Multivariate: Mixed models

Table of contents

1 Goals

1.1 Goals

1.1.1 Goals of this section

- **Multiple measures** of the same thing or related things as an **outcome**
  - Possibly over time
- Want the variables **separate**: Not PCA / FA
- In this section:
  - Last time: MANOVA and repeated measures ANOVA
  - Mixed models (this week)

1.1.2 Goals of this lecture

- **Mixed models** as an approach to repeated measures
  - Focus on *individual* change
  - Fewer problems with *missing data*
  - Continuous and unevenly spaced *time*
  - Flexible with *predictors* (continuous and categorical)
- **Way more complex & interesting than we have time to talk about!**
  - Take another class: Longitudinal, Multilevel, Categorical, SEM

2 Linear mixed model

2.1 Linear mixed model

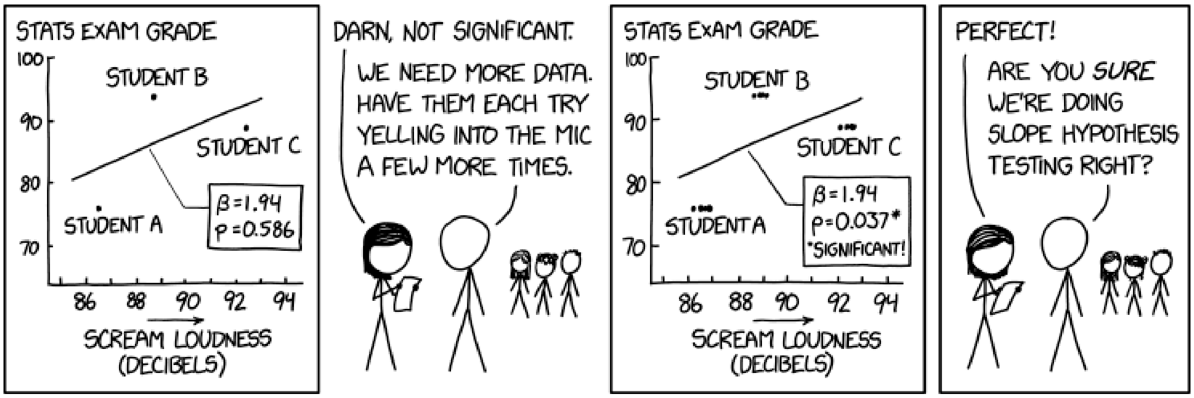
2.1.1 Linear mixed model

- **Also known as** random coefficient models, multilevel models, nested models, hierarchical linear models, random effects models
- Developed in *different disciplines*
  - Random coefficient models from *statistics and biostatistics*
  - Multilevel models from *education*

2.1.2 Linear mixed model

- Model for **non-independent observations**
  - Cross-sectional
    * Multiple schoolchildren with the same teacher
    * Employees who work in teams or workgroups
  - Longitudinal
    * **Multiple observations from the same individual over time**
- Observations from *same* class/team/person are **more similar to one another** than observations from *different* classes/teams/persons

2.1.3 How not to do it



https://xkcd.com/2533

2.1.4 Non-independence

- **Non-independence** means that there is some redundancy (or *correlation*) between observations
  - **Effective** sample size is smaller that the **actual** sample size
    - Collect 100 observations but we only have (for example) 72 obs' worth of information, due to correlations between obs
- Smaller effective sample size means standard error is **underestimated** if you ignore non-independence
  - How much the standard errors are underestimated depends on how much the observations are related to one another

2.1.5 Linear mixed model: Motivation

- Linear mixed model (LMM): Extension of general linear model (GLM)
  - Partitions **variation**, just like ANOVA and regression
  - But more ways to partition and more control over the form

2.1.6 Linear mixed model: Motivation

- **How are observations related to one another?**
  - Linear regression: They're not (Independence)

  - Between-subjects ANOVA: They're not (Independence)
  - Repeated-measures ANOVA: According to "compound symmetry"
  - LMM has two ways to do this
    * Random effects (this class)
    * Correlated residuals (not this class)

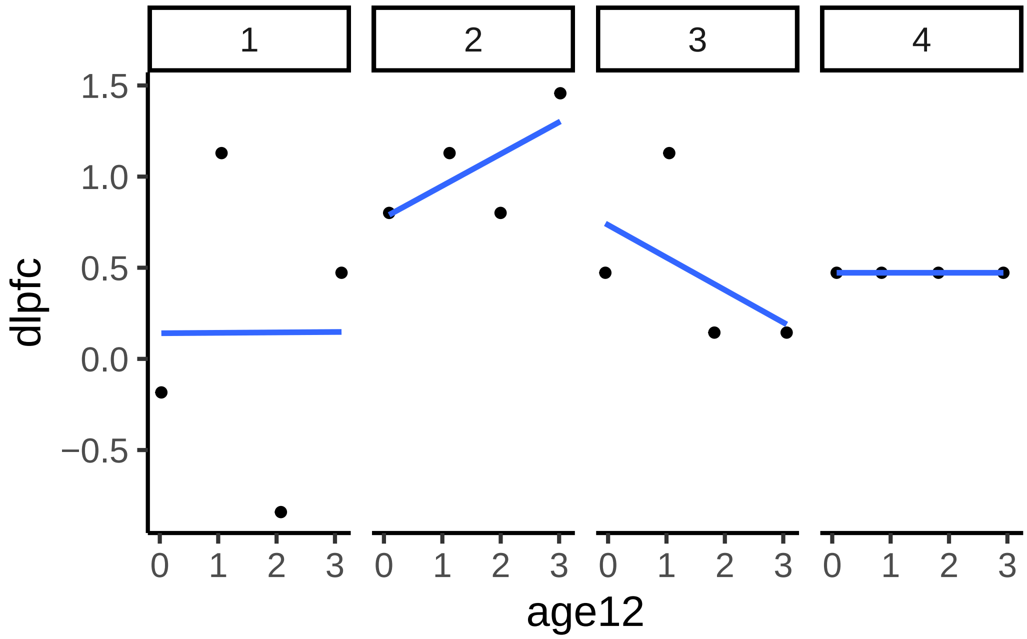
2.1.7 Random effects

- Relationship between time and outcome
  - X axis = time
  - Y axis = outcome
- Random effects: **Relationship can be different for each person**
  - Differences in intercept = random intercepts
  - Differences in change over time = random slopes
  - Can have one or other or both

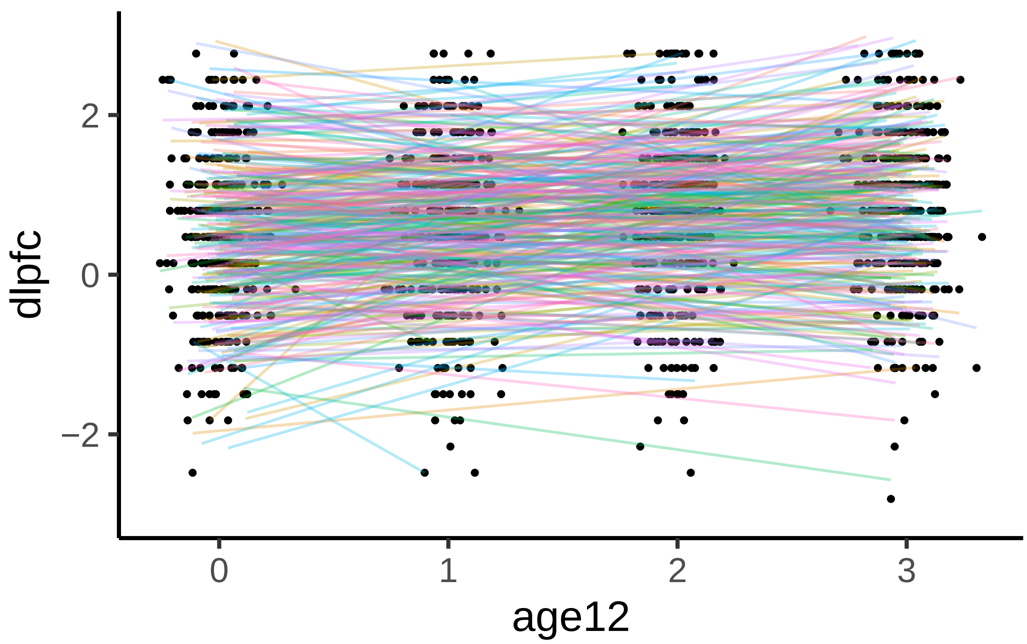
2.1.8 Data: Executive functioning dataset

id	sex	tx	wave	dlpfc	ef	age	age12
1	1	0	1	-0.184	2.167	12.027	0.027
1	1	0	2	1.129	1.806	13.058	1.058
1	1	0	3	-0.840	1.444	14.074	2.074
1	1	0	4	0.472	2.889	15.112	3.112
2	1	0	1	0.801	0.722	12.089	0.089
2	1	0	2	1.129	1.444	13.124	1.124
2	1	0	3	0.801	1.806	13.997	1.997
2	1	0	4	1.457	2.528	15.021	3.021
3	1	1	1	0.472	3.250	11.953	-0.047
3	1	1	2	1.129	3.250	13.048	1.048
3	1	1	3	0.144	2.528	13.820	1.820
3	1	1	4	0.144	2.528	15.058	3.058
4	0	0	1	0.472	3.611	12.076	0.076
4	0	0	2	0.472	4.333	12.845	0.845
4	0	0	3	0.472	3.972	13.818	1.818
4	0	0	4	0.472	3.972	14.931	2.931

2.1.9 Individual trajectories for first 4 people



2.1.10 All individual trajectories: Spaghetti plot



2.1.11 Assumptions

- Linear regression assumes **independence** of observations
  - Definitely not true here
  - What should we do?
- We can still use the regression lines for each person
  - Non-independence only a problem for **estimating standard errors**
  - Can use the estimates of individual **intercepts** and **slopes**

2.1.12 So I just report 100 intercepts and slopes?

- What do we do with all those regression lines?
  - **Fixed effects**: Average of **intercepts** and **slopes**
  - **Random effects**: Variance of **intercepts** and **slopes**

> **!** Important
>
> - You have control over the model: Everyone **can** have a different *slope* but they don't **have to**
>   - Both random intercepts and slopes: **People can have different change over time**
>   - Only random intercepts: **Everyone has the same change over time**

2.1.13 Mixed model: Output in R

```
Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
  method [lmerModLmerTest]
Formula: dlpfc ~ 1 + age12 + tx + age12 * tx + (1 + age12 | id)
   Data: ef_uni

     AIC      BIC   logLik deviance df.resid
  3502.5   3543.9  -1743.3   3486.5     1296

Scaled residuals:
     Min       1Q   Median       3Q      Max
-2.89823 -0.50229 -0.02245  0.51797  2.80891

Random effects:
```

```
 Groups   Name        Variance Std.Dev. Corr
 id       (Intercept) 0.67110  0.8192
          age12       0.05248  0.2291   -0.37
 Residual             0.48857  0.6990
Number of obs: 1304, groups:  id, 342

Fixed effects:
             Estimate Std. Error        df t value Pr(>|t|)
(Intercept)   0.58230    0.07792 341.09983   7.473 6.63e-13 ***
age12         0.11158    0.03059 335.26580   3.648 0.000307 ***
tx           -0.06801    0.10933 342.28563  -0.622 0.534346
age12:tx      0.01185    0.04298 338.26053   0.276 0.782972
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
         (Intr) age12  tx
age12    -0.550
tx       -0.713  0.392
age12:tx  0.391 -0.712 -0.552
```

2.1.14 Mixed model: Output in SPSS

Mixed Model Analysis

Model Dimension[a]

		Number of Levels	Covariance Structure	Number of Parameters	Subject Variables
Fixed Effects	Intercept	1		1	
	age12	1		1	
	tx	1		1	
	age12 * tx	1		1	
Random Effects	Intercept + age12[b]	2	Unstructured	3	id
Residual				1	
Total		6		8	

a. Dependent Variable: dlpfc.

b. As of version 11.5, the syntax rules for the RANDOM subcommand have changed. Your command syntax may yield results that differ from those produced by prior versions. If you are using version 11 syntax, please consult the current syntax reference guide for more information.

Information Criteria[a]

-2 Log Likelihood	3486.503
Akaike's Information Criterion (AIC)	3502.503
Hurvich and Tsai's Criterion (AICC)	3502.615
Bozdogan's Criterion (CAIC)	3551.889
Schwarz's Bayesian Criterion (BIC)	3543.889

The information criteria are displayed in smaller-is-better form.

a. Dependent Variable: dlpfc.

Fixed Effects

Type III Tests of Fixed Effects[a]				
Source	Numerator df	Denominator df	F	Sig.
Intercept	1	341.102	55.853	.000
age12	1	335.258	13.305	.000
tx	1	342.288	.387	.534
age12 * tx	1	338.253	.076	.783

a. Dependent Variable: dlpfc.

Estimates of Fixed Effects[a]							
						95% Confidence Interval	
Parameter	Estimate	Std. Error	df	t	Sig.	Lower Bound	Upper Bound
Intercept	.582297	.077915	341.102	7.473	.000	.429043	.735552
age12	.111581	.030590	335.258	3.648	.000	.051408	.171754
tx	-.068006	.109332	342.288	-.622	.534	-.283052	.147040
age12 * tx	.011849	.042984	338.253	.276	.783	-.072700	.096399

a. Dependent Variable: dlpfc.

Covariance Parameters

Estimates of Covariance Parameters[a]								
						95% Confidence Interval	
Parameter		Estimate	Std. Error	Wald Z	Sig.	Lower Bound	Upper Bound
Residual		.488568	.027363	17.855	.000	.437777	.545253
Intercept + age12 [subject = id]	UN (1,1)	.671099	.080192	8.369	.000	.530975	.848201
	UN (2,1)	-.070179	.025878	-2.712	.007	-.120898	-.019460
	UN (2,2)	.052476	.012977	4.044	.000	.032320	.085204

a. Dependent Variable: dlpfc.

Random Effect Covariance Structure (G)[a]		
	Intercept | id	age12 | id
Intercept | id	.671099	-.070179
age12 | id	-.070179	.052476

Unstructured

a. Dependent Variable: dlpfc.

2.1.15 Fixed effects: Means or averages

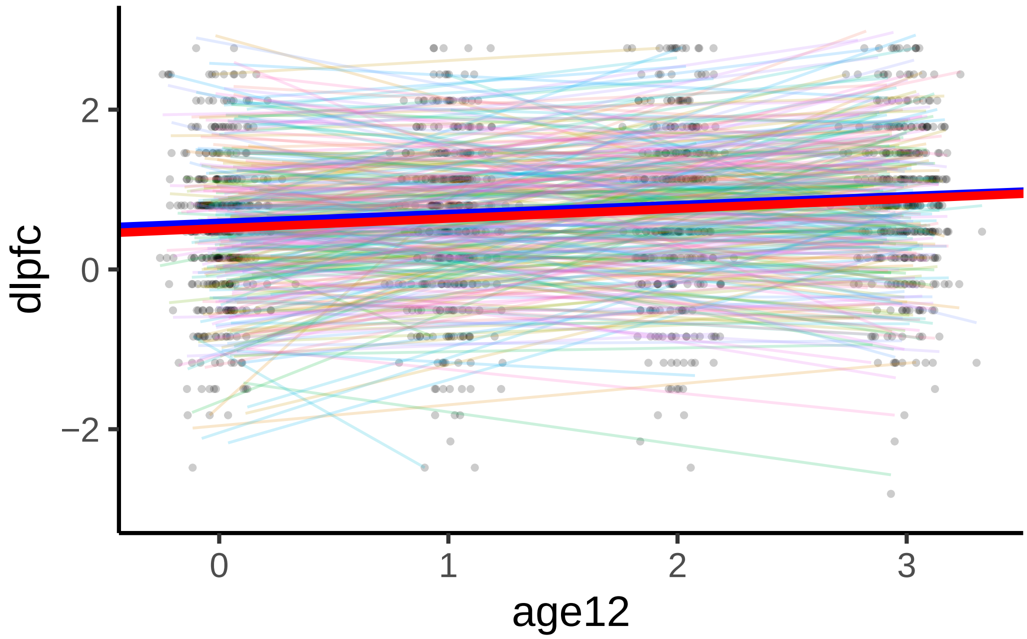
term	estimate	std.error	statistic	df	p.value
(Intercept)	0.582	0.078	7.473	341.100	0.000
age12	0.112	0.031	3.648	335.266	0.000
tx	-0.068	0.109	-0.622	342.286	0.534
age12:tx	0.012	0.043	0.276	338.261	0.783

- R: t-tests with no df or p-values with just **lme4** package
  - df and p-values with **lmerTest** package
- SPSS: t-tests

2.1.16 Fixed effects: Interpretation

- $Y = 0.582 + 0.112(age12) + -0.068(tx) + 0.012(age12 * tx)$
  - `tx` $= 0$ (in *blue*): $Y = 0.582 + 0.112(age12)$
    - Intercept: Expected `dlpfc` when `age12` $= 0$, for group `tx` $= 0$
    - Slope: Change in `dlpfc` for 1 unit change in `age12`, for group `tx` $= 0$
  - `tx` $= 1$ (in *red*): $Y = 0.514 + 0.123(age12)$
    - Intercept: Expected `dlpfc` when `age12` $= 0$, for group `tx` $= 1$
    - Slope: Change in `dlpfc` for 1 unit change in `age12`, for group `tx` $= 1$

2.1.17 Fixed effects: Figure



2.1.18 Random effects: Variances

term	estimate
var___(Intercept)	0.671
cov___(Intercept).age12	-0.070
var___age12	0.052
var___Observation	0.489

- R: No test statistics
- SPSS: z-tests (p-value/2 for **variances**)

2.1.19 Random effects: Interpretation

- **Intercept variance**: *Variance* of individual *intercepts*
- **Slope variance**: *Variance* of individual *slopes*
- **Correlation** between intercept and slope: Correlation between individual *intercepts and slopes*
  - How is a person's intercept related to their slope?
- **Residual**: Error
  - How well we do at predicting the individual trajectory

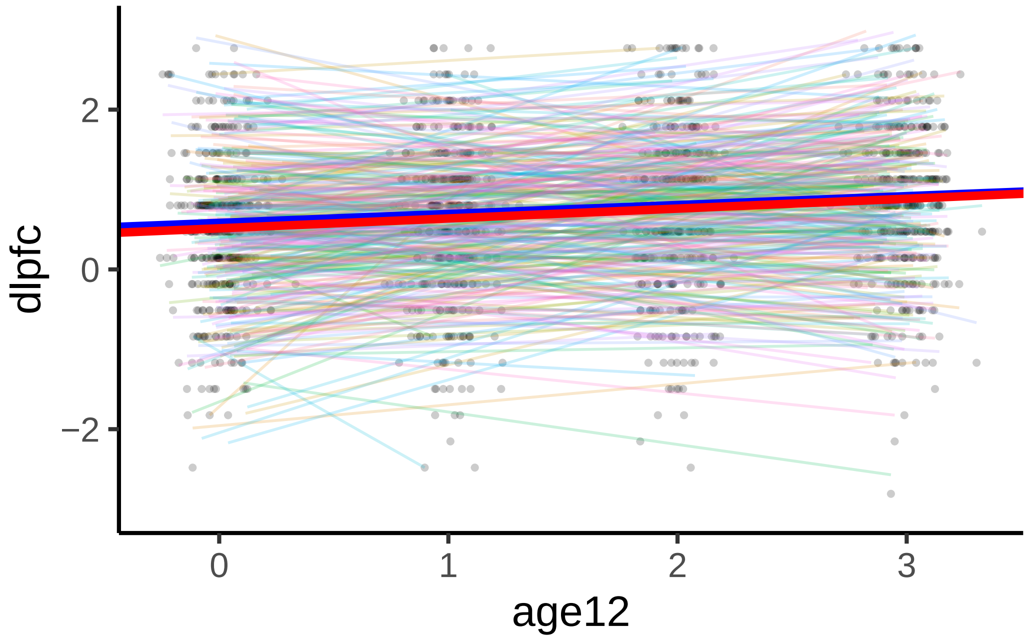
2.1.20 Prediction interval: Fixed + random

- **Average effects** with **individual variation**
  - What do typical **individual effects** look like?
  - Assume normally distributed variance: $estimate \pm 1.96 \times SD$
- **Prediction intervals**
  - Interval for likely values of **individual** intercepts and slopes
- **Not** confidence intervals
  - Sampling distribution of test statistic
  - You get those for fixed effects

2.1.21 Prediction interval: Fixed + random

- Average **intercept** (for `tx = 0`) = 0.582
  - $1.96 \times SD = 1.96 \times 0.819 = 1.606$
  - 95% of **individual intercepts** are in [−1.023, 2.188]
- Average **slope** (for `tx = 0`) = 0.112
  - $1.96 \times SD = 1.96 \times 0.229 = 0.449$
  - 95% of **individual slopes** are in [−0.337, 0.561]

2.1.22 Means + individual variability



3 LMM: Some more details

3.1 Equations

3.1.1 Linear mixed model: Equations

$$Y = \underbrace{\mathbf{X}\beta}_{\text{fixed effects}} + \underbrace{\mathbf{Z}\gamma}_{\text{random effects}} + \underbrace{\epsilon}_{\text{residual}}$$

- **Fixed effects**: Average effects of predictors
  - Like regression coefficients
- **Random effects**: Individual variation around those averages
  - Variances and covariances
- **Residual**: Error in predicting individual trajectories
  - Also a variance (but usually not interpreted)

3.1.2 Linear mixed model: Equations

- Random *intercept* and *slope*
  - $Y_{ij} = \underbrace{\beta_{00} + \beta_{10}(age12_{ij}) + \beta_{01}(tx_i) + \beta_{11}(age12_{ij})(tx_i)}_{\text{fixed effects}} + \underbrace{r_{0i} + r_{1i}(age12_{ij})}_{\text{random effects}} + \underbrace{e_{ij}}_{\text{residual}}$
- Random effects are normally distributed with mean 0 and variance-covariance matrix **G**
  - $\gamma \sim N(0, \mathbf{G})$
  - $\mathbf{G} = \begin{bmatrix} {\sigma_{r_{0i}}}^2 & \\ \sigma_{r_{0i}r_{1i}} & {\sigma_{r_{1i}}}^2 \end{bmatrix}$

3.1.3 Linear mixed model: Equations

- Random *intercept* only
  - $Y_{ij} = \underbrace{\beta_{00} + \beta_{10}(age12_{ij}) + \beta_{01}(tx_i) + \beta_{11}(age12_{ij})(tx_i)}_{\text{fixed effects}} + \underbrace{r_{0i}}_{\text{random effects}} + \underbrace{e_{ij}}_{\text{residual}}$
- Random effects are normally distributed with mean 0 and variance-covariance matrix **G**
  - $\gamma \sim N(0, \mathbf{G})$
  - **No random slopes**, so $\mathbf{G} = \begin{bmatrix} {\sigma_{r_{0i}}}^2 \end{bmatrix}$

3.1.4 Example data: Equations

- Random *intercept* and *slope*
  - $Y_{ij} = \underbrace{0.582 + 0.112(age12_{ij}) - 0.068(tx) + 0.012(tx)(age12_{ij})}_{\text{fixed effects}} + \underbrace{r_{0i} + r_{1i}(age12_{ij})}_{\text{random effects}} + \underbrace{e_{ij}}_{\text{residual}}$
- Random effects are normally distributed with mean 0 and variance-covariance matrix **G**
  - $\mathbf{G} = \begin{bmatrix} {\sigma_{r_{0i}}}^2 & \sigma_{r_{0i}r_{1i}} \\ \sigma_{r_{0i}r_{1i}} & {\sigma_{r_{1i}}}^2 \end{bmatrix} = \begin{bmatrix} 0.671 & -0.07 \\ -0.07 & 0.052 \end{bmatrix}$
  - Can convert variances and covariances into SDs and correlations for interpretation
    * e.g., $\sqrt{0.671} = 0.819$

3.2 Predictors

3.2.1 Adding predictors to the model

- The example model has two predictors
  - Treatment (`tx`): Categorical (dummy code)
    - Time-**invariant** predictor: **Same value** at all times
  - Age (`age12`): *Continuous*
    - Time-**varying** predictor: **Different value** at each time
- Two types of predictors are entered into model in different ways
  - Peugh, J. L. (2010). A practical guide to multilevel modeling. *Journal of school psychology, 48(1)*, 85-112.

3.2.2 Data: Executive functioning dataset

id	sex	tx	wave	dlpfc	ef	age	age12
1	1	0	1	-0.184	2.167	12.027	0.027
1	1	0	2	1.129	1.806	13.058	1.058
1	1	0	3	-0.840	1.444	14.074	2.074
1	1	0	4	0.472	2.889	15.112	3.112
2	1	0	1	0.801	0.722	12.089	0.089
2	1	0	2	1.129	1.444	13.124	1.124
2	1	0	3	0.801	1.806	13.997	1.997
2	1	0	4	1.457	2.528	15.021	3.021
3	1	1	1	0.472	3.250	11.953	-0.047
3	1	1	2	1.129	3.250	13.048	1.048
3	1	1	3	0.144	2.528	13.820	1.820
3	1	1	4	0.144	2.528	15.058	3.058
4	0	0	1	0.472	3.611	12.076	0.076
4	0	0	2	0.472	4.333	12.845	0.845
4	0	0	3	0.472	3.972	13.818	1.818
4	0	0	4	0.472	3.972	14.931	2.931

3.2.3 LMM = Multilevel model

- LMM is also a **multilevel** model where
  - Level 1: Occasions
  - Level 2: Person

    - Multiple occasions are nested within each person (L1 w/in L2)
- LMM can be re-written in terms of L1 and L2
    - **Time-varying predictors** go in **L1** (occasions) part
    - **Time-invariant predictors** go in **L2** (person) part

3.2.4 Multilevel models: Adding predictors

- Level 1: Trajectories
    - $Y_{ij} = \pi_{0i} + \pi_{1i}(age12_{ij}) + e_{ij}$
- Level 2: People
    - $\pi_{0i} = \beta_{00} + \beta_{01}(tx_i) + r_{0i}$
    - $\pi_{1i} = \beta_{10} + \beta_{11}(tx_i) + r_{1i}$
- Combined: Put them together
    - $Y_{ij} = \beta_{00} + \beta_{10}(age12_{ij}) + \beta_{01}(tx_i) + \beta_{11}(age12_{ij})(tx_i) +$ $r_{0i} + r_{1i}(age12_{ij}) + e_{ij}$

3.3 Centering

3.3.1 Why center predictors?

1. **Interactions**: Reduce collinearity
2. Interactions and **more generally**: Improve interpretability
    - Intercept: Predicted outcome when $X = 0$
    - What if $X = 0$ doesn't exist or is somewhere useless?
3. **Mixed models**: Unconflate time-specific (L1) and person-specific (L2) relationships

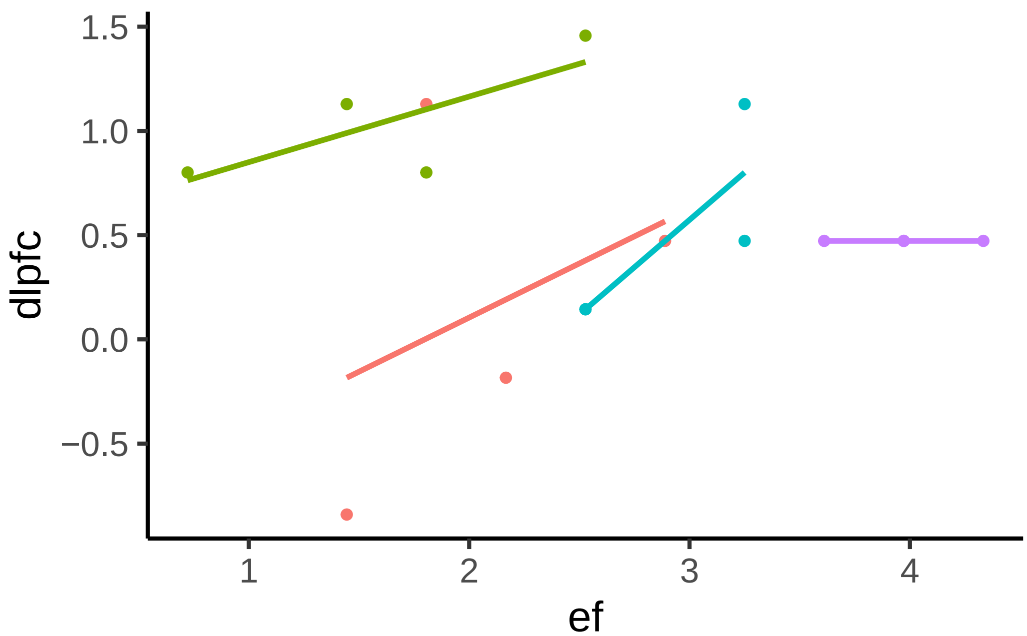
3.3.2 Why does this matter so much for mixed models?

- Level 1 (occasion) observations have **two kinds of information**
    - Occasion (L1)
    - Person (L2)
- If you ask me one day if I'm depressed, that gives you information about
    - How depressed I am **that day** (occasion, L1)
    - How depressed I **generally** am (person, L2)

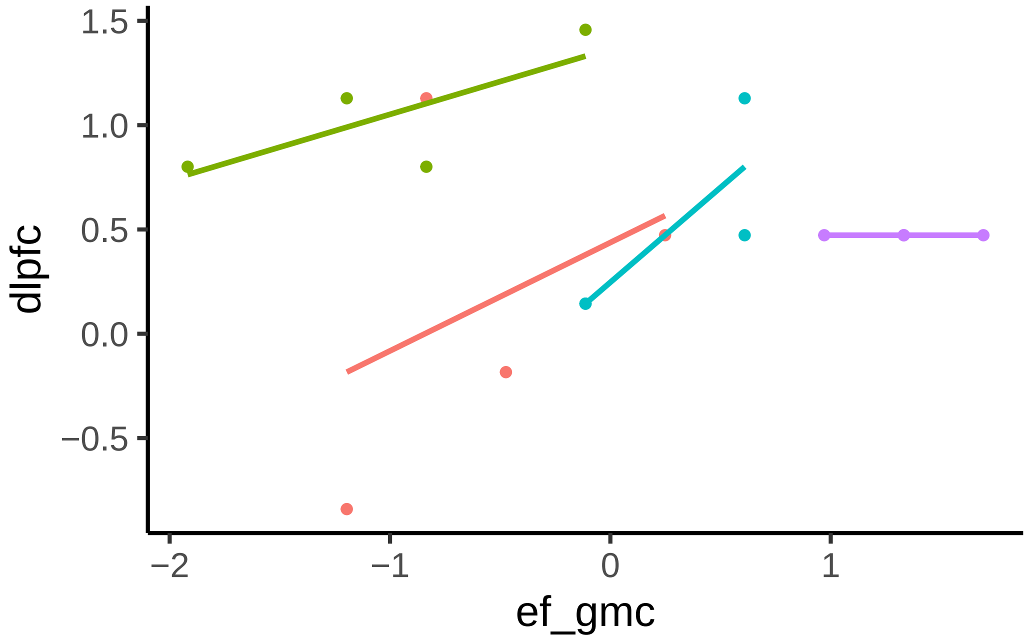
3.3.3 Centering is more complicated

- Grand mean centering (GMC)
  - Center all observations at the **grand mean** of **all observations**
  - *Doesn't change* the relationships among variables
- Centering within cluster (CWC)
  - Center each person's observations at the **mean of that person**
  - *Does change* the relationships among variables

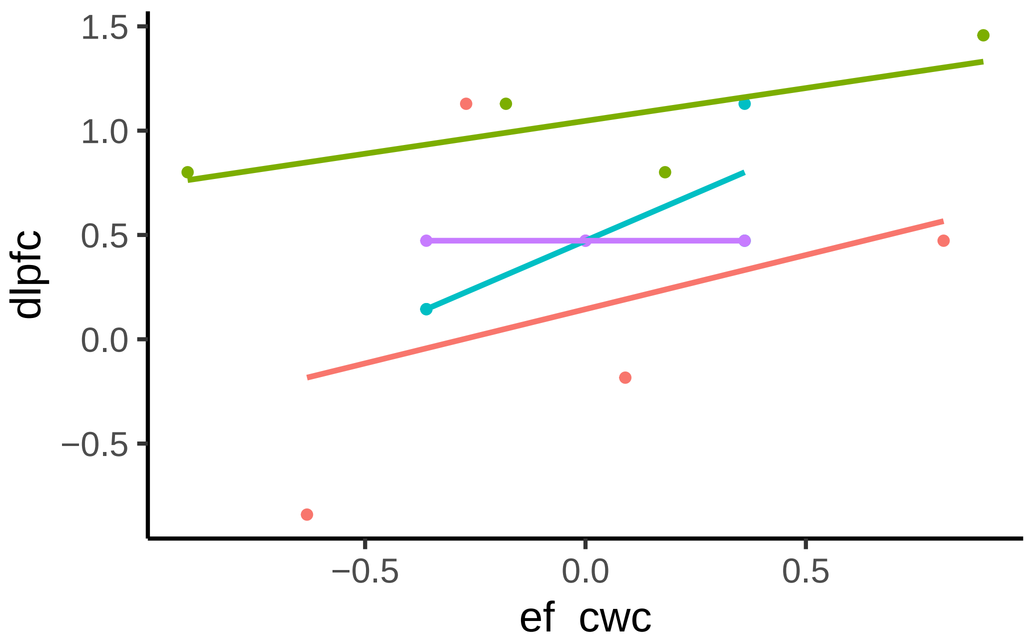
3.3.4 Figure: Uncentered



3.3.5 Figure: GMC



3.3.6 Figure: CWC



3.3.7 GMC vs CWC

- Centering changes the **context** for the different clusters (L2: People)
  - GMC **maintains mean differences** between people on L1 predictor
    - What is a person like *compared to other people*?
  - CWC l**eliminates differences** between people on L1 predictor
    - What are people like *compared to their own mean*?
- *Different contexts* means *different interpretations* for both level 1 and level 2 predictors

3.3.8 Centering predictors: Some references

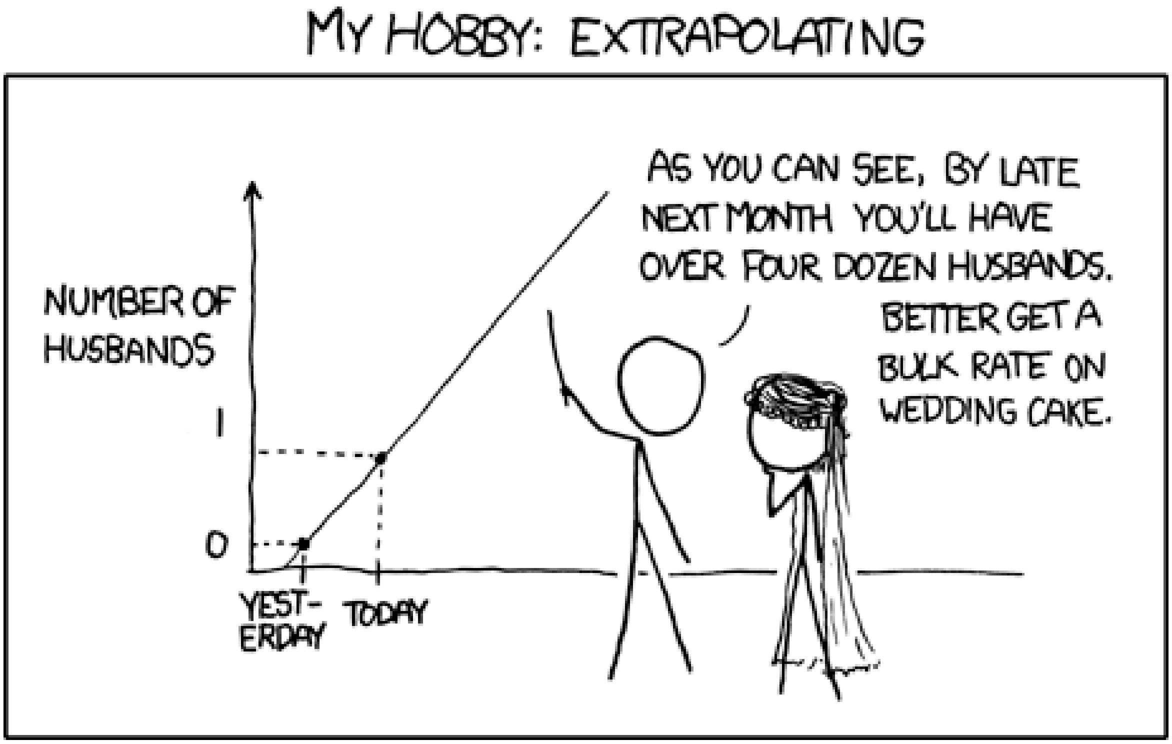
- Yaremych, H. E., Preacher, K. J., & Hedeker, D. (2021). Centering categorical predictors in multilevel models: Best practices and interpretation. *Psychological Methods.*
- Rights, J. D., Preacher, K. J., & Cole, D. A. (2020). The danger of conflating level-specific effects of control variables when primary interest lies in level-2 effects. *British Journal of Mathematical and Statistical Psychology, 73*, 194-211.
- Hamaker, E. L., & Muthén, B. (2020). The fixed versus random effects debate and how it relates to centering in multilevel modeling. *Psychological methods, 25(3)*, 365.
- Hoffman, L. (2019). On the interpretation of parameters in multivariate multilevel models across different combinations of model specification and estimation. *Advances in methods and practices in psychological science, 2(3)*, 288-311.
- West, S. G., Ryu, E., Kwok, O. M., & Cham, H. (2011). Multilevel modeling: Current and future applications in personality research. *Journal of personality, 79(1)*, 2-50.
- Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: a new look at an old issue. *Psychological methods, 12(2)*, 121.

3.3.9 Centering time

- "Time" is a special predictor
  - Center time variable so that 0 is at a meaningful point
  - Intercept: Expected value of outcome when $X = 0$
    - Age = 0?
    - Baseline?
    - Centered age at specific time

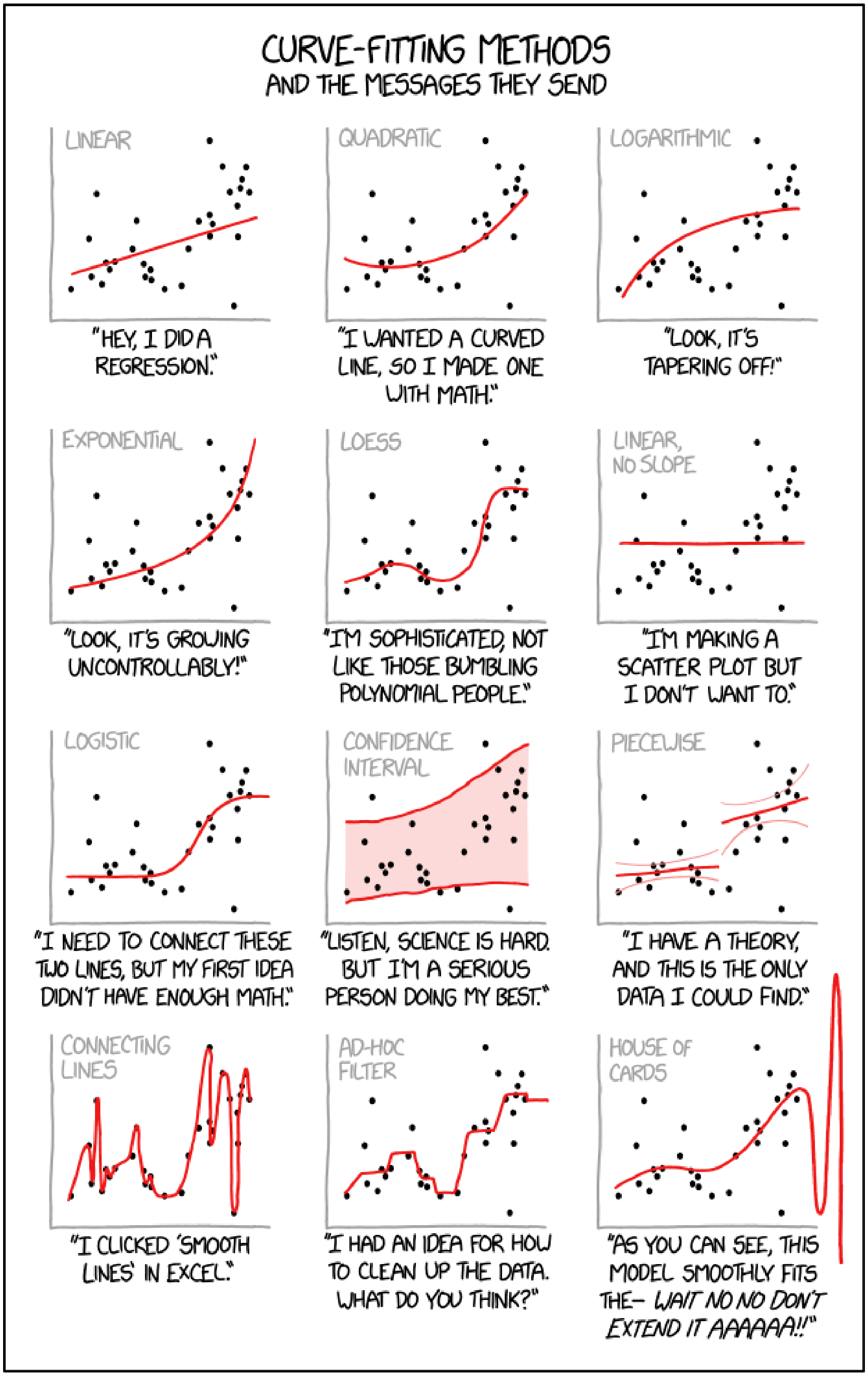
3.4 Shape of change

3.4.1 Is it linear?



http://www.xkcd/605

3.4.2 If not linear, then what?



https://xkcd.com/2048/

3.4.3 Shape of change: L1 equation

- Linear change
  - $Y_{ij} = \pi_{0i} + \pi_{1i}(age12_{ij}) + e_{ij}$
- Quadratic change
  - $Y_{ij} = \pi_{0i} + \pi_{1i}(age12_{ij}) + \pi_{2i}(age12_{ij})^2 + e_{ij}$
- Logarithmic change
  - $Y_{ij} = \pi_{0i} + \pi_{1i}(ln(age12_{ij})) + e_{ij}$

3.4.4 Non-linear change

- Many phases of development or change are non-linear
  - Increase followed by plateau / maintanence
  - Decrease to a set point
  - Sometimes reflect floor or ceiling effects
- Non-linear change in inherently more complex
  - Straight lines are easy

3.5 Other stuff

3.5.1 Intraclass correlation (ICC)

- Quantifies **non-independence** in repeated outcome
- Use "random effects ANOVA" or "unconditional mixed model"
  - Like "no predictors" model from logistic regression
- Ratio of L1 and L2 variability:
  - $ICC = \frac{\sigma^2_{r_{0i}}}{\sigma^2_{r_{0i}} + \sigma^2_e}$
  - Proportion of variance due to differences between people

3.5.2 Variance explained and variance reduction

- When you compare your model to the **unconditional model**
  - Variance *explained*
  - How much variance does my model explain?
  - Like R^2
- When you compare your model to **some other (simpler) model**
  - Variance *reduction*
  - How much is (error) variance reduced by adding whatever you added?
  - Like R^2_{change}

3.5.3 Variance explained and variance reduction

- Model 1 is simpler, Model 2 is more complex
  - The model you "care about" is Model 2
- Reduction in variance =

$$\frac{\sigma_e^2(Model1) - \sigma_e^2(Model2)}{\sigma_e^2(Model1)}$$

3.5.4 Missing data

- Missing on outcome: OK (assuming MAR)
  - Uses all observations for a person to create trajectories
- Missing on predictor: Case is dropped
  - Make sure no missing or use multiple imputation

3.5.5 Extensions of mixed models

- Change in multiple variables at once
  - Baldwin et al. (2014): Complicated but possible
- Nonnormal outcomes
  - More difficult in unexpected ways when outcomes are non-normal

- SEM framework (Latent growth models)
  - Growth as a predictor, simultaneous growth of multiple processes, and other more complex models

4 Summary

4.1 Summary

4.1.1 Summary of this week

- **Mixed models** as an approach to repeated measures
  - Focus on *individual* change
  - Fewer problems with *missing data*
  - Continuous and unevenly spaced *time*
  - Flexible with *predictors* (continuous and categorical)
- **Way more complex & interesting than we have time to talk about!**
  - Adding predictors, shape of change, multiple outcomes
  - Take Longitudinal or Multilevel models or Categorical or SEM

4.1.2 Summary: RM ANOVA vs mixed models

- RM ANOVA focuses on **group level differences** in means at each time point
  - Only uses *complete cases* on the outcome
  - *Categorical* predictors only
- LMM focuses on **individual trajectories** over time
  - Uses *all observations available* on the outcome
  - *Continuous* or categorical predictors
- In general, I would always use some version of a mixed model