Categorical: Repeated measures

Table of contents

1 Goals

1.1 Goals

1.1.1 Goals of this lecture

- Some additional crucial details

- Estimation
- Model comparisons
- Adding and centering predictors

2 Review

2.1 Data

2.1.1 Schizophrenia over time

- Schizophrenia treatment effects over the course of 7 weeks ($N = 437$), measured by the Inpatient Multidimensional Psychiatric Scale (IMPS)
  - `id`: ID variable
  - `imps79`: Continuous measure of schizophrenia (1 to 7)
  - `imps79b`: Binary measure of schizophrenia (3.5+)
  - `imps79o`: Ordinal measure of schizophrenia (Cuts: 2.5+, 4.5+, 5.5+)
  - `tx`: Placebo (0) or treatment (1)
  - `week`: Week of study (0, 1, 3, 6)

2.1.2 Data

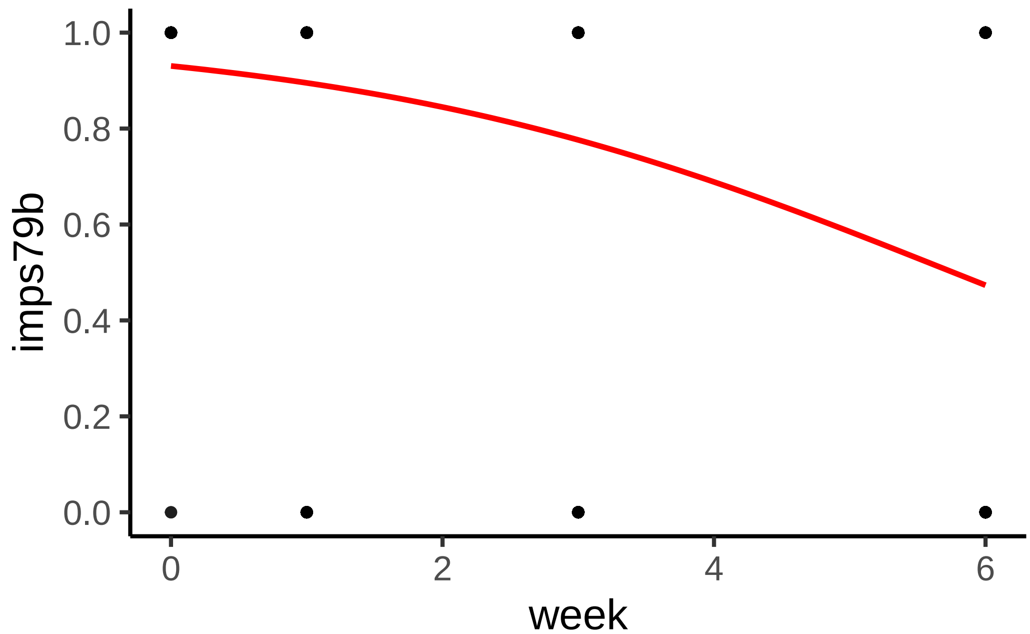
id	imps79	imps79b	imps79o	tx	week
1103	5.5	1	4	1	0
1103	3.0	0	2	1	1
1103	2.5	0	2	1	3
1103	4.0	1	2	1	6
1104	6.0	1	4	1	0
1104	3.0	0	2	1	1
1104	1.5	0	1	1	3
1104	2.5	0	2	1	6
1105	4.0	1	2	1	0
1105	3.0	0	2	1	1
1105	1.0	0	1	1	3
1105	NA	NA	NA	1	6

2.2 Marginal model

2.2.1 Marginal model

```
Call:
geeglm(formula = imps79b ~ 1 + week, family = binomial("logit"),
    data = schizx1, id = schizx1$id, corstr = "unstructured")

 Coefficients:
            Estimate  Std.err  Wald            Pr(>|W|)
(Intercept)  2.59459  0.11876 477.3 <0.0000000000000002 ***
week        -0.45017  0.02767 264.7 <0.0000000000000002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation structure = unstructured
Estimated Scale Parameters:

            Estimate Std.err
(Intercept)   0.9646  0.1007
  Link = identity

Estimated Correlation Parameters:
          Estimate Std.err
alpha.1:2  0.05390 0.05343
alpha.1:3 -0.02855 0.03265
alpha.1:4 -0.01890 0.03342
alpha.2:3  0.56341 0.10114
alpha.2:4  0.15242 0.06617
alpha.3:4  0.51550 0.07964
Number of clusters:   437  Maximum cluster size: 4
```

2.2.2 $ln\left(\frac{\hat{p}}{1-\hat{p}}\right) = 2.59 - 0.45(week)$



2.2.3 Population-averaged effects

$$ln\left(\frac{\hat{p}}{1-\hat{p}}\right) = b_0 + b_1(week) = 2.59 - 0.45(week)$$

- Basically treats `week` as a non-repeated measures predictor
  - Does not link observations from the same person together
- Odds ratio: $e^{-0.45} = 0.64$
  - Each week, the *odds of diagnosis* (`imps79b`) is multiplied by 0.64

2.2.4 Population-averaged effects

$$ln\left(\frac{\hat{p}}{1-\hat{p}}\right) = b_0 + b_1(week) = 2.59 - 0.45(week)$$

- Basically treats `week` as a non-repeated measures predictor
  - Does not link observations from the same person together
- Predicted probabilities

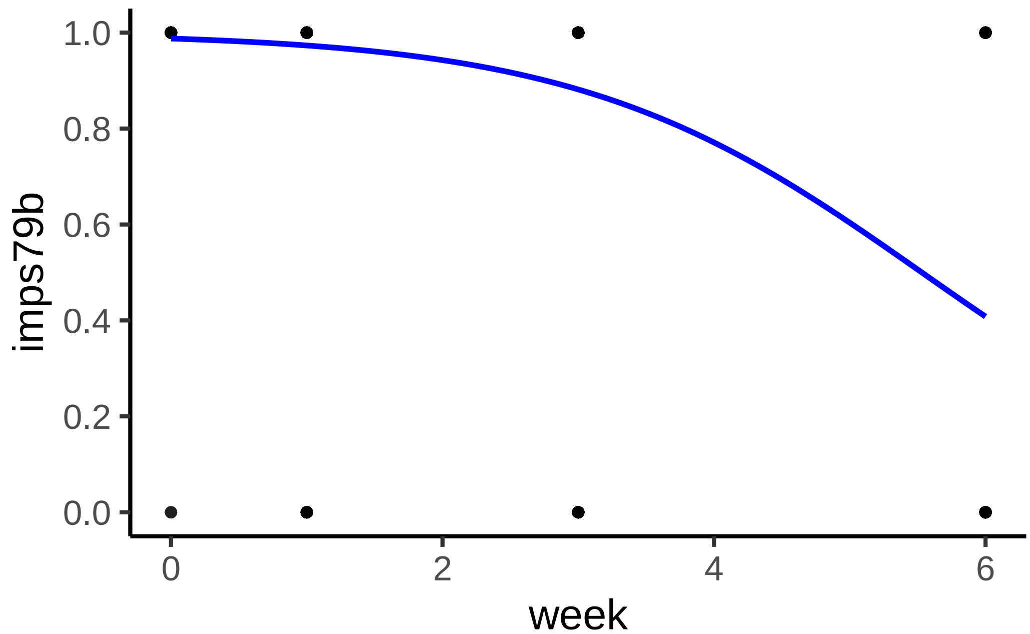
week	prob
0	0.93
1	0.89
3	0.78
6	0.47

2.3 Conditional model

2.3.1 Conditional model

```
Generalized linear mixed model fit by maximum likelihood (Laplace
  Approximation) [glmerMod]
 Family: binomial  ( logit )
Formula: imps79b ~ 1 + week + (1 + week | id)
   Data: schizx1

     AIC      BIC   logLik deviance df.resid
  1291.6   1318.4   -640.8   1281.6     1564

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.8446  0.0898  0.1139  0.2646  1.0822

Random effects:
 Groups Name        Variance Std.Dev. Corr
 id     (Intercept) 4.413    2.101
        week        0.711    0.843    -0.13
Number of obs: 1569, groups:  id, 437

Fixed effects:
            Estimate Std. Error z value            Pr(>|z|)
(Intercept)    4.386      0.539    8.14 0.00000000000000041 ***
week          -0.793      0.118   -6.71 0.00000000001954283 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
     (Intr)
week -0.817
```

2.3.2 $ln\left(\frac{\hat{p}}{1-\hat{p}}\right) = 4.39 - 0.79(week)$



2.3.3 Person-specific effects

$$ln\left(\frac{\hat{p}}{1-\hat{p}}\right) = b_0 + b_1(week) = 4.39 - 0.79(week)$$

- Models each person's trajectory *separately*
  - *Averages* intercepts to get average intercept, slopes to get average slope
- Odds ratio: $e^{-0.79} = 0.45$
  - Each week, the *odds of diagnosis* (`imps79b`) is multiplied by 0.45

2.3.4 Person-specific effects

$$ln\left(\frac{\hat{p}}{1-\hat{p}}\right) = b_0 + b_1(week) = 4.39 - 0.79(week)$$

- Models each person's trajectory *separately*
  - *Averages* intercepts to get average intercept, slopes to get average slope
- Predicted probabilities

week	prob
0	0.99
1	0.97
3	0.88
6	0.41

2.4 Comparison

2.4.1 Comparison

- Marginal model ignores individual variability
  - Only cares about correlations among repeated measures
- Remember the more complex conditional models we looked at last time
  - We couldn't include a **random slope** in the model with `tx` and `week`
  - Implies that people don't really vary in their slopes over time
- The models are quite similar
  - So maybe, in this case, it doesn't matter much to ignore the individual

3 Estimation

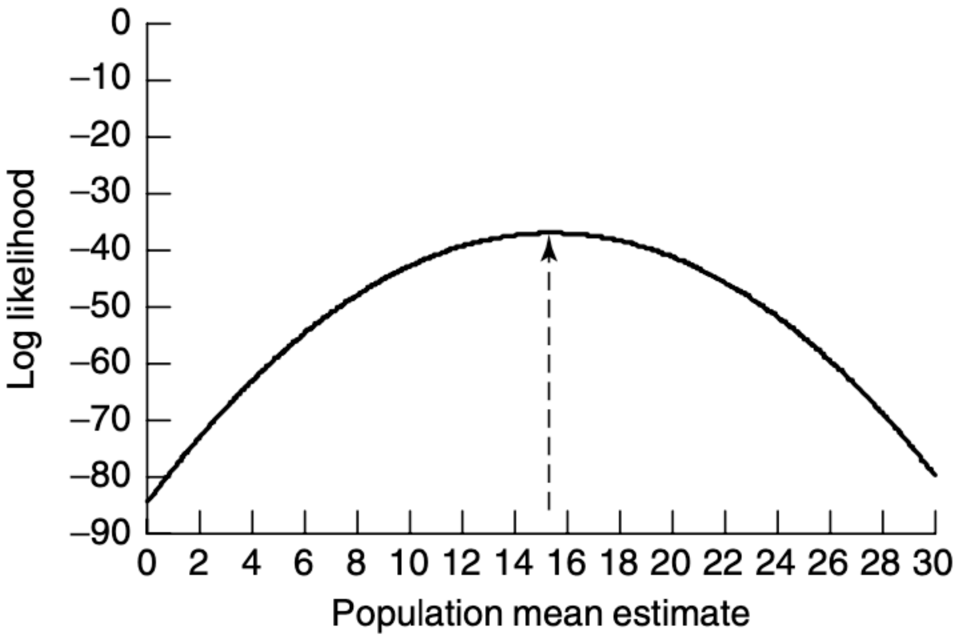
3.1 Approaches

3.1.1 Review: Maximum likelihood estimation (MLE)

- Models (linear, logistic regression) have a **likelihood function** that gives the *likelihood* of different parameter estimates
  - Likelihood ≈ probability
- "Maximum likelihood estimates" are the **parameter estimates** (e.g., regression coefficients, etc) that are **most likely** given the data
  - Uses *calculus* (derivatives) to find it
- Traditional MLE requires *joint distributions* for the model
  - Which we don't always have for categorical outcome models

3.1.2 Figure: Maximum likelihood estimation

- $(n+1)$-dimensional mountain
  - where n is the # of parameters you're estimating
- Peak of the mountain is the **maximum likelihood estimate**
- Right: Figure 2 from Enders, C. K. (2005). Maximum likelihood estimation. *Encyclopedia of statistics in behavioral science.*



The maximum of the log likelihood function is found at $\mu = 15.4$

3.1.3 Estimation approaches

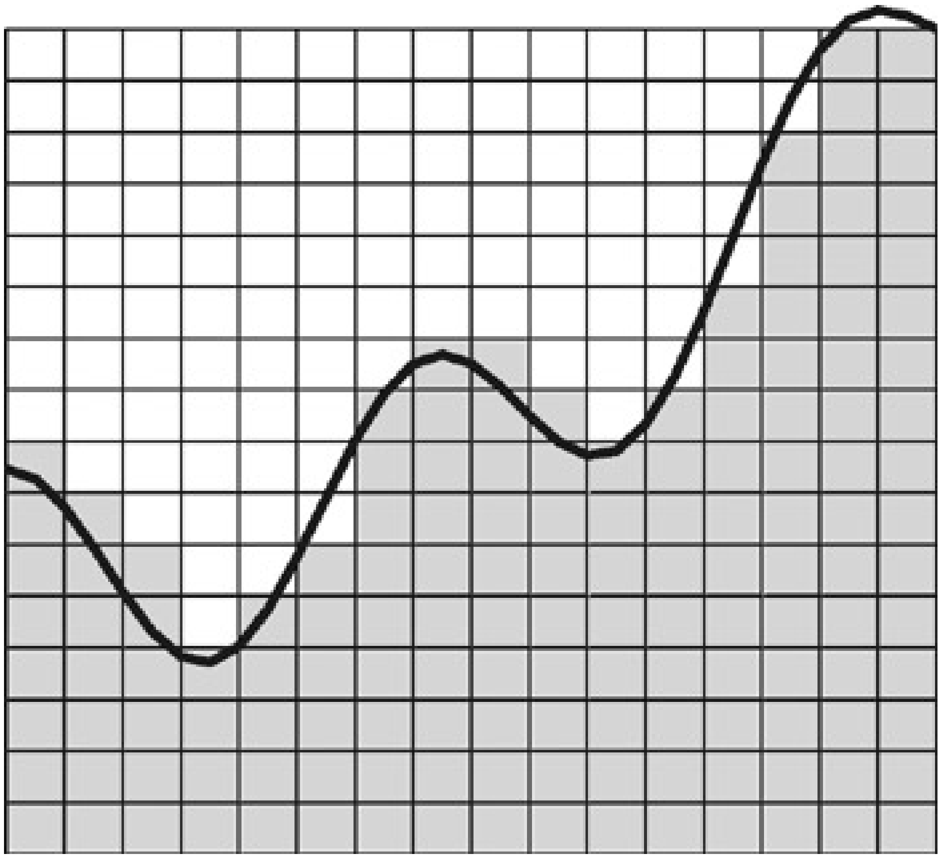
- **Marginal models**
  - Generalized estimating equations (GEE)
- **Conditional models**
  - Likelihood approximation
    - Integral approximation, numerical integration, adaptive quadrature
  - Linearization methods
    - Pseudo-likelihood, quasi-likelihood

3.1.4 Generalized estimating equations (GEE)

- Not likelihood functions
  - Type of quasi-likelihood, so no -2LL, AIC, LR tests, etc
- Only require *marginal* distributions
  - Not *joint* as for traditional MLE
- Technically GEE is the estimation method
  - But often used to refer to marginal models in general

3.1.5 Likelihood approximation

- Preferred method for accuracy
  - Estimates the likelihood
- Slow and computationally intensive
  - Doesn't always work
- "Integral approximation"
  - Find area under the curve



Picture came from here

3.1.6 Linearization methods

- Tries to turn the non-linear problem into a linear problem
  - Uses Taylor series expansion
- Pseudo- or quasi-likelihood approach
  - No -2LL, AIC, LR tests, etc
- Only option with SPSS `genlinmixed`
  - Default with SAS `glimmix` (but likelihood approximation too)

3.1.7 GLMM estimation in R

- `glmer()` function in **lme4** package

- Default: Laplace approximation
- Option: Adaptive Gauss-Hermite quadrature

- Other R packages that can run GLMMs have other options

4 Model comparisons

4.1 Model comparisons

4.1.1 Model comparisons

- Whether and how you can compare models depends on
  - Which models they are
    - Conditional vs marginal
  - How they were estimated
    - Likelihood method or not

4.1.2 Marginal models

- Estimated using generalized estimating equations
  - A quasi-likelihood approach
    - No LL: No AIC, no LR tests
  - QIC is a "quasi" information criteria
    - Can be used to compare nested or un-nested models
    - Similar to AIC

4.1.3 Conditional models

- Estimated with a **linearization** method (not preferred)
  - A quasi-likelihood approach
    - No LL: No AIC, no LR tests
    - Can use QIC (if available) similarly to AIC
- Estimated with **likelihood approximation** methods (preferred)
  - You get a log-likelihood and everything that comes with it: AIC, LR tests
  - Compare nested and non-nested models as usual

5 Predictors

5.1 Predictors

5.1.1 Predictors

- Marginal models
  - Predictors are predictors
  - Everything is at one level
- Conditional models
  - *Predictors can be at level 1 or level 2*
    - Longitudinal: Level 1 = observation, level 2 = person
    - Cross-sectional: Level 1 = person, level 2 = class, company, etc.
  - Entered into different parts of the model

5.1.2 Predictors in the model

- Two predictors: `week` (L1: Observation) and `tx` (L2: Person)
- Level 1: Within-person equation
  - $\eta_i = \pi_{0i} + \pi_{1i}(week_{ij}) + e_{ij}$
- Level 2: Between-person equation
  - $\pi_{0i} = \beta_{00} + \beta_{01}(tx_i) + r_{0i}$
  - $\pi_{1i} = \beta_{10} + \beta_{11}(tx_i) + r_{1i}$

5.1.3 Why do we care?

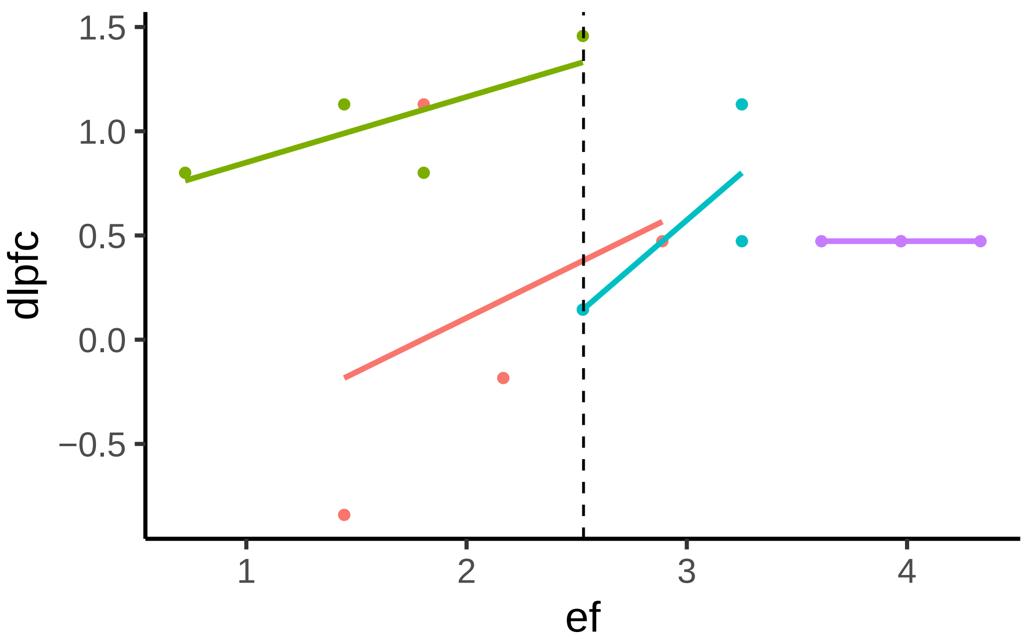
- Level 1 observations have both level 1 and level 2 information
  - Longitudinal: Occasion and person
  - Cross-sectional: Person and class, company, neighborhood
- If you ask me one day if I'm depressed, that gives you information about
  - How depressed I am **that day** (occasion, L1)
  - How depressed I **generally** am (person, L2)
- How can we disentangle those two kinds of information?
  - Centering

5.2 Centering

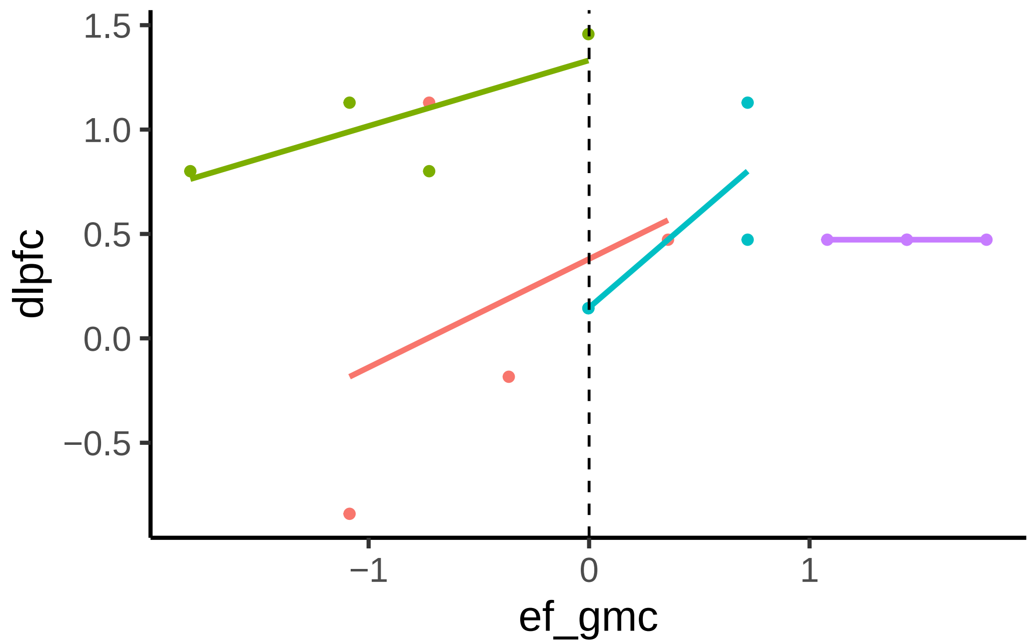
5.2.1 Centering in multi-level models

- 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

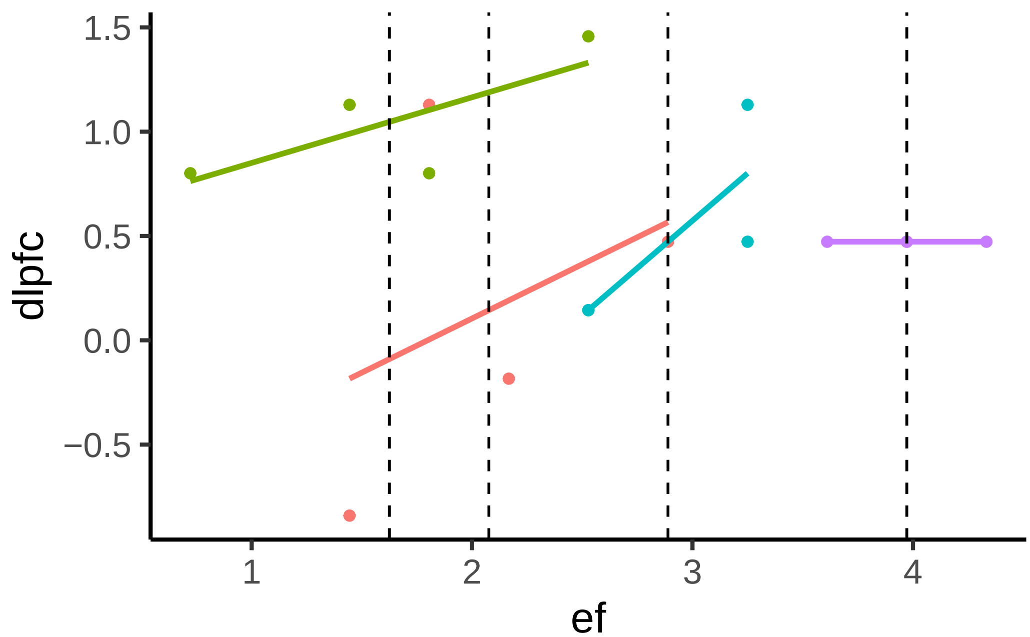
5.2.2 Figure: Uncentered with grand mean



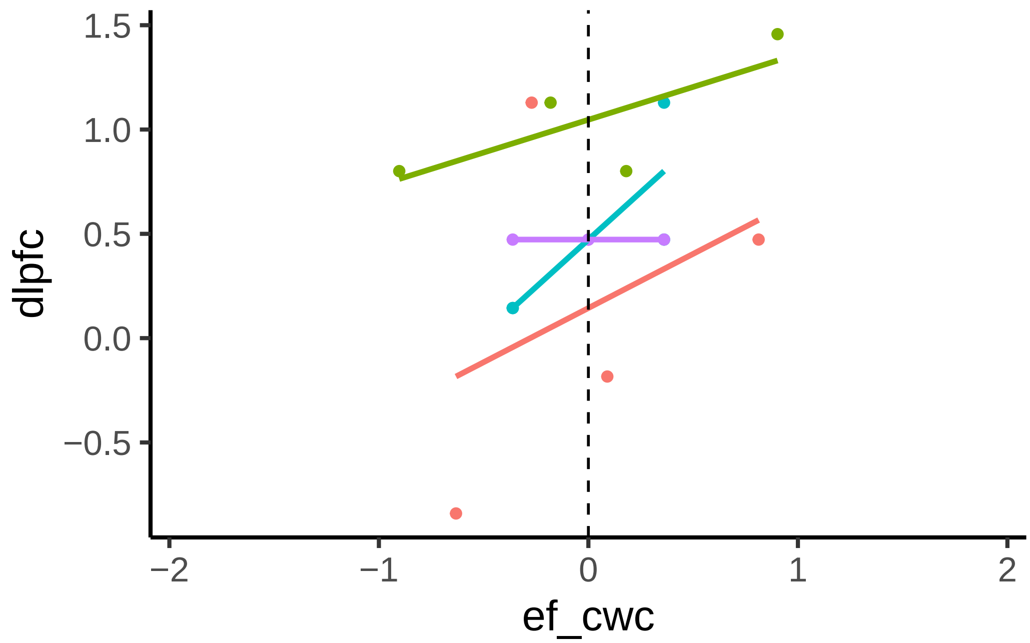
5.2.3 Figure: Grand mean centered



5.2.4 Figure: Uncentered with person (cluster) means



5.2.5 Figure: Centered within cluster



5.2.6 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 **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

5.2.7 Fully unconflated model

- Just centering doesn't **fully** unconflate level 1 and level 2
- When you have predictors at level 1 and you **center within cluster**
  - Removed the cluster-level means: L1 and L2 are still conflated
- What to do?
  - Add cluster mean of level 1 predictor back as a level 2 predictor
- Less commonly done for longitudinal
  - More common for cross-sectional

5.2.8 Fully unconflated model

- Two predictors: `week` (L1: Observation) and `tx` (L2: Person)
  - Add L1 predictor (`L1pred`), which is centered within cluster (person)
- Level 1: Within-person equation
  - $\eta_i = \pi_{0i} + \pi_{1i}(week_{ij}) + \pi_{2i}(L1pred_{ij} - \overline{L1pred}_i) + e_{ij}$
- Level 2: Between-person equation
  - $\pi_{0i} = \beta_{00} + \beta_{01}(tx_i) + \beta_{02}(\overline{L1pred}_i) + r_{0i}$
  - $\pi_{1i} = \beta_{10} + \beta_{11}(tx_i)$
  - $\pi_{2i} = \beta_{20} + \beta_{21}(tx_i)$

5.2.9 Centering predictors: Some references

- Curran, P. J., & Bauer, D. J. (2011). The disaggregation of within-person and between-person effects in longitudinal models of change. *Annual review of psychology, 62*, 583–619.
- 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.
- 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.
- Hayes, T. B. (under review). Individual-Level Probabilities and Cluster-Level Proportions: Toward Interpretable Level- 2 Estimates in Unconflated Multilevel Models for Binary and Ordinal Outcomes.
- 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.
- 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.
- 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.
- Yaremych, H. E., Preacher, K. J., & Hedeker, D. (2021). Centering categorical predictors in multilevel models: Best practices and interpretation. *Psychological Methods.*

6 Summary

6.1 Summary

6.1.1 Summary of this week

- Reviewed marginal and conditional models
  - Different interpretations, different numbers
- Estimation
- Model comparison
- Predictors and centering

6.1.2 Summary of this section

- Repeated measures models for categorical outcomes
  - Marginal: **R** matrix, population averaged, GEE, cluster robust
  - Conditional: **G** matrix, cluster-specific, generalized linear mixed models (GLMM)
- Additional complexities: marginal and conditional are not the same, estimation is more difficult, model comparison is more difficult

6.1.3 Next weeks

- Next week: No class, but work on final project
  - Sign up for a meeting with me if you want to chat about anything
  - Email me if you have any questions about anything
  - Last article discussion (4/9) and homework 4 (4/16)
- 2 weeks from now: No class
  - Record presentations (4/23)
  - Watch presentations (4/26)
  - Comment on presentations (4/26)
  - Final paper (4/28)