A New Approach for Detecting Differential Item Functioning in IRT Models

Carolin Strobl

LMU

CEST
Center for Empirical Studies

Outline

- Example I: differential item functioning (DIF) in a Rasch model
- Examples II: different worth parameters in a Bradley-Terry model
- Statistical/computational framework
- Outlook and open questions

Example I:

data from the SPIEGEL "Students-PISA" survey

- open-access online survey on general education
- each partcipant was randomly assigned one of 24 questionnaires, consisting of 45 items from 6 topics: politics, history, economics, culture and natural sciences
- questions were either multiple-choice or open
- recorded response: correct/wrong

results presented here are for one exemplary questionnaire,

N = 30 188

Example I:

example questions

- politics: "What is the capital of Rheinland-Pfalz?"
- history: "In which century did the Thirty Years' War take place?"
- economics: "Which internet-company took over the media-group Time Warner?"
- culture: "Which city is the setting for the novel 'Buddenbrooks'?"
- natural sciences: "Which sensory cells in the human eye are responsible for color vision?"

Example I:

example questions

- politics: "What is the capital of Rheinland-Pfalz?" Mainz
- history: "In which century did the Thirty Years' War take place?"
- economics: "Which internet-company took over the media-group Time Warner?"
- culture: "Which city is the setting for the novel 'Buddenbrooks'?"
- natural sciences: "Which sensory cells in the human eye are responsible for color vision?"

Example I:

example questions

- politics: "What is the capital of Rheinland-Pfalz?" Mainz
- history: "In which century did the Thirty Years' War take place?" 17th
- economics: "Which internet-company took over the media-group Time Warner?"
- culture: "Which city is the setting for the novel 'Buddenbrooks'?"
- natural sciences: "Which sensory cells in the human eye are responsible for color vision?"

Example I:

example questions

- politics: "What is the capital of Rheinland-Pfalz?" Mainz
- history: "In which century did the Thirty Years' War take place?" 17th
- economics: "Which internet-company took over the media-group Time Warner?" AOL
- culture: "Which city is the setting for the novel 'Buddenbrooks'?"
- natural sciences: "Which sensory cells in the human eye are responsible for color vision?"

Example I:

example questions

- politics: "What is the capital of Rheinland-Pfalz?" Mainz
- history: "In which century did the Thirty Years' War take place?" 17th
- economics: "Which internet-company took over the media-group Time Warner?" AOL
- culture: "Which city is the setting for the novel 'Buddenbrooks'?" Lübeck
- natural sciences: "Which sensory cells in the human eye are responsible for color vision?"

Example I:

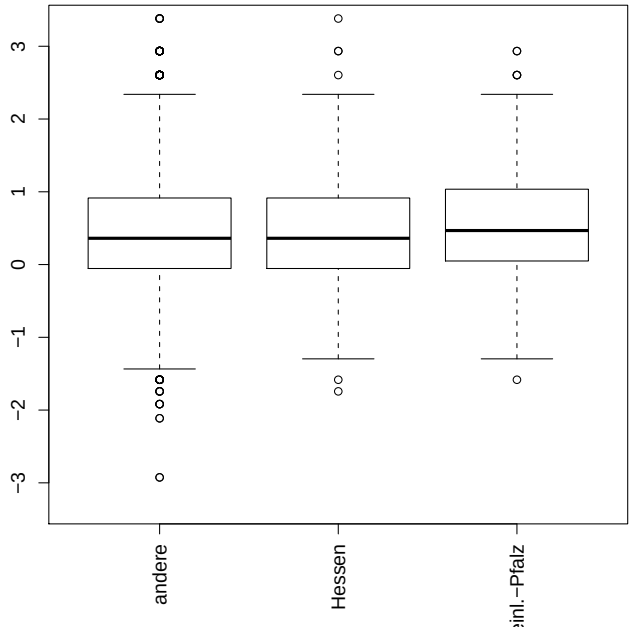
example questions

- politics: "What is the capital of Rheinland-Pfalz?" Mainz
- history: "In which century did the Thirty Years' War take place?" 17th
- economics: "Which internet-company took over the media-group Time Warner?" AOL
- culture: "Which city is the setting for the novel 'Buddenbrooks'?" Lübeck
- natural sciences: "Which sensory cells in the human eye are responsible for color vision?" cones

Example I:

curious finding:

those participants who received their Abitur in Rheinland-Pfalz perform significantly better in the test

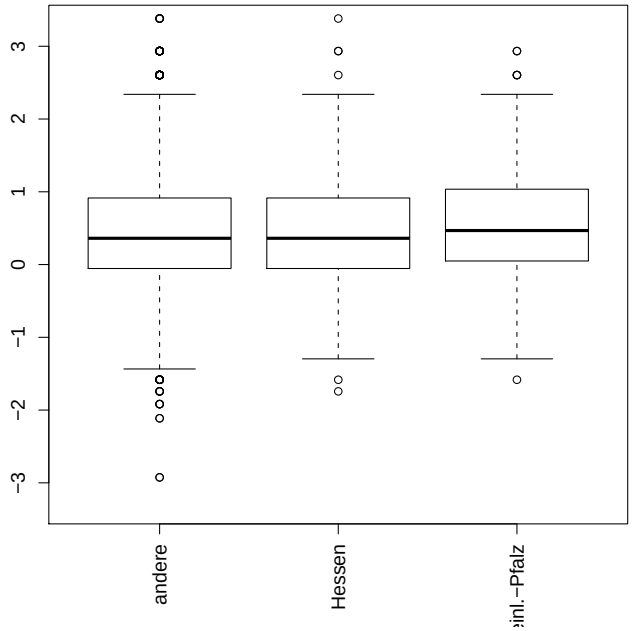


possible explanations:

Example I:

curious finding:

those participants who received their Abitur in Rheinland-Pfalz perform significantly better in the test



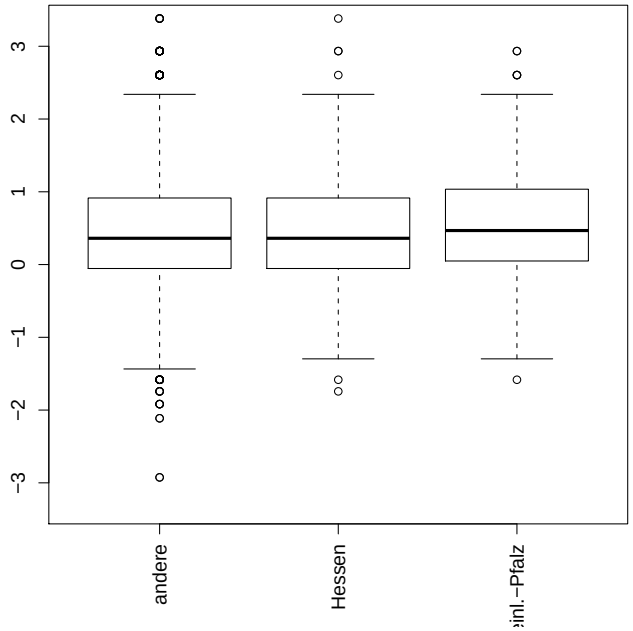
possible explanations:

- they are just smarter

Example I:

curious finding:

those participants who received their Abitur in Rheinland-Pfalz perform significantly better in the test



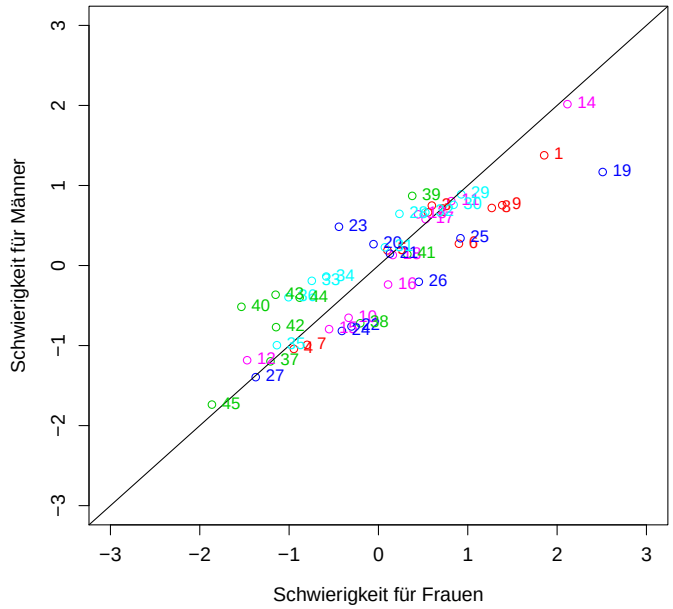
possible explanations:

- they are just smarter
- they have an unfair advantage ⇒ DIF

Example I:

ways to detect DIF in the Rasch Model:

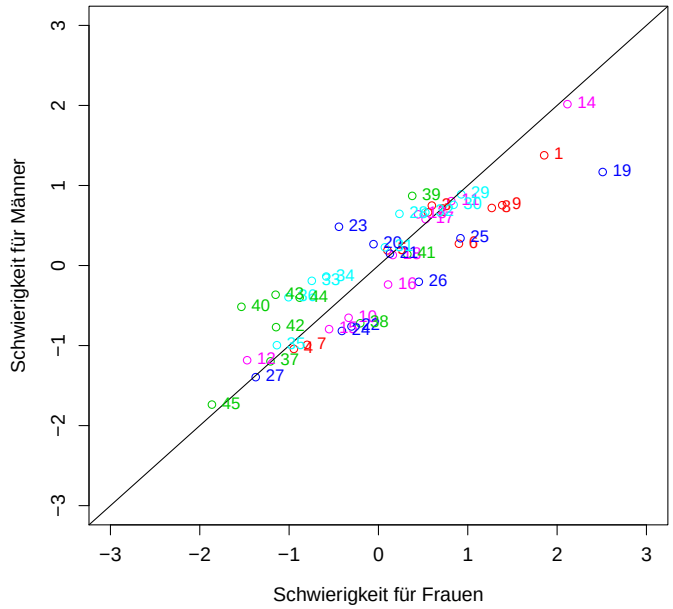
- graphical test (for two given groups)



Example I:

ways to detect DIF in the Rasch Model:

- graphical test (for two given groups)

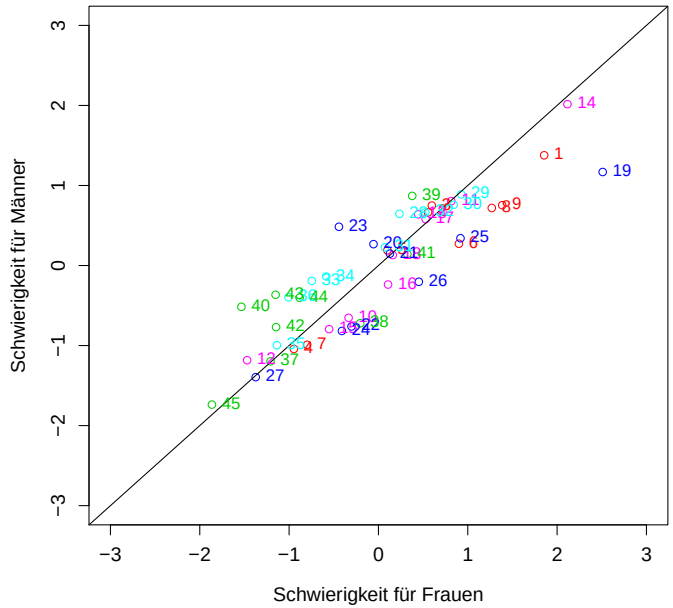


- LR-test (for k given groups)
- ...

Example I:

ways to detect DIF in the Rasch Model:

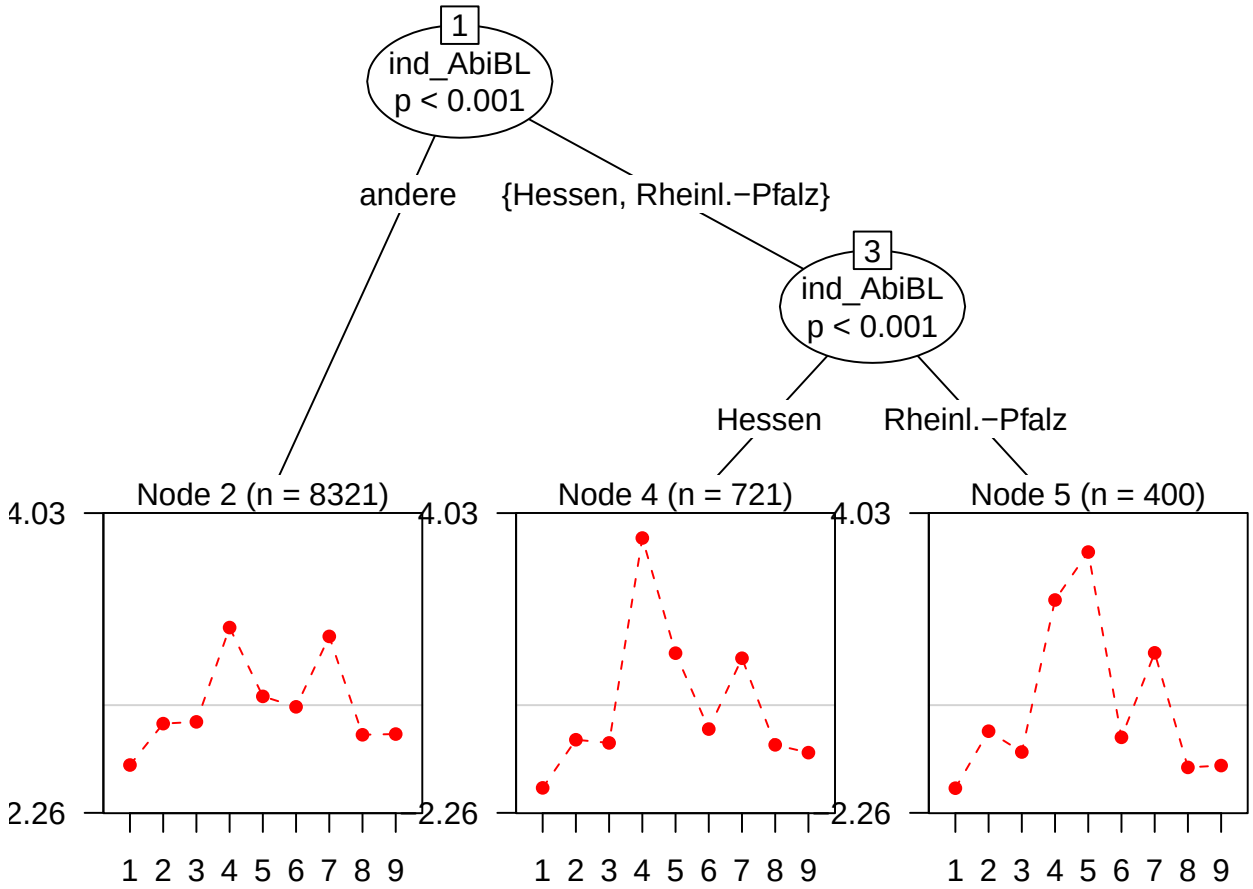
- graphical test (for two given groups)



- LR-test (for k given groups)
- ...
- our way: model-based recursive partitioning

Example I:

looking only at the politics items and the covariate ind_AbiBL



Nr. 4: Where is Hessen? (indicate location on a map)

Nr. 5: What is the capital of Rheinland-Pfalz?

Example I:

psychological impact of DIF

- test is no longer specifically objective
- fair comparisons between the groups are impossible

$\Rightarrow$ eliminate DIF-items from the test

(ideally in the pretest-phase)

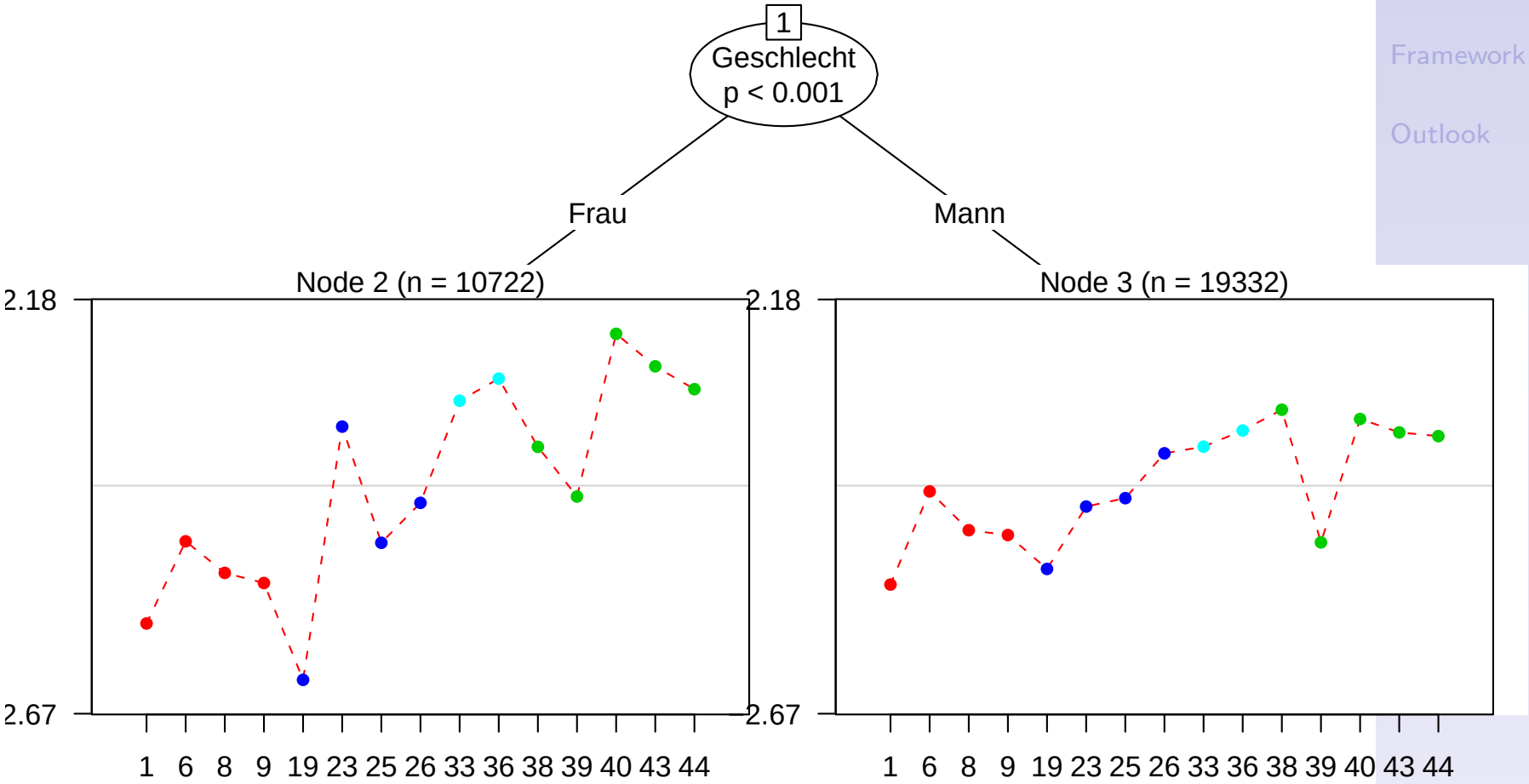
in our example:

eliminating items 4 and 5 eliminates group differences

Example I:

the 15 most differential items for the covariate Gender

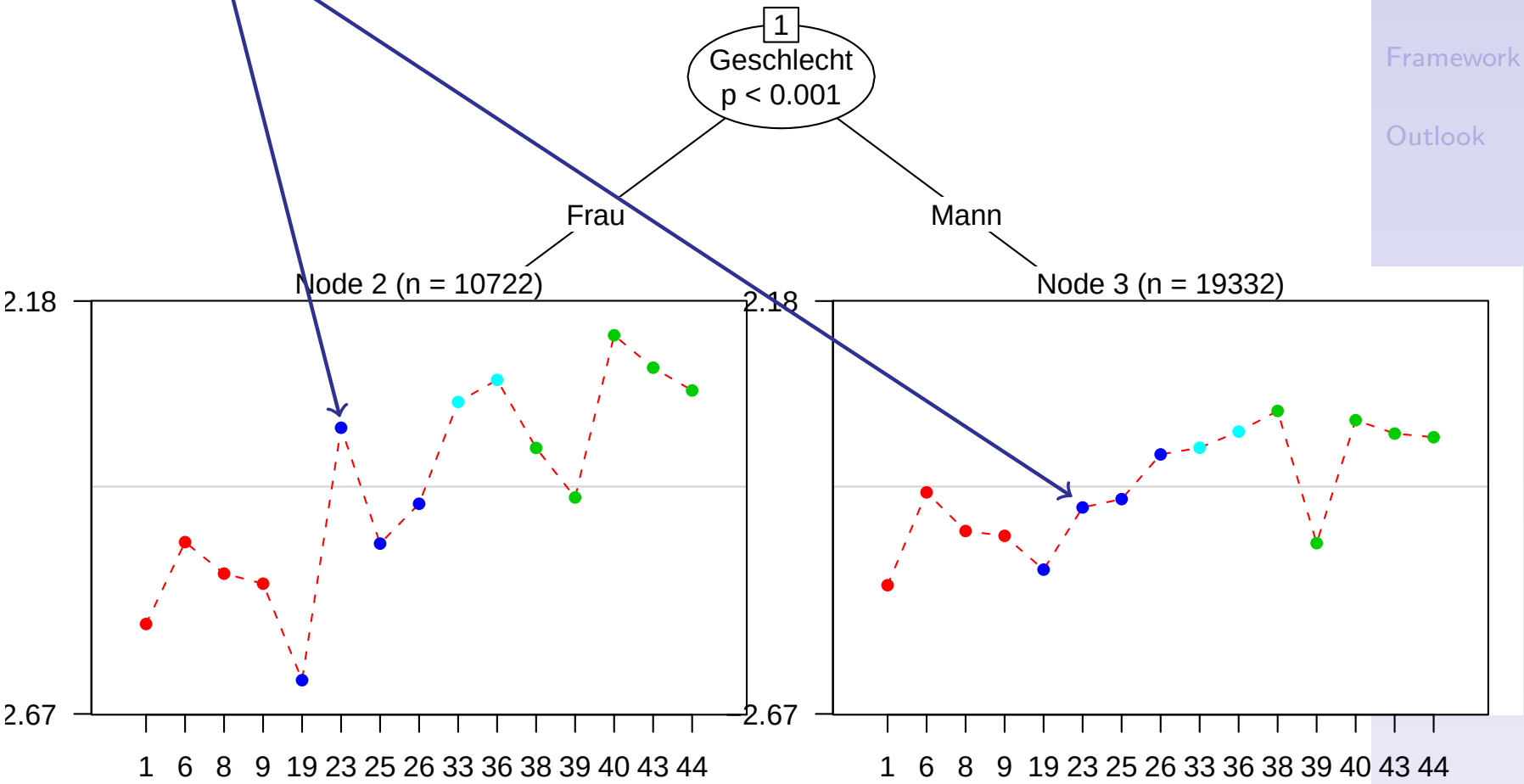
Nr. 23: bio logo – Nr. 36: Mozart opera – Nr. 38: ultrasound



Example I:

the 15 most differential items for the covariate Gender

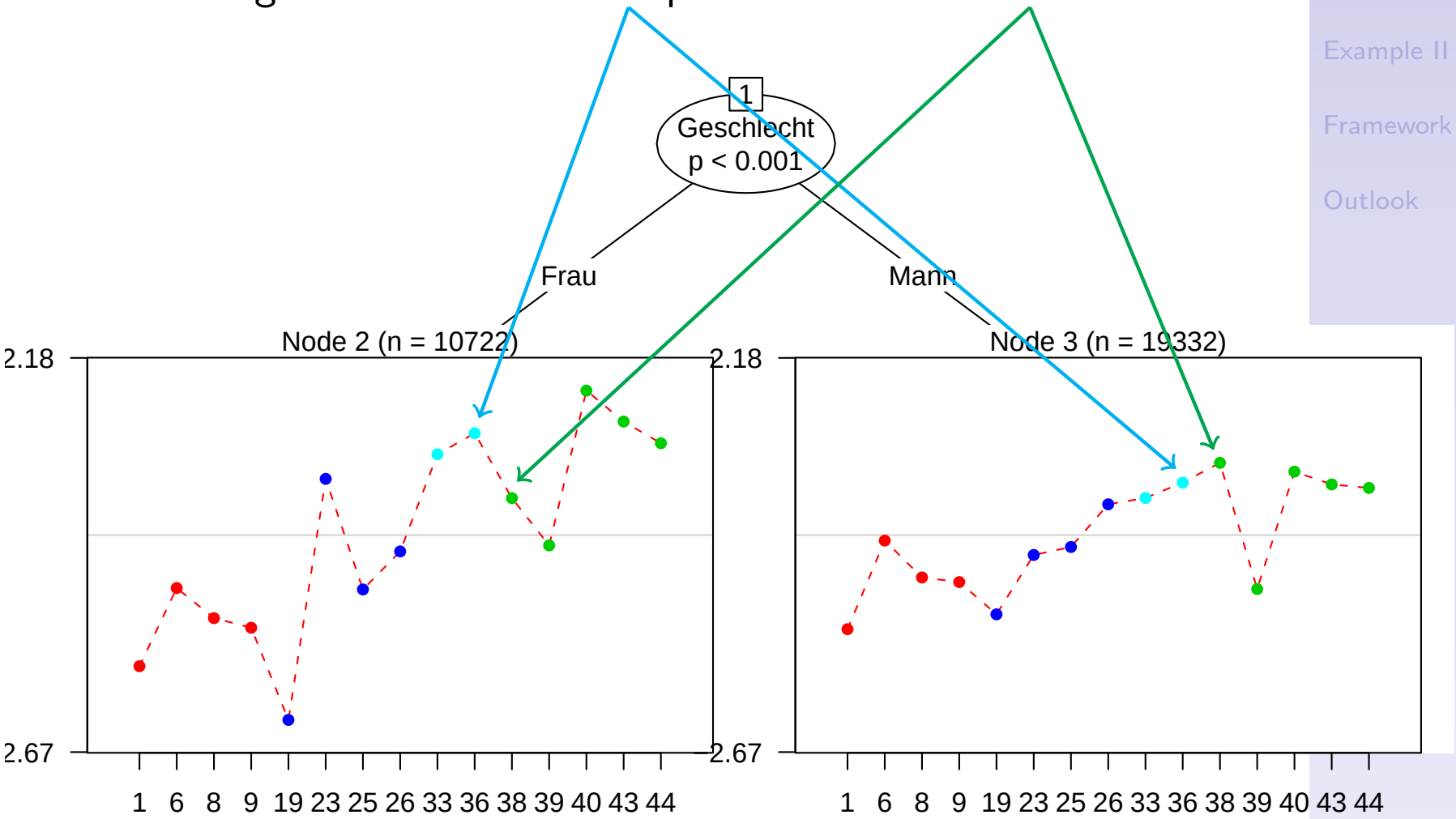
Nr. 23: bio logo – Nr. 36: Mozart opera – Nr. 38: ultrasound



Example I:

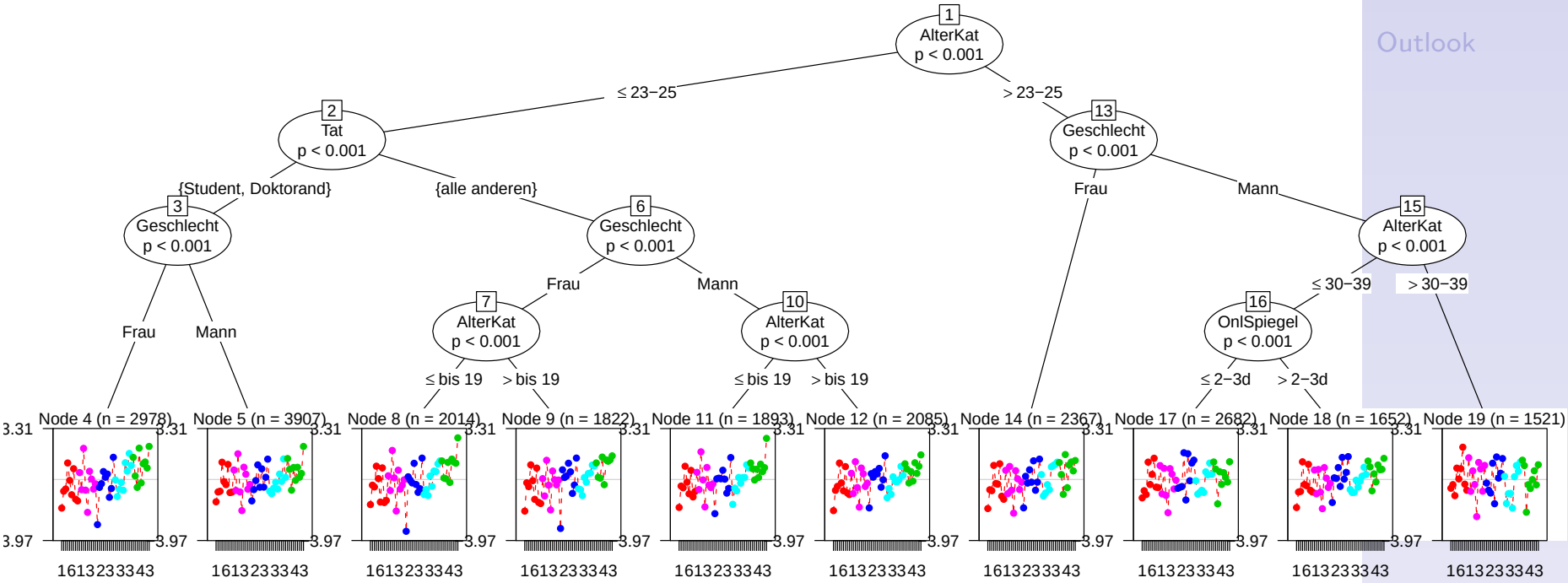
the 15 most differential items for the covariate Gender

Nr. 23: bio logo – Nr. 36: Mozart opera – Nr. 38: ultrasound



Example I:

all items and all covariates

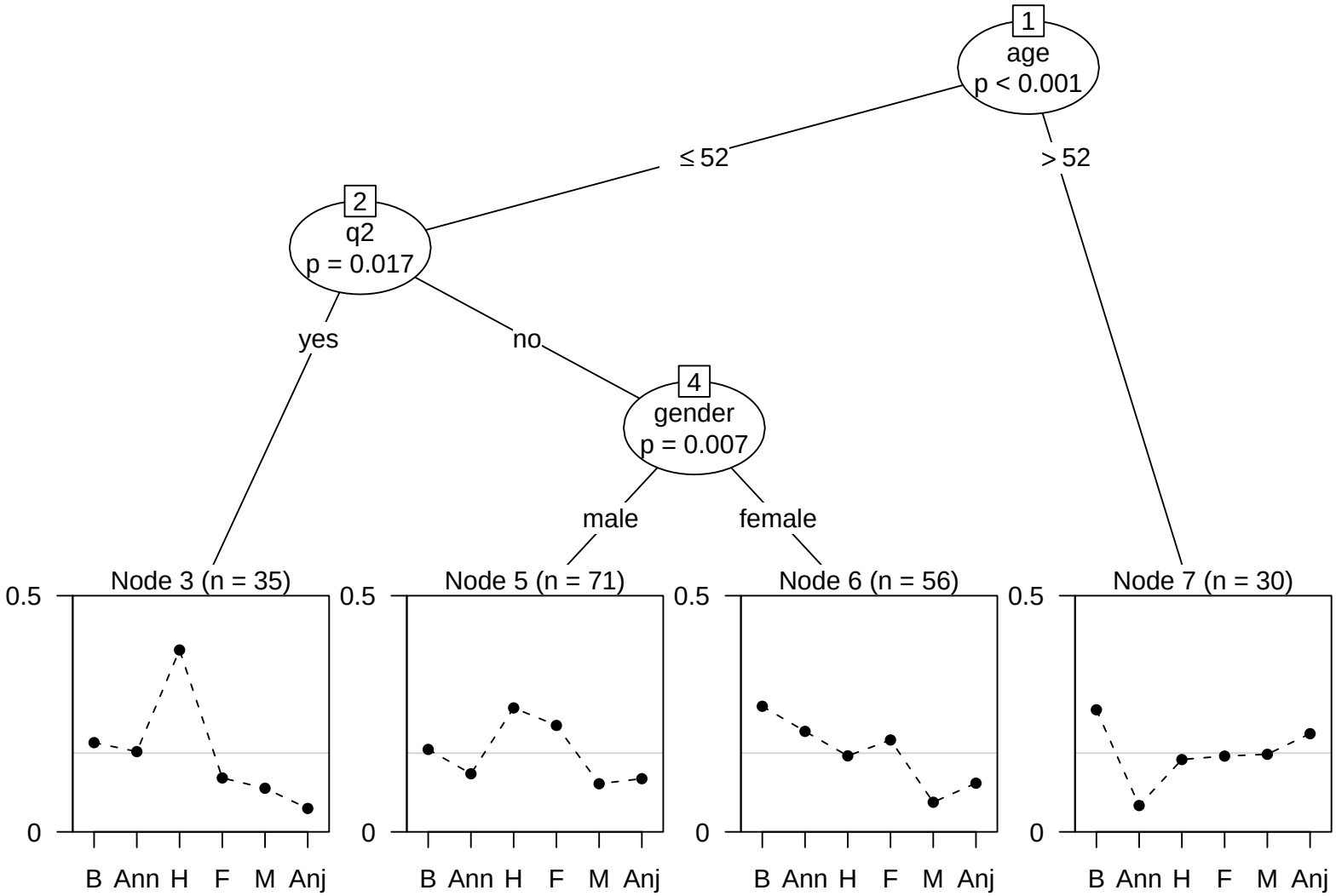


Example II:

- N = 192 subjects (friends and family of psychology students from Tübingen)
- paired comparisons of 6 candidates of the TV-show "Germany's Next Topmodel" 2007
- covariates: age, gender, regularly watched the show etc.

Example II:

Example II:


1
age
p < 0.001
≤ 52
> 52
2
q2
p = 0.017
yes
no
4
gender
p = 0.007
male
female
Node 3 (n = 35)
Node 5 (n = 71)
Node 6 (n = 56)
Node 7 (n = 30)
0.5
0
B Ann H F M Anj

Statistical/computational framework

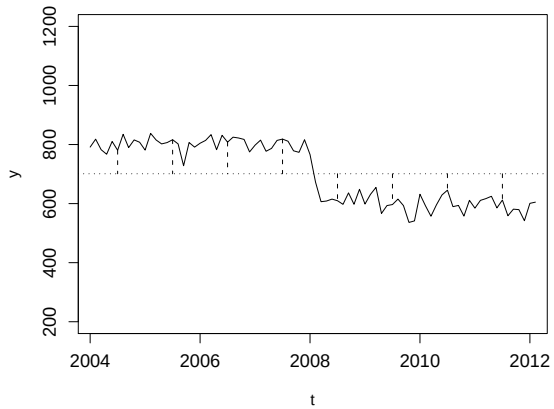
model-based recursive partitioning:

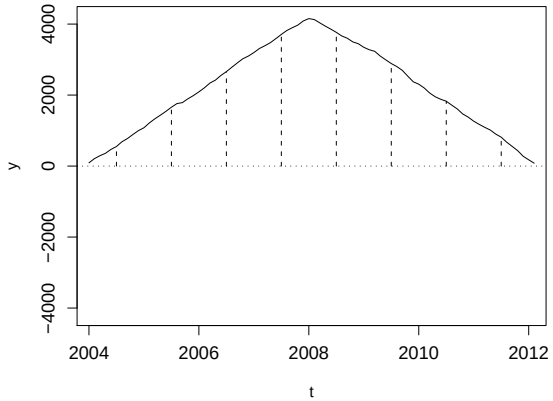
1. fit joint model
2. test for instability in model parameters over all covariates
3. split sample in the covariate and cutpoint inducing the strongest parameter instability
4. repeat steps 1–3 recursively until some stopping criterion is met

Tests for parameter instability

- individual contributions to the score-funktion

$$\psi(y_i, \theta) = \frac{\partial \Psi(y_i, \theta)}{\partial \theta}$$





- cumulated over all values of the covariate ℓ

$$W_\ell(t) = \hat{V}^{-1/2} n^{-1/2} \sum_{i=1}^{\lfloor n \cdot t \rfloor} \psi(y_{(i|\ell)}, \hat{\theta})$$

- under H_0 the path $W_\ell(t)$ randomly fluctuates around zero ($\rightarrow$ Brownian bridge)

(Zeileis und Hornik, 2007, *Statistica Neerlandica*)

Tests for parameter instability

test statistics

- for continuous covariates:

$$S_\ell = \max_{i=\underline{i},\ldots,\bar{i}} \left(\frac{i}{n} \cdot \frac{n-i}{n}\right)^{-1} \left\| W_\ell\left(\frac{i}{n}\right) \right\|_2^2$$

- for categorical covariates:

$$S_\ell = \sum_{q=1}^{Q} n \left(\sum_{i=1}^{n} I(x_{i\ell} = q)\right)^{-1} \left\| \Delta_q W_\ell\left(\frac{i}{n}\right) \right\|_2^2$$

with known distributions (Zeileis, Hothorn und Hornik, 2008, *Journal of Computational and Graphical Statistics*)

Statistical/computational framework

so it all depends on the individual contributions to the score-funktion $\psi(y_i, \theta)$

- for the Bradley-Terry model: closed form (Strobl, Wickelmaier und Zeileis, 2010, *Journal of Educational and Behavioral Statistics*)
- for the Rasch model: CML approach with Liou's algorithm for computing the derivatives of the symmetric functions (will start writing when back home...)

R-package `psychotree` on CRAN/R-Forge

(ask Achim for details, he has done all the work!)

Outlook and open questions

Outlook and open questions

- keep doing this for other IRT models
  - Partial Credit model
  - Birnbaum/2 and 3 PL models $\Rightarrow$ MML

    (esp. guessing parameters for multiple choice items)

Outlook and open questions

- keep doing this for other IRT models
  - Partial Credit model
  - Birnbaum/2 and 3 PL models $\Rightarrow$ MML

    (esp. guessing parameters for multiple choice items)

- "post-hoc tests" – which items have significant DIF?