, resulting in estimates that don't generalize well.
- *Inefficient data use*: Not sharing information between individuals leads to less efficient use of data and less precise estimates.
- *Interpretation challenges*: Separate models for each individual make it hard to draw overall conclusions about treatment effectiveness.

CONSTRUCTING A LONGITUDINAL MODEL (HOFFMAN)

1. Building an unconditional model of change
 - Decide what your metric of time will be
 - Decide at what occasion time 0 should be located
2. Plot individual trajectories over time
 - What kind of change, if any, do you see on average?
 - Do you see individual differences in that pattern of change?
3. Fit a null model (random intercept)
 - Calculate the ICC. If it is close to 1.0, there is no longitudinal variation to model.
4. Add level-1 Fixed predictors
5. Add level-2 Explanatory Variables
6. Examine Whether a Particular slope varies between groups
7. Add cross-level interactions to explain variation in the slope.

class: mainTopic, middle, center

SECTION: CATEGORICAL TIME INVARIANT PREDICTORS

TIME VARYING AND TIME INVARIANT PREDICTORS

- A predictor is **time varying** when it is measured at multiple points in time, just as is the outcome variable.
 - In the context of education, a time-varying predictor might be the number of hours in the previous 30 days a student has spent studying.
 - time-varying predictors will appear at level 1 because they are associated with specific measurements
- On the other hand, a predictor is **time invariant** when it is measured at only one point in time, and its value does not change across measurement occasions.
 - An example of this type of predictor would be gender,
 - whereas time-invariant predictors will appear at level 2 or higher, because they are associated with the individual
- Unless time-varying predictors are group-mean centered, BP and WP variance may not be cleanly separated.

- **Time-varying predictors**

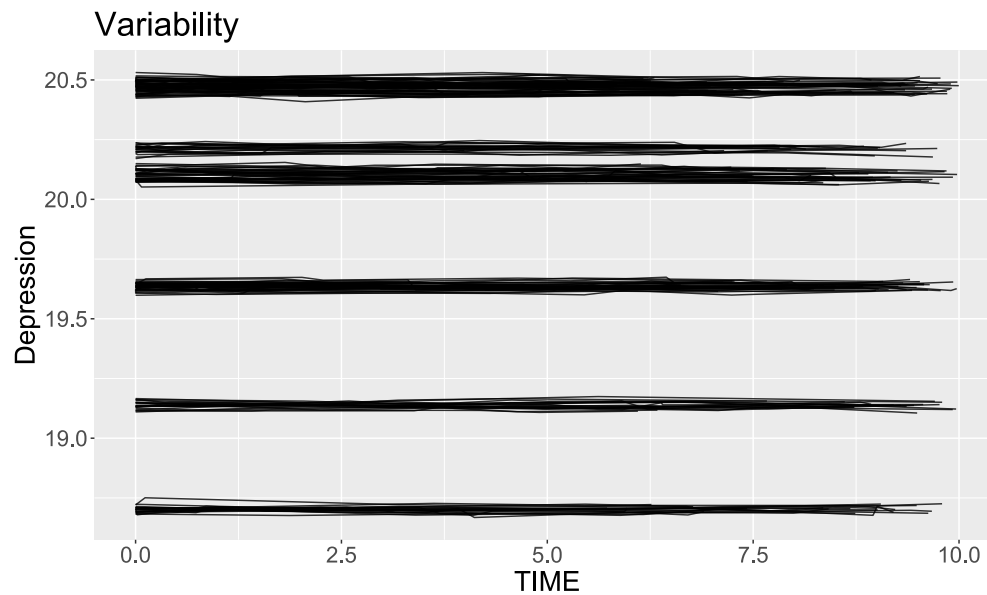
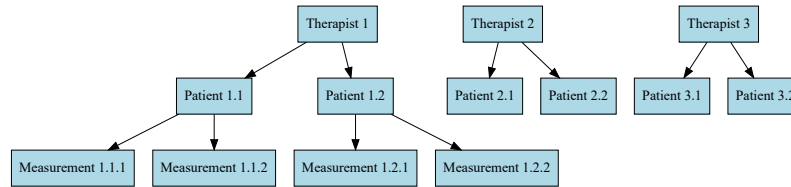
- A predictor is **time varying** when it is measured at multiple points in time, just as is the outcome variable.
- Change over time within individuals
- Predict within-person fluctuations
- Can be used to model individual trajectories
- Need to consider potential lagged effects, autocorrelation, and time-dependent confounders

- **Time-invariant predictors**

- A predictor is **time invariant** when it is measured at only one point in time, and its value does not change across measurement occasions.
- Do not change over time within individuals
- Predict between-person differences
- Can be used to model group differences in trajectories
- Need to consider potential multicollinearity with other time-invariant predictors

SECTION: HIGHER ORDER LONGITUDINAL MODELS

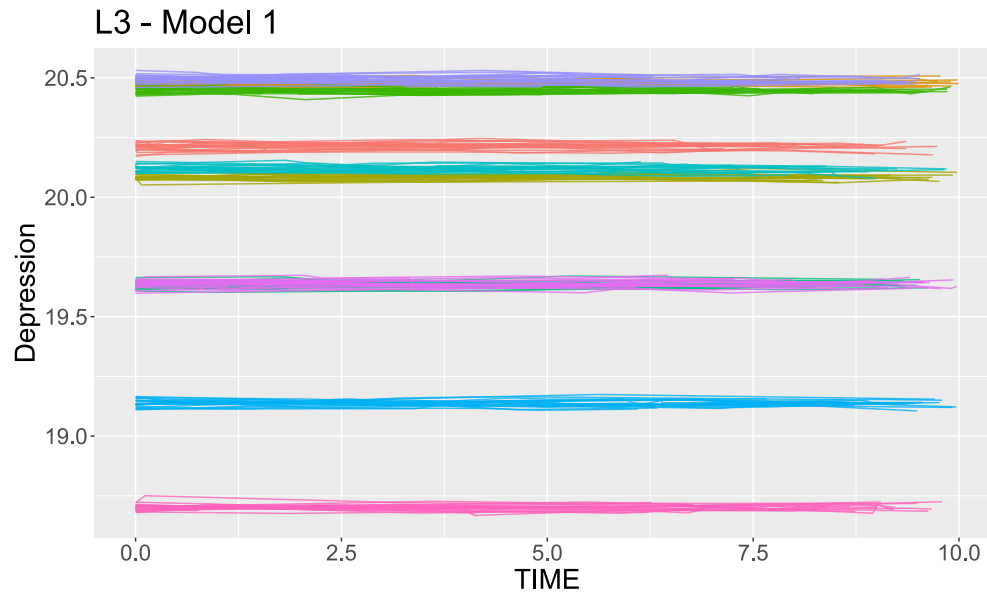
INTRODUCING HIGHER LEVEL VARIABILITY



THIRD ORDER LINEAR MIXED MODEL: DEPRESSION OVER TIME

- **Outcome:** Depression score over time
- **Data structure:**
 - Repeated measures of depression for each patient
 - Patients nested within therapists
- **Model components:**
 1. **Fixed effects:** Population-level effects of predictors (e.g., time, patient characteristics, therapist characteristics)
 2. **Random effects:** Individual-level variability in intercepts (baseline depression) and slopes (depression change over time)
 3. **Third-level random effects:** Therapist-level variability in intercepts and slopes
- **Model interpretation:**
 - Fixed effects describe the average relationships between predictors and depression scores
 - Random effects capture the variability in depression scores and change over time across patients

L3 - MODEL 1



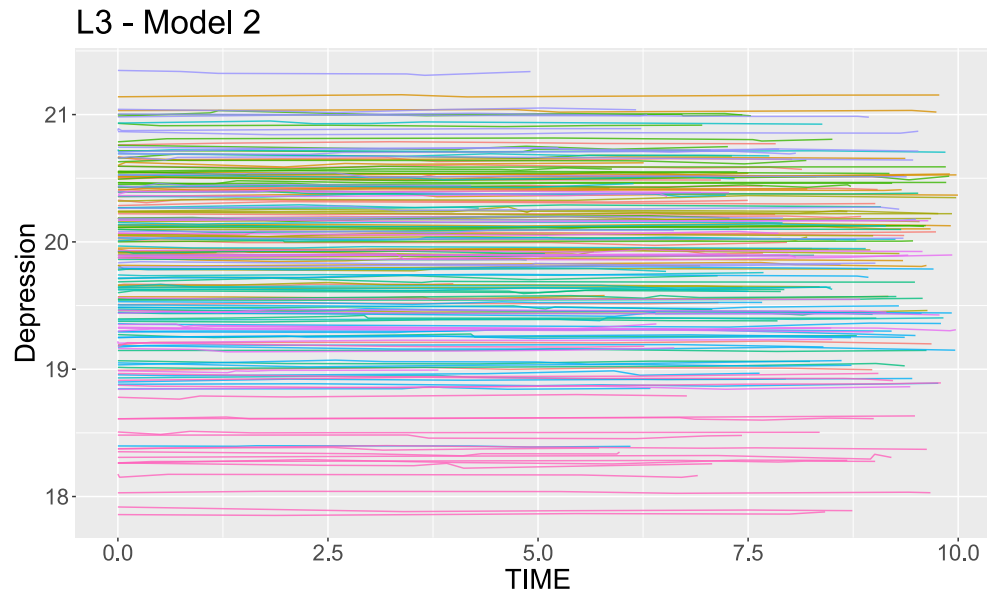
Level 3: $\gamma_{00b} = \delta_{000} + V_{00b}$

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

Level 1: $y_{tib} = \beta_{0ib} + \epsilon_{tib}$

Composite: $y_{tib} = \delta_{000} + V_{00b} + U_{0ib} + \epsilon_{tib}$

L3 - MODEL 2

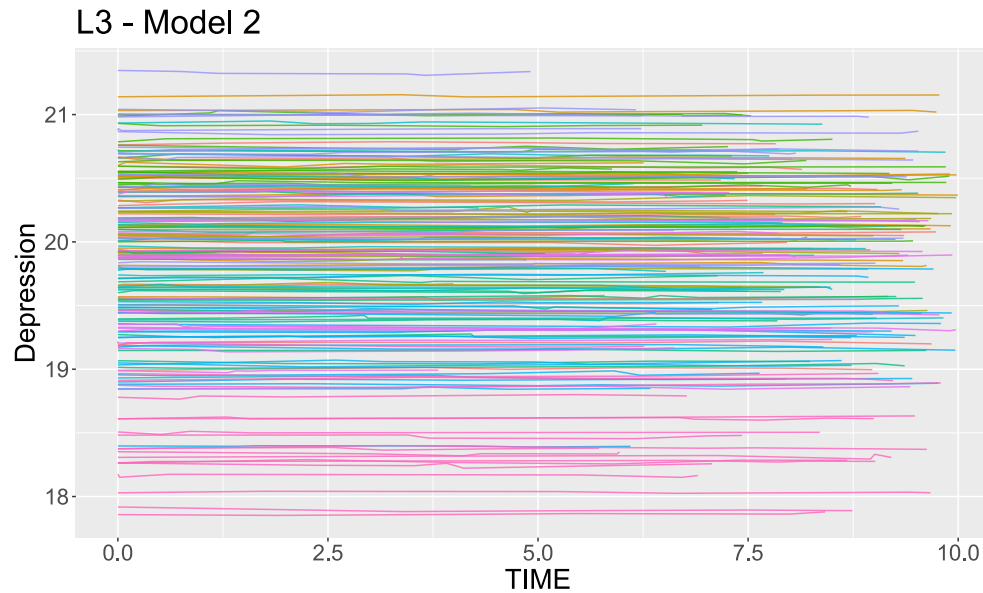


Level 3: $\gamma_{00b} = \delta_{000} + V_{00b}$

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

Level 1: $y_{tib} = \beta_{0ib} + \epsilon_{tib}$

Composite: $y_{tib} = \delta_{000} + V_{00b} + U_{0ib} + \epsilon_{tib}$



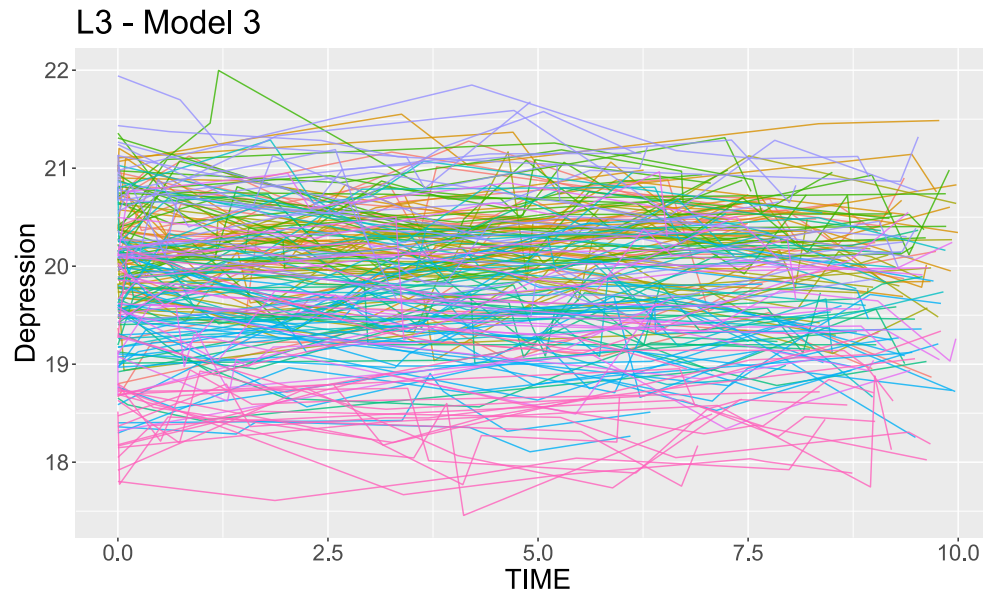
Level 3: $\gamma_{00b} = \delta_{000} + V_{00b}$

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

Level 1: $y_{tib} = \beta_{0ib} + \epsilon_{tib}$

Composite: $y_{tib} = \delta_{000} + V_{00b} + U_{0ib} + \epsilon_{tib}$

L3 - MODEL 3

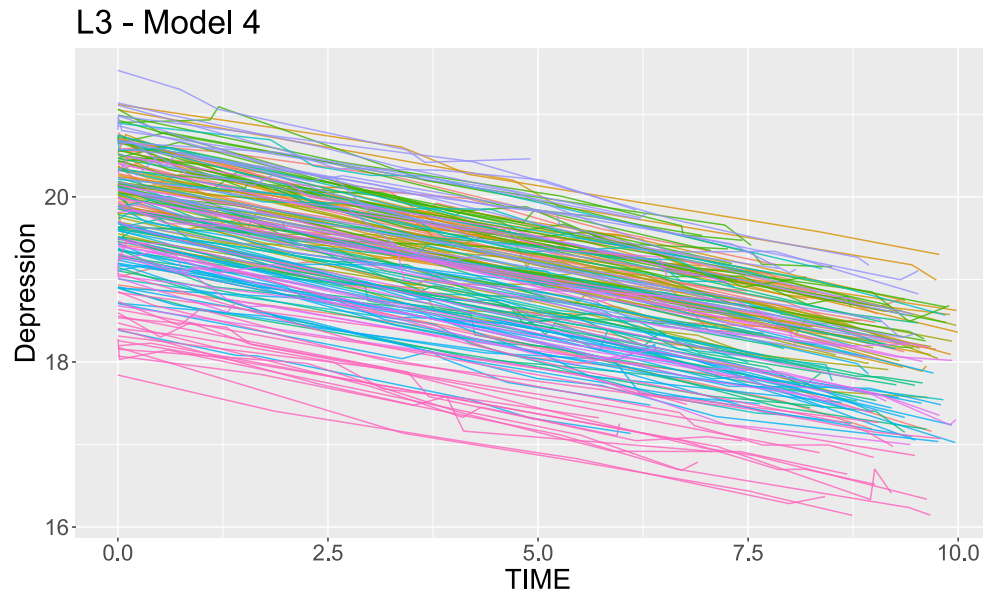


Level 3: $\gamma_{00b} = \delta_{000} + V_{00b}$

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

Level 1: $y_{tib} = \beta_{0ib} + \epsilon_{tib}$

Composite: $y_{tib} = \delta_{000} + V_{00b} + U_{0ib} + \epsilon_{tib}$



Level 3: $\gamma_{00b} = \delta_{000} + V_{00b}$

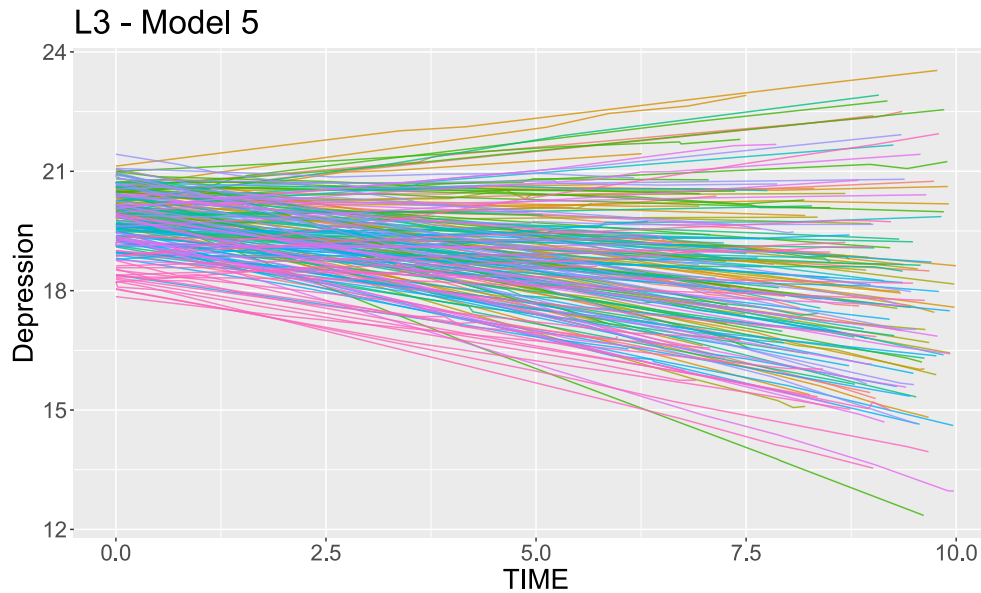
Level 2: $\beta_{0ib} = \gamma_{00b} + U_{0ib}$

Level 2: $\beta_{1ib} = \gamma_{100}$

Level 1: $y_{tib} = \beta_{0ib} + \beta_{1ib} \cdot \text{Time} + \epsilon_{tib}$

Composite: $y_{tib} = \delta_{000} + V_{00b} + U_{0ib} + \epsilon_{tib}$

L3 - MODEL 5



Level 3: $\gamma_{00b} = \delta_{000} + V_{00b}$

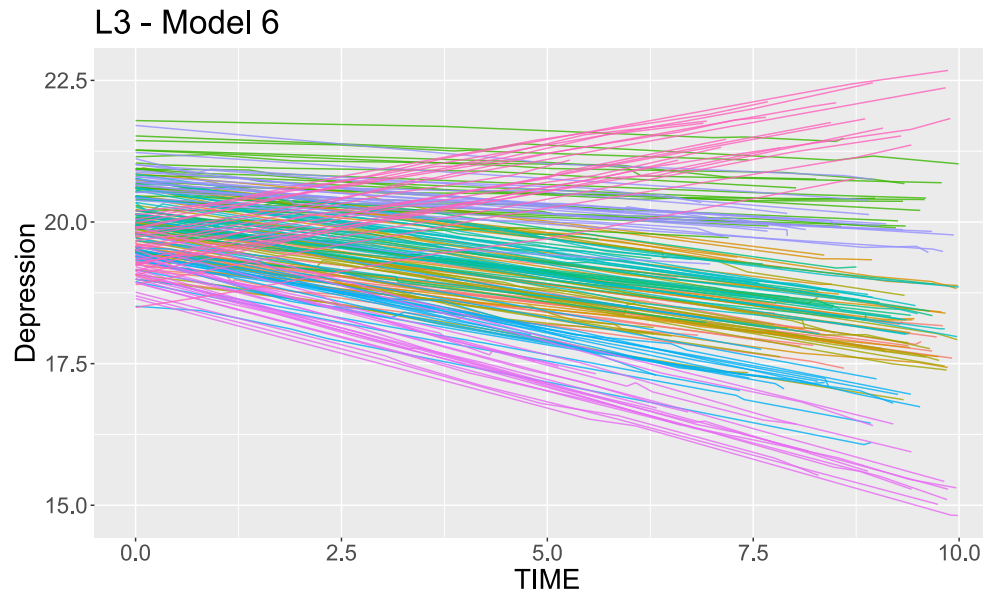
Level 2: $\beta_{0ib} = \gamma_{00b} + U_{0ib}$

Level 2: $\beta_{1ib} = \gamma_{100} + U_{1ib}$

Level 1: $y_{tib} = \beta_{0ib} + \beta_{1ib} \cdot \text{Time} + \epsilon_{tib}$

Composite: $y_{tib} = \delta_{000} + V_{00b} + U_{0ib} + \epsilon_{tib}$

L3 - MODEL 6



Level 3: $\gamma_{00b} = \delta_{000} + V_{00b}$

Level 3: $\gamma_{10b} = \delta_{100} + V_{00b}$

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

Level 2: $\beta_{1ib} = \gamma_{100}$

Level 1: $y_{tib} = \beta_{0ib} + \beta_{1ib} \cdot \text{Time} + \epsilon_{tib}$

Composite: $y_{tib} = \delta_{000} + V_{00b} + U_{0ib} + \epsilon_{tib}$

Table 5.3 Results of multilevel analysis of GPA, fixed effects

Model	M1: null model	M2: + occas.	M3: + job stat	M4: + high school GPA, gender
Fixed part	Coeff. (s.e.)	Coeff. (s.e.)	Coeff. (s.e.)	Coeff. (s.e.)
Intercept				
Occasion				
Job status				
GPA highschl				
Gender				
Random part				
σ_e^2				
σ_{u0}^2				
Deviance				
AIC				
BIC				

GPA EXERCISE (2)

Model	M5: + occasion random	M6: + cross-level interaction	M6: standardized
Fixed part	Coeff. (s.e.)	Coeff. (s.e.)	
Intercept			
Occasion			0.38
Job status			-0.14
GPA highschl			
Gender			0.10
Occas*Gender			0.13
Random part			
σ_e^2			
σ_{u0}^2			
σ_{u1}^2			
σ_{u01}			
r_{u01}			
Deviance			
AIC			
BIC			

LEVEL-3 EXERCISE (2)

Table 11.2 Results from longitudinal models for persons nested in time-invariant groups.
Bold values are $p < .05$.

Parameters	Unconditional Time Model (11.4)			Final Conditional Model (11.9)		
	Est	SE	$p <$	Est	SE	$p <$
Model for the Means						
γ_{000} Intercept				4.211	0.058	.001
γ_{100} Linear Time Slope (0 = Wave 1)				0.131	0.030	.001
γ_{010} Student Gender (0 = Boy, 1 = Girl)				0.252	0.068	.001
γ_{110} Student Gender by Time				-0.001	0.039	.987
γ_{003} Class Mean Gender (0 = .50 Girls)				0.545	0.444	.224
γ_{103} Class Mean Gender by Time				-1.112	0.235	.001
γ_{004} Class Mean Emotional Support (0 = 5)				0.108	0.074	.155
γ_{104} Class Mean Emotional Support by Time				0.170	0.038	.001
γ_{009} Class Mean Emotional Support by Class Mean Gender				-2.349	0.808	.006
γ_{400} Level-1 Victimization (0 = Student Mean)				-0.075	0.045	.092
γ_{020} Level-2 Victimization (0 = Class Mean)				-0.053	0.052	.308
γ_{420} Level-2 Victimization by Level-1 Victimization				0.109	0.045	.016
γ_{040} Level-2 Victimization by Student Gender				-0.152	0.073	.038
γ_{021} Level-2 Victimization by Class Mean Gender				-0.096	0.371	.797
γ_{005} Level-3 Victimization (0 = 3)				-0.054	0.160	.738
γ_{405} Level-3 Victimization by Level-1 Victimization				0.157	0.132	.235
$\gamma_{00.11}$ Level-3 Victimization by Class Mean Emotional Support				-0.778	0.278	.009
Model for the Variance						
$\tau_{V_{00}}^2$ Class Random Intercept Variance				0.019	0.015	.105
$\tau_{V_{10}}^2$ Class Random Linear Time Slope Variance				0.005	0.005	.153
$\tau_{V_{00,10}}$ Class Intercept-Time Slope Covariance				0.020	0.006	.001
$\tau_{U_{00}}^2$ Student Random Intercept Variance				0.395	0.043	.001
$\tau_{U_{10}}^2$ Student Random Linear Time Slope Variance				0.044	0.016	.003
$\tau_{U_{40}}^2$ Student Random L1 Victimization Slope Variance				0.095	0.032	.001
$\tau_{U_{00,10}}$ Student Intercept-Time Slope Covariance				-0.007	0.020	.710
$\tau_{U_{00,40}}$ Student Intercept-L1 Victimization Slope Covariance				0.017	0.027	.525
$\tau_{U_{10,40}}$ Student Time-L1 Victimization Slope Covariance				-0.013	0.013	.339
σ_e^2 Residual Variance				0.297	0.020	.001

SECTION: TRANSFORMING DATA FROM WIDE TO LONG FORMAT

WIDE VS. LONG FORMAT

	student	sex	highgpa	gpa1	gpa2	gpa3	gpa4	gpa5	gpa6	job1	job2
1	1	1	2.8	2.3	2.1	3.0	3.0	3.0	3.3	2	2
2	2	0	2.5	2.2	2.5	2.6	2.6	3.0	2.8	2	3
3	3	1	2.5	2.4	2.6	3.0	2.8	3.3	3.4	2	2
4	4	0	3.8	2.5	2.7	2.4	2.7	2.9	2.7	3	2
5	5	0	3.1	2.8	2.8	2.8	3.0	2.9	3.1	2	2
6	6	1	2.9	2.5	2.4	2.4	2.3	2.7	2.8	2	3
7	7	0	2.3	2.4	2.4	2.8	2.6	3.0	3.0	3	2
8	8	1	3.0	2.8	2.8	3.1	3.3	3.3	3.4	2	2
9	9	0	2.0	2.8	2.7	2.7	3.1	3.1	3.5	2	2
10	10	0	2.8	2.8	2.8	3.0	2.7	3.0	3.0	2	2
11	11	1	3.9	2.6	2.9	3.2	3.6	3.6	3.8	2	3
12	12	1	2.9	2.6	3.0	2.3	2.9	3.1	3.3	3	2
13	13	0	3.7	2.6	3.1	3.5	3.6	3.9	3.9	2	2

	student	occas	gpa	job	sex	highgpa
1	1	0	2.3	2	1	2.8
2	1	1	2.1	2	1	2.8
3	1	2	3.0	2	1	2.8
4	1	3	3.0	2	1	2.8
5	1	4	3.0	2	1	2.8
6	1	5	3.3	2	1	2.8
7	2	0	2.2	2	0	2.5
8	2	1	2.5	3	0	2.5
9	2	2	2.6	2	0	2.5
10	2	3	2.6	2	0	2.5
11	2	4	3.0	2	0	2.5
12	2	5	2.8	2	0	2.5
13	3	0	2.4	2	1	2.5
14	3	1	2.9	2	1	2.5
15	3	2	3.0	2	1	2.5
16	3	3	2.8	3	1	2.5
17	3	4	3.3	2	1	2.5
18	3	5	3.4	2	1	2.5
19	4	0	2.5	3	0	3.8
20	4	1	2.7	2	0	3.8
21	4	2	2.4	2	0	3.8
22	4	3	2.7	2	0	3.8
23	4	4	2.9	2	0	3.8
24	4	5	2.7	2	0	3.8
25	5	0	2.8	2	0	3.1
26	5	1	2.8	2	0	3.1

Left: Wide format; each measure of the same variable is in separate column.

Right: Long format; each measure of the same variable is in separate row.

- The *tidyverse* is a collection of open source R packages that share an underlying design philosophy, grammar, and data structures.
- In the tidyverse approach to R syntax, these two functions are used to transform a dataset from wide to long format.
 - `pivot_longer()` makes datasets longer by increasing the number of rows and decreasing the number of columns.
 - `pivot_wider()` is the opposite of `pivot_longer()`: it makes a dataset wider by increasing the number of columns and decreasing the number of rows.
- See these pages for documentation and examples:
 - <https://cran.r-project.org/web/packages/tidyr/vignettes/pivot.html>

CREATING A SYNTHETIC WIDE DATASET

```
library(tidyverse)
dt<-tibble(
  id=1:20, male=sample(0:1, 20, replace=TRUE),
  dep1=rnorm(20), dep2=rnorm(20), dep3=rnorm(20),
  anx1=rnorm(20), anx2=rnorm(20), anx3=rnorm(20)
)

print(dt,n=10)
```

```
## # A tibble: 20 x 8
##       id  male  dep1    dep2    dep3    anx1    anx2    anx3
##   <int> <int> <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>
## 1     1     1 -0.709  0.0301  0.189  0.357  0.407  0.199
## 2     2     0 -1.10  -0.696 -0.0357  1.37  -0.0255  0.880
## 3     3     0 -1.06  -0.104 -0.315  0.226  0.585 -1.17
## 4     4     1  0.490  1.42   0.0567 -0.182 -0.347  1.17
## 5     5     1  0.149 -1.25   0.284 -0.354  0.0755  0.948
## 6     6     0  1.39  -0.828  0.392 -0.197  1.55   0.847
## 7     7     0 -1.41  -1.94  -2.02  -0.0964 -0.182  1.58
## 8     8     0 -0.592 -0.356 -1.30   0.434 -0.600  0.352
## 9     9     0  1.03   0.359  0.0586  0.980  0.921  0.793
## 10    10     1 -1.02   1.81   0.776  0.252  0.556 -0.589
## # i 10 more rows
```

PIVOT FROM WIDE TO LONG

```
dt_l ←
  dt %>% pivot_longer(
    cols=dep1:anx3,
    names_sep=3,
    names_to = c("disorder", "number")
  ) %>% pivot_wider(names_from="disorder") %>%
  rename("wave"="number",
         "anxiety"="anx",
         "depression"="dep")

print(dt_l,n=10)
```

```
## # A tibble: 60 x 5
##       id  male wave  depression anxiety
##   <int> <int> <chr>      <dbl>    <dbl>
## 1     1     1  1 1        -0.709    0.357
## 2     1     1  1 2         0.0301   0.407
## 3     1     1  1 3         0.189    0.199
## 4     2     0  1 1        -1.10     1.37
## 5     2     0  2 2        -0.696   -0.0255
## 6     2     0  3 3        -0.0357   0.880
## 7     3     0  1 1        -1.06     0.226
## 8     3     0  2 2        -0.104    0.585
## 9     3     0  3 3        -0.315   -1.17
## 10    4     1  1 1         0.490   -0.182
## # i 50 more rows
```