Longitudinal Data Analysis I

PSYC 575

October 3, 2020 (updated: 15 October 2022)

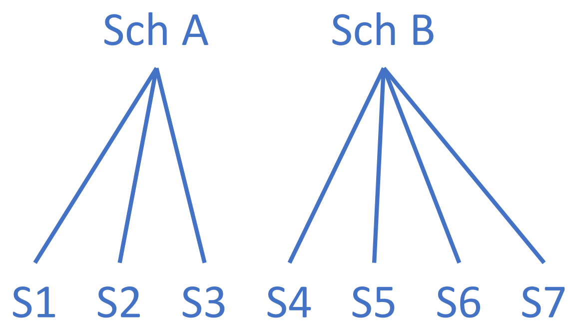
Learning Objectives

- Describe the similarities and differences between **longitudinal data** and cross-sectional clustered data
- Perform some basic attrition analyses
- Specify and run **growth curve analysis**
- Analyze models with **time-invariant covariates** (i.e., lv-2 predictors) and interpret the results

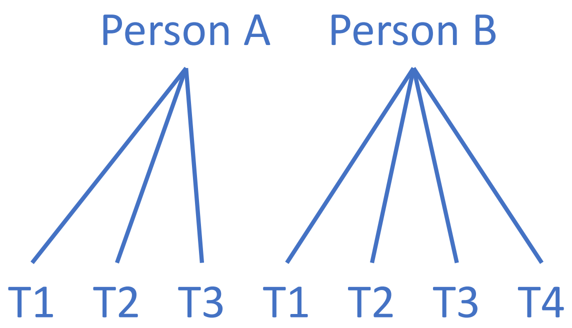
Longitudinal Data and Models

Data Structure

- Students in Schools



- Repeated measures within individuals



Types of Longitudinal Data

- Panel data
  - Everyone measured at the same time (e.g., every two years)
- Intensive longitudinal data
  - Each person measured at many time points
  - E.g., daily diary, ecological momentary assessment (EMA)

Two Different Goals of Longitudinal Models

- Trend
  - Growth modeling
  - Stable pattern
  - E.g., trajectory of cognitive functioning over five years

- Fluctuations
  - Clear trend not expected
  - E.g., fluctuation of mood in a day

Example

Children’s Development in Reading Skill and Antisocial Behavior

- 405 children within first two years entering elementary school
- 2-year intervals between 1986 and 1992
- Age = 6 to 8 years at baseline

Same Multilevel Structure

- At first, it may not be obvious looking at the data (in wide format)

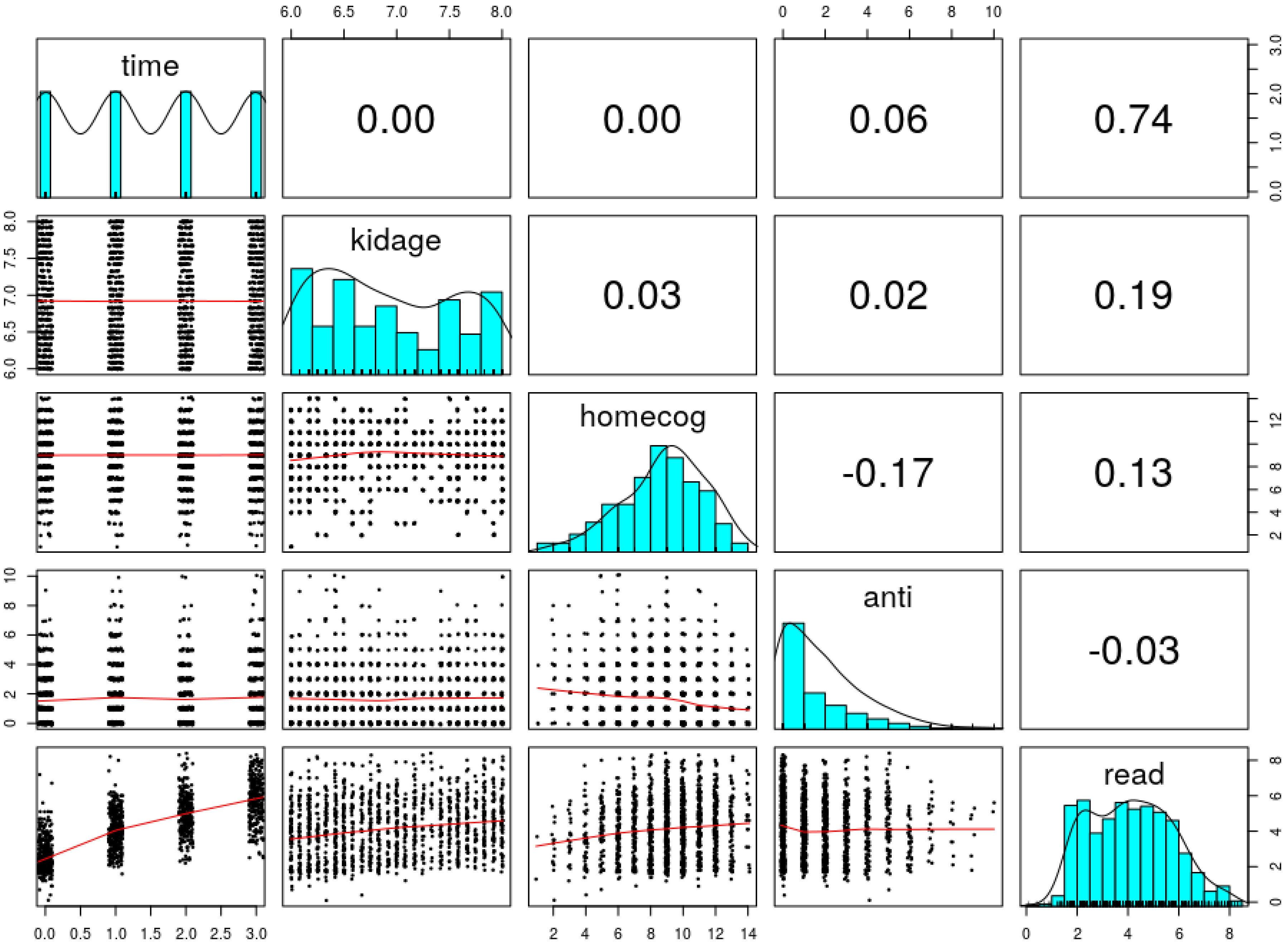
id <dbl>	anti1 <dbl>	anti2 <dbl>	anti3 <dbl>	anti4 <dbl>	read1 <dbl>	read2 <dbl>	read3 <dbl>	read4 <dbl>
22	1	2	NA	NA	2.1	3.9	NA	NA
34	3	6	4	5	2.1	2.9	4.5	4.5
58	0	2	0	1	2.3	4.5	4.2	4.6
122	0	3	1	1	3.7	8.0	NA	NA
125	1	1	2	1	2.3	3.8	4.3	6.2
133	3	4	3	5	1.8	2.6	4.1	4.0
163	5	4	5	5	3.5	4.8	5.8	7.5
190	0	NA	NA	0	2.9	6.1	NA	NA
227	0	0	2	1	1.8	3.8	4.0	NA
248	1	2	2	0	3.5	5.7	7.0	6.9
	T1	T2	T3	T4	T1	T2	T3	T4

Restructuring!

- Long format

“Cluster” 22

id	anti	read	time
<dbl>	<dbl>	<dbl>	<dbl>
22	1	2.1	1
22	2	3.9	2
22	NA	NA	3
22	NA	NA	4
34	3	2.1	1
34	6	2.9	2
34	4	4.5	3
34	5	4.5	4
58	0	2.3	1
58	2	4.5	2
58	0	4.2	3
58	1	4.6	4
122	0	3.7	1
122	3	8.0	2
122	1	NA	3
122	1	NA	4
125	1	2.3	1
125	1	3.8	2
125	2	4.3	3
125	1	6.2	4
133	3	1.8	1
133	4	2.6	2
133	3	4.1	3
133	5	4.0	4
163	5	3.5	1
163	4	4.8	2
163	5	5.8	3
163	5	7.5	4
190	0	2.9	1
190	NA	6.1	2



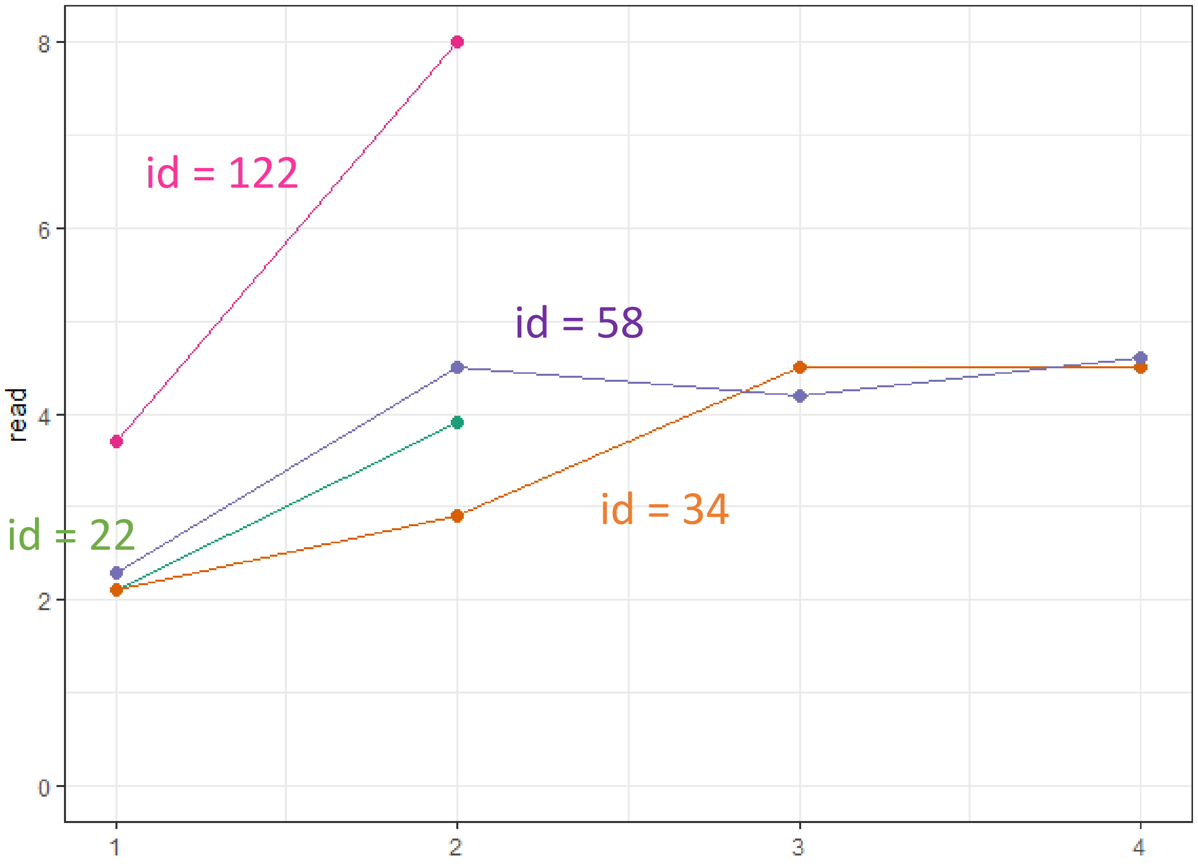
time
kidage
homecog
anti
read
0.00
0.00
0.06
0.74
0.03
0.02
0.19
-0.17
0.13
-0.03

Attrition Analysis

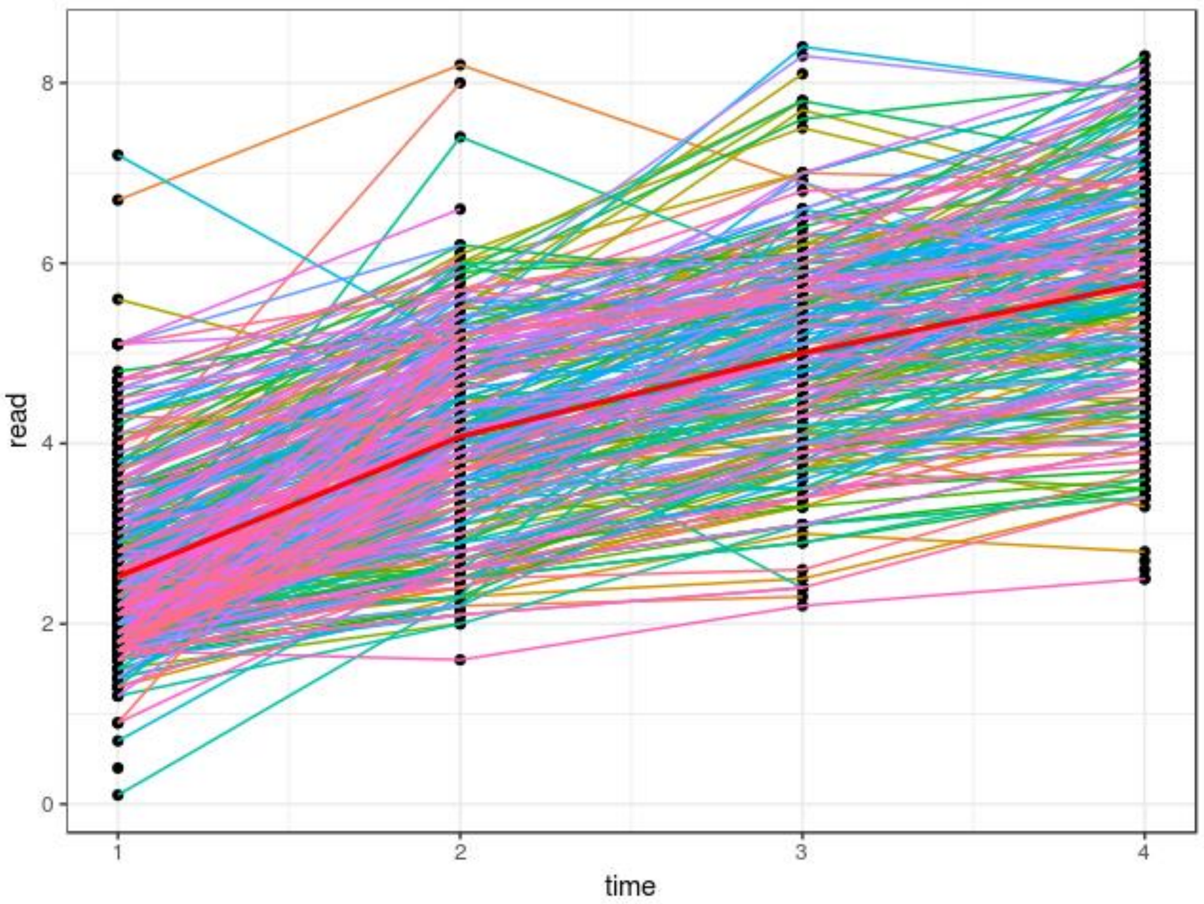
- Whether those who dropped out differ in important characteristics from those who stayed
- Design: Collect information on predictors of attrition, and perceived likelihood of dropping out
- Limited generalizability
- Missing data handling techniques
  - E.g., Multiple imputation, pattern mixture models

	complete		incomplete	
	Mean	SD	Mean	SD
anti1	1.49	1.54	1.89	1.78
read1	2.50	0.88	2.55	0.99
kidgen	0.52	0.50	0.48	0.50
momage	25.61	1.85	25.42	1.92
kidage	6.90	0.62	6.97	0.66
homecog	9.09	2.46	8.63	2.70
homeemo	9.35	2.23	9.01	2.41

Visualizing Some “Clusters”



Spaghetti Plot



Growth Curve Modeling

MLM for Longitudinal Data

	Student i in School j	Repeated measures at time t for Person i
Lv-1 model	$\text{MATH}_{ij} = \beta_{0j} + \beta_{1j}\ \text{SES}_{ij} + e_{ij}$	$\text{READ}_{ti} = \beta_{0i} + \beta_{1i}\ \text{TIME}_{ti} + e_{ti}$
Lv-2 model	$\beta_{0j} = \gamma_{00} + u_{0j}$ $\beta_{1j} = \gamma_{10} + u_{1j}$	$\beta_{0i} = \gamma_{00} + u_{0i}$ $\beta_{1i} = \gamma_{10} + u_{1i}$
Random effects	$\text{Var}\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} = \begin{bmatrix} \tau_0^2 & \tau_{01} \\ \tau_{01} & \tau_1^2 \end{bmatrix}$ $\text{Var}(e_{ij}) = \sigma^2$ τ_0^2, τ_1^2 = intercept & slope variance *between* schools σ^2 = *within*-school variation (across students)	$\text{Var}\begin{pmatrix} u_{0i} \\ u_{1i} \end{pmatrix} = \begin{bmatrix} \tau_0^2 & \tau_{01} \\ \tau_{01} & \tau_1^2 \end{bmatrix}$ $\text{Var}(e_{ti}) = \sigma^2$ τ_0^2, τ_1^2 = intercept & slope variance *between* persons σ^2 = *within*-person variation (across time)

Random Intercept Model (with **brms**)

```
> m00 <- brm(read ~ (1 | id), data = curran_long)
> summary(m00)

Group-Level Effects:
~id (Number of levels: 405)
              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sd(Intercept)     0.54      0.08     0.39     0.68 1.00     1131     1866

Family Specific Parameters:
      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sigma     1.55      0.04     1.48     1.62 1.00     2310     2707
```

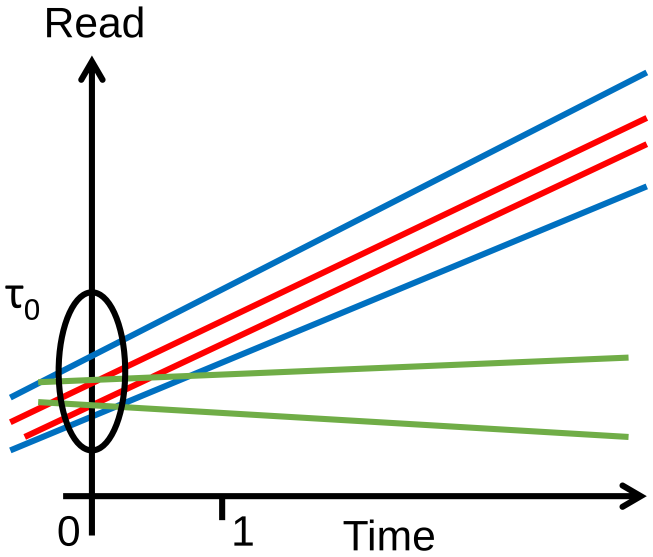
- Estimated ICC = 0.11

Linear Growth Model

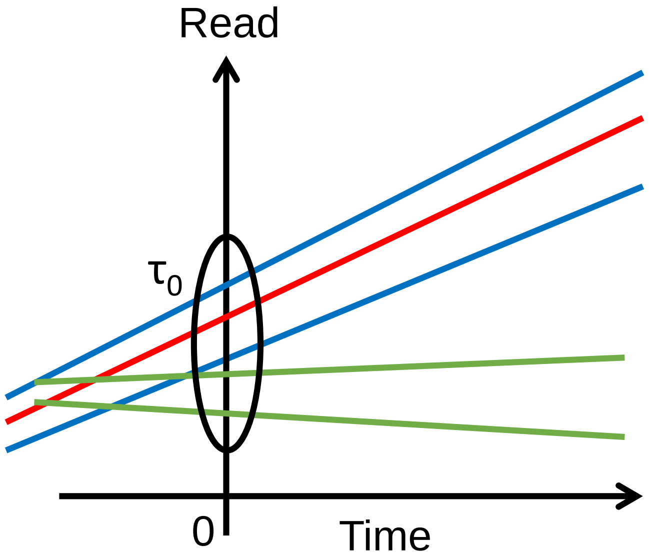
- Here time is treated as a continuous variable
  - Can handle varying occasions
  - Assume time is an *interval* variable
- Fit a linear regression line between time and outcome for each "cluster" (individual)

(Grand) Centering of Time

- Time = 1, 2, 3, 4



- Time = 0, 1, 2, 3



Compared to Repeated Measures ANOVA

- MLM and RM-ANOVA are the same in some basic situations
- Some advantages of MLM
  - Handles missing observations for individuals
    - Larger statistical power
  - Accommodates varying occasions
  - Allows clustering at a higher level (i.e., 3-level model)
  - Can include time-varying or time-invariant predictor variables

Random Slope of Time

- It is uncommon to expect the growth trajectory is the same for every person
- Therefore, usually the <u>baseline model</u> in longitudinal data analysis is the <u>random coefficient model of time</u>

R Output (brms)

```
Formula: read ~ time + (time | id)
   Data: curran_long (Number of observations: 1325)

Population-Level Effects:
          Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept     2.70      0.05     2.61     2.79 1.00     1970     2810
time          1.12      0.02     1.08     1.16 1.00     3568     3404
```

The estimated mean of read at time = 0 is **γ_{00} = 2.70 (*SE* = 0.05)**

The model predicts that the constant growth rate per 1 unit increase in time (i.e., **2 years**) is **γ_{10} = 1.12 (*SE* = 0.02)** units in read

```
Group-Level Effects:
~id (Number of levels: 405)
                   Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sd(Intercept)          0.76      0.04     0.68     0.84 1.00     1527     2500
sd(time)               0.27      0.03     0.22     0.32 1.00      741     1497
cor(Intercept,time)    0.30      0.12     0.07     0.54 1.00      828     1082
```

What do the *SD*s mean?

Piecewise Growth

Alternative Growth Shape

- For many problems, a linear growth model is at best an approximation
- Other common models (need 3+ time points)
  - Piecewise
  - Polynomial
  - Exponential, spline, etc

Piecewise Growth Model

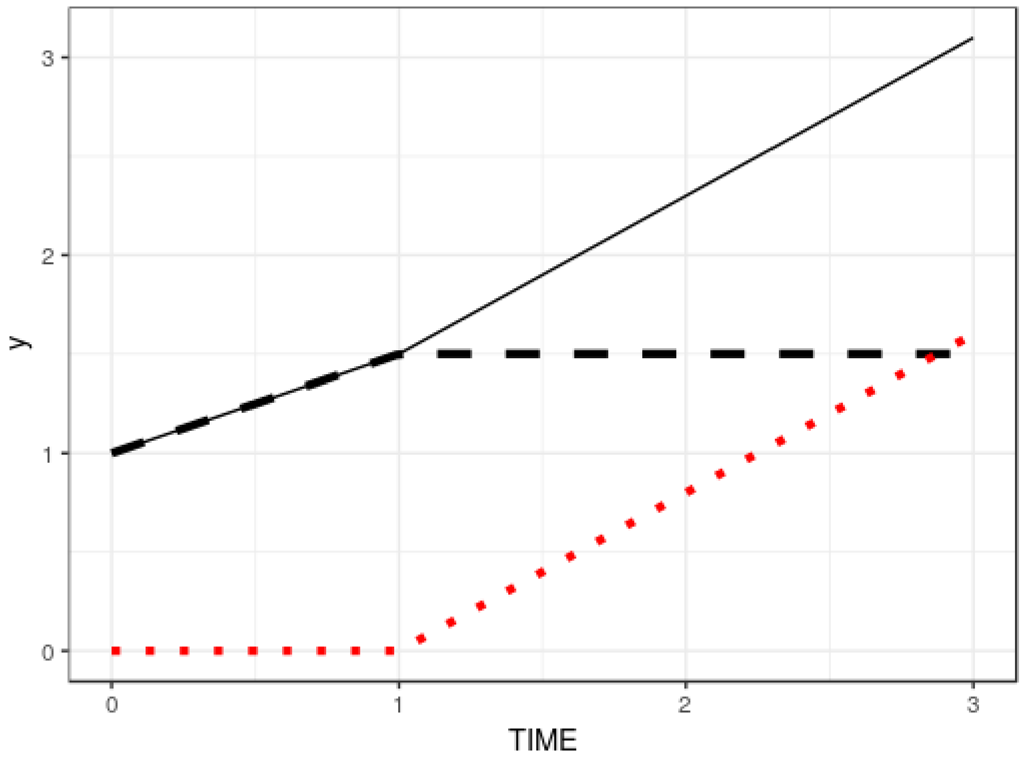
- Piecewise linear function
  - $Y = \beta_0 + \beta_1$ TIME, if TIME $\leq$ TIME^c
  - $Y = \beta_0 + \beta_1$ TIME^c + β_2 (TIME – TIME^c), if TIME > TIME^c
- β_0 = initial status (when TIME = 0)
- β_1 = phase 1 growth rate (up until TIME^c)
- β_2 = phase 2 growth rate (after TIME^c)

Coding of Time

time	phase1	phase2
0	0	0
1	1	0
2	1	1
3	1	2

$b_0 = 1, b_0 = 0.5, b_2 = 0.8$

- Dashed line: Phase 1
- Dotted line: Phase 2
- Combined: Linear piecewise growth



R Output

```
Formula: read ~ piece(time0, node = 1) + (piece(time0, node = 1) | id)

># Population-Level Effects:
>#                    Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
># Intercept              2.52      0.05     2.43     2.62 1.00     2193     2767
># piecetime0nodeEQ11     1.56      0.04     1.48     1.65 1.00     4774     3531
># piecetime0nodeEQ12     0.88      0.03     0.83     0.93 1.00     5974     3254
```

The model suggests that the average growth rate in phase 1 is 1.56 unit per unit time (*SE* = .04), but the growth rate decreases to 0.88 unit/time (*SE* = .03) subsequently.

R Output

```
Group-Level Effects:
~id (Number of levels: 405)
                                      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sd(Intercept)                             0.78      0.04     0.71     0.86 1.00     1737     2420
sd(piecetime0nodeEQ11)                    0.50      0.05     0.40     0.60 1.00     1065     2027
sd(piecetime0nodeEQ12)                    0.25      0.03     0.18     0.31 1.00     1226     2330
cor(Intercept,piecetime0nodeEQ11)         0.11      0.11    -0.10     0.34 1.00     1752     2886
cor(Intercept,piecetime0nodeEQ12)        -0.11      0.13    -0.35     0.15 1.00     3198     3331
cor(piecetime0nodeEQ11,piecetime0nodeEQ12) 0.76     0.15     0.41     0.97 1.00      587     1266
```

SD of the phase 1 growth rate is 0.50. So majority of children have growth rates between
1.56 +/- 0.50 = [1.06, 2.06]

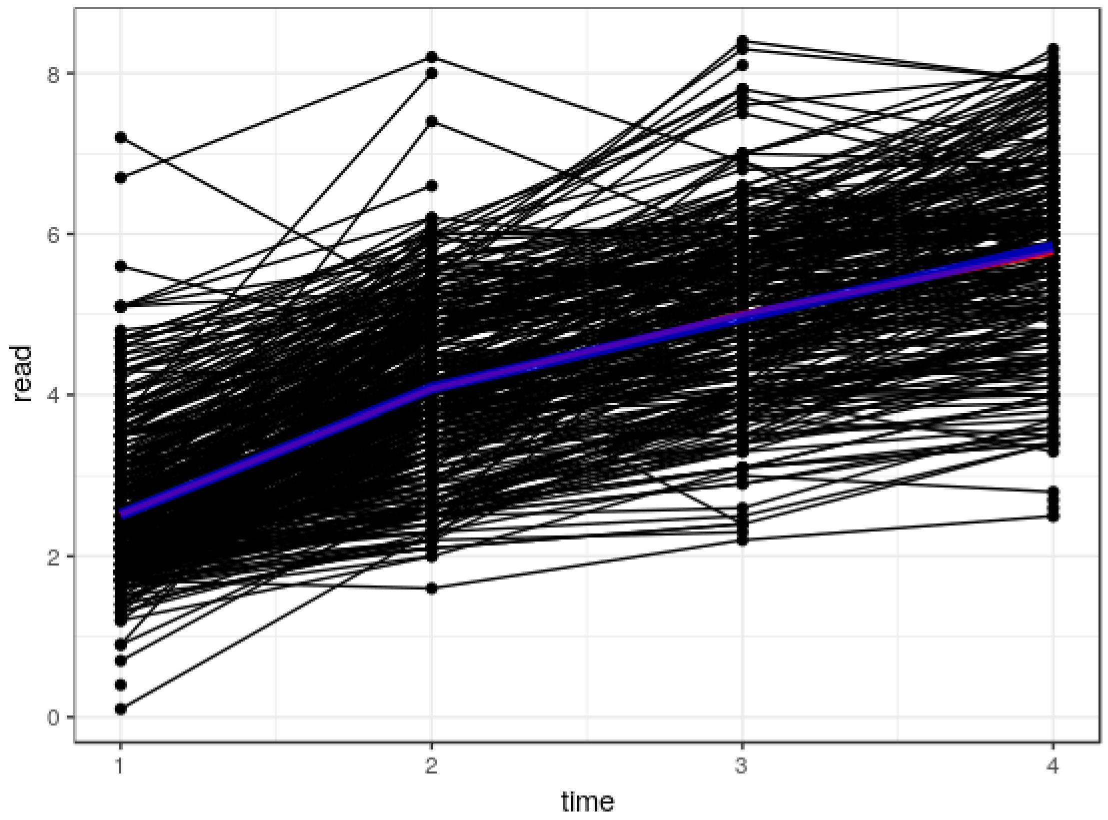
SD of the phase 2 growth rate is 0.25. So majority of children have growth rates between 0.88 +/- 0.25 = [0.63, 1.13]

Model Comparison

```
> loo(m_gca, m_pw)

Output of model 'm_gca':

looic      2953.8 67.1

Output of model 'm_pw':

looic      2658.1 70.3
```

- The model with lower LOOIC should be preferred
  - Note: the LOO in this example is not very stable due to the non-normality of the outcome

Predicted Average Trajectory



Including Predictors

Time-Invariant vs. Time-Varying Covariates

- Time-invariant predictor: Lv-2
- Time-varying predictor: Lv-1 (to be discussed next week)
  - "Cluster"-mean centering is generally recommended
  - However, usually not meaningful for "time." *Why?*

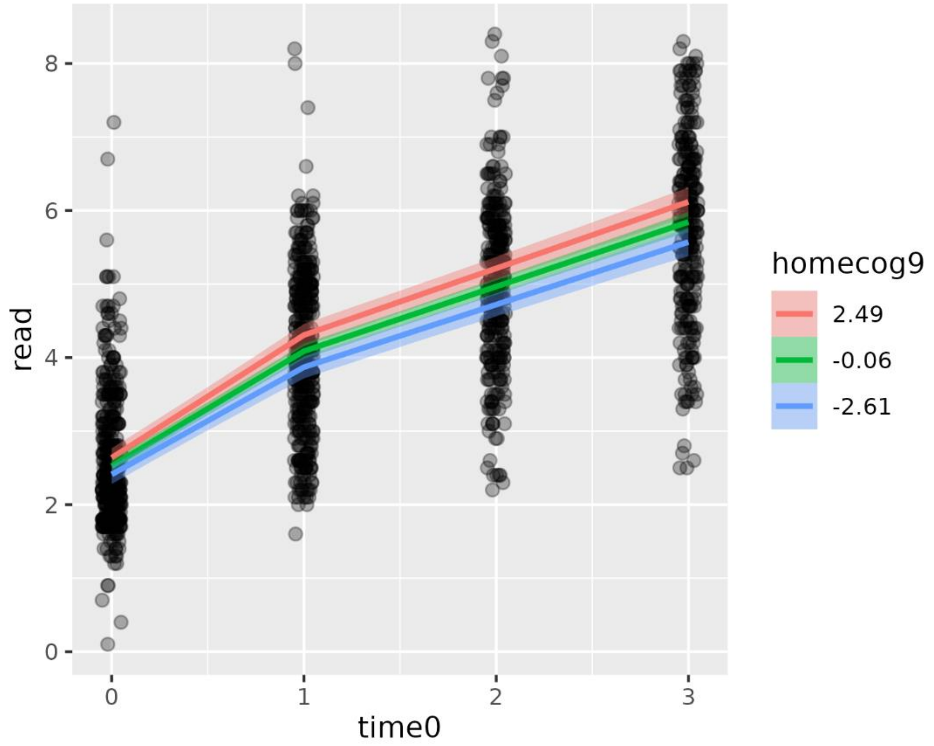
Time-Invariant Covariate

- Time-invariant predictor: Lv-2
  - Homecog (1-14): mother's cognitive stimulation at baseline
    - Centered at 9

Population-Level Effects:

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
Intercept	2.53	0.05	2.43	2.62	1.00	2954	3286
piecetime0nodeEQ11	1.57	0.04	1.48	1.65	1.00	4742	3038
piecetime0nodeEQ12	0.88	0.03	0.83	0.93	1.00	5749	3047
homecog9	0.04	0.02	0.01	0.08	1.00	2717	2356
piecetime0nodeEQ11:homecog9	0.04	0.02	0.01	0.07	1.00	5482	3328
piecetime0nodeEQ12:homecog9	0.01	0.01	-0.01	0.03	1.00	5759	3224

Cross-Level Interactions

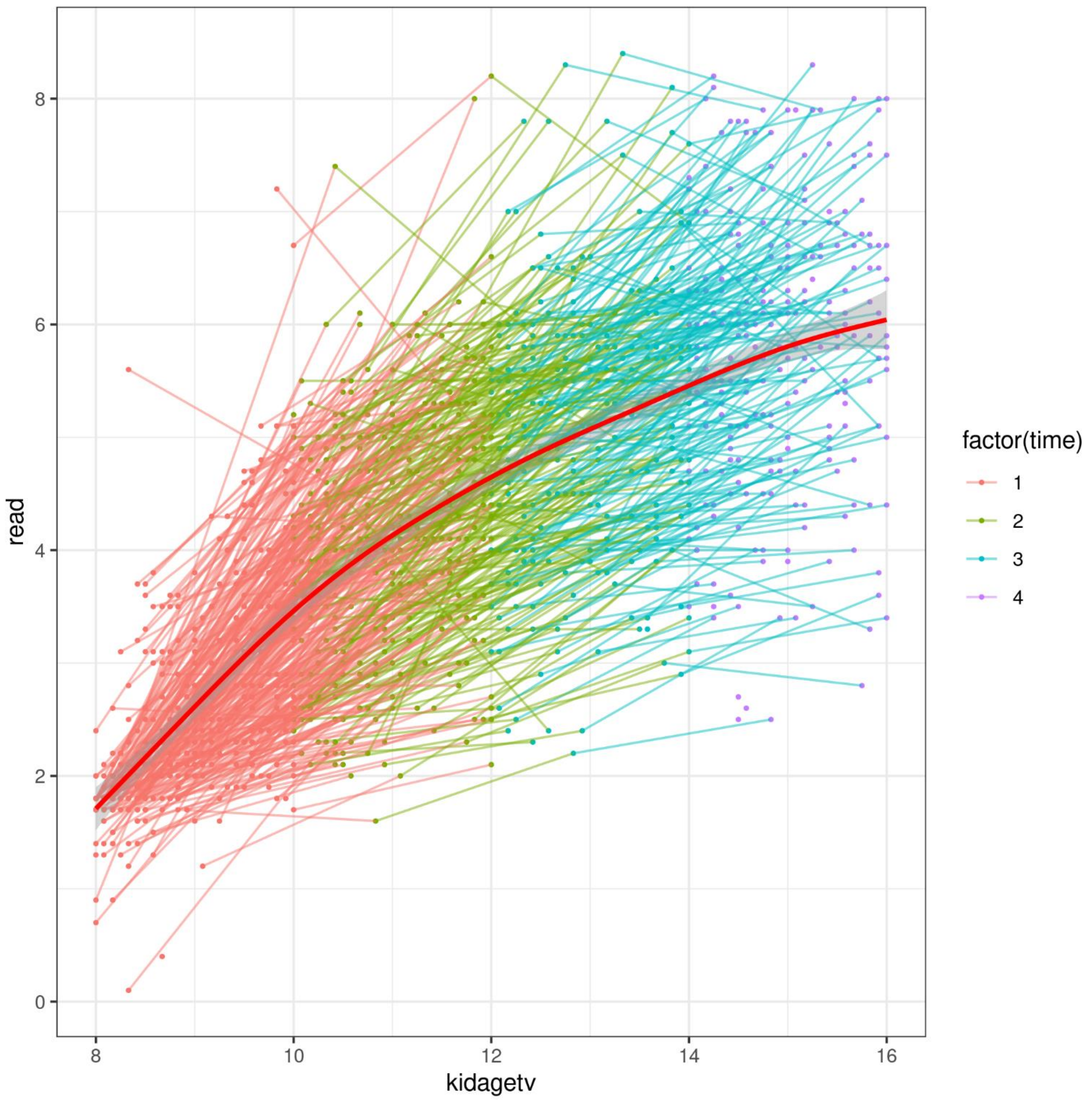


Handling Varying Occasions

Different “Time” Variables

- So far we model changes as a function of time passage from a common fixed point in history
  - I.e., when the study started
- In developmental research, one may be more interested in changes as a function of age
  - I.e., time passage from each person’s date of birth

- An advantage of MLM is that it does not require equal time intervals
  - Person 1: age 7 → age 9 → age 10
  - Person 2: age 5 → age 6.5 → age 8



read
kidagetv
0
2
4
6
8
8
10
12
14
16
factor(time)
1
2
3
4

Handling Varying Occasions

- Age as predictor (see textbook)

```
# Subtract age by 6
curran_long <- curran_long %>%
  mutate(kidagetv = kidage + time * 2,
         # Compute the age for each time point
         kidage6tv = kidagetv - 6)
# Fit the model
m_agesq <- brm(read ~ kidage6tv + I(kidage6tv^2) + (kidage6tv + I(kidage6tv^2) | id),
               data = curran_long)
summary(m_agesq)
```

```
Group-Level Effects:
~id (Number of levels: 405)
                           Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sd(Intercept)                  0.51      0.11     0.28     0.74 1.00     1871     2192
sd(kidage6tv)                  0.40      0.05     0.31     0.49 1.00      573      784
sd(Ikidage6tvE2)               0.03      0.00     0.02     0.04 1.01      309      522
cor(Intercept,kidage6tv)      -0.92      0.07    -0.99
cor(Intercept,Ikidage6tvE2)    0.81      0.12     0.52
cor(kidage6tv,Ikidage6tvE2)   -0.95      0.02    -0.98

Population-Level Effects:
             Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept       -0.32      0.10    -0.51    -0.13 1.00
kidage6tv        1.13      0.04     1.05     1.21 1.00
Ikidage6tvE2    -0.05      0.00    -0.06    -0.04 1.00
```

The model suggests that the average initial growth rate is 1.13 unit per year (*SE* = .04)

The growth rate slows down by .05 every year. Therefore, at Wave 2 (two years later), the growth rate is 1.02

```
Group-Level Effects:
~id (Number of levels: 405)
                              Estimate Est.Error l-95% CI u-
sd(Intercept)                     0.51      0.11     0.28
sd(kidage6tv)                     0.40      0.05     0.31     0.49 1.00      573      784
sd(Ikidage6tvE2)                  0.03      0.00     0.02     0.04 1.01      309      522
cor(Intercept,kidage6tv)         -0.92      0.07    -0.99
cor(Intercept,Ikidage6tvE2)       0.81      0.12     0.52
cor(kidage6tv,Ikidage6tvE2)      -0.95      0.02    -0.98
```

The 68% plausible range of the initial growth rate is 1.13 +/- 0.44 = [0.69, 1.57]

The 68% plausible range of the change in growth rate is -0.05 +/- 0.03 = [-0.02, -0.08]

```
Population-Level Effects:
             Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept       -0.32      0.10    -0.51    -0.13 1.00     5713     3244
kidage6tv        1.13      0.04     1.05     1.21 1.00     4869     3388
Ikidage6tvE2    -0.05      0.00    -0.06    -0.04 1.00     5945     3299
```