universität
innsbruck

Examining Exams Using Rasch Models and Assessment of Measurement Invariance

Achim Zeileis

https://www.zeileis.org/

Prologue

Dedication: To the memory of Fritz Leisch.

Historical notes: About Fritz' impact on this work.

Mixed topics: Combination of methods, software, application – and of teaching and research.

Photo: DSC 2007, Whangārei, New Zealand.

Prologue: Structural change tests

Started: 2000.

R package: *strucchange*.

Key publication: Zeileis A, Leisch F, Hornik K, Kleiber C (2002). “strucchange: An R Package for Testing for Structural Change in Linear Regression Models.” *Journal of Statistical Software*, **7**(2), 1–38. doi:10.18637/jss.v007.i02

Photo: DSC 2001, Vienna, Austria.

Prologue: Recursive partitioning

Started: 2003.

R packages: *party*, *partykit*, . . .

Key publication: Zeileis A, Hothorn T, Hornik K (2008). "Model-Based Recursive Partitioning." *Journal of Computational and Graphical Statistics*, **17**(2), 492–514. doi:10.1198/106186008X319331

Photo: AASC 2004, Pecol, Italy.

Prologue: Psychometric computing

Started: 2008.

R packages: *psychotools*, *psychotree*, *psychomix*.

Key publication: Frick H, Strobl C, Leisch F, Zeileis A (2012). “Flexible Rasch Mixture Models with Package psychomix.” *Journal of Statistical Software*, **48**(7), 1–25. doi:10.18637/jss.v048.i07

Photo: useR! 2006, Vienna, Austria.

Prologue: R/exams

Started: 2007 (v1), 2012 (v2).

R package: *exams*.

Key publication: Zeileis A, Umlauf N, Leisch F (2014). “Flexible Generation of E-Learning Exams in R: Moodle Quizzes, OLAT Assessments, and Beyond.” *Journal of Statistical Software*, **58**(1), 1–36. doi:10.18637/jss.v058.i01

Photo: AASC 2009, Tragöß, Austria.

Large-scale exams

Motivation:

- Statisticians often teach large-scale courses for other fields.
- Multiple-choice exams typically evaluated and graded automatically.
- Little further examination of results (if any).

Potential questions:

- Ability of students.
- Difficulty of exercises (or items).
- Differential item functioning (DIF).
- Unidimensionality.
- Fairness.

Large-scale exams

Course: Mathematics for first-year business and economics students at Universität Innsbruck.

Format: Biweekly online tests (conducted in OpenOlat) and two written exams for 500 to 1,000 students per semester.

Here: Individual results from an end-term exam.

- 729 students.
- 13 single-choice items with five answer alternatives.
- Two groups with partially different item pools (on the same topics). Individual versions of items generated via *exams* in R.

Large-scale exams

Variables: In `MathExam14W`.

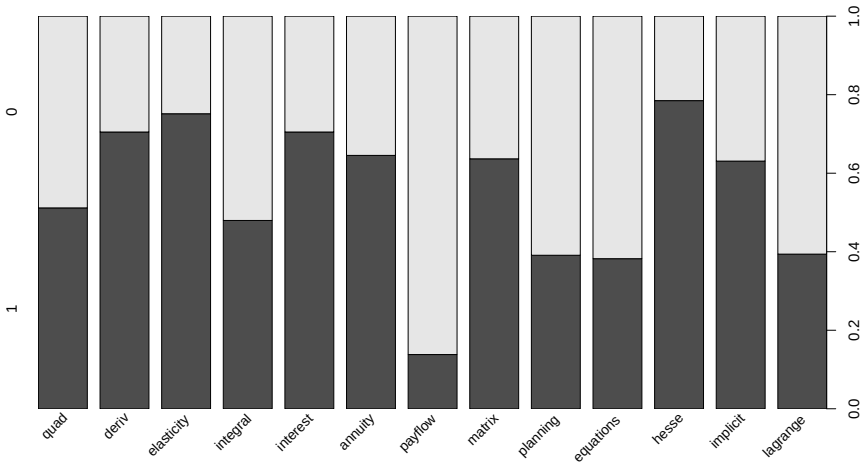
- `solved`: Item response matrix (1/0 coding).
- `group`: Factor for group (1 vs. 2).
- `tests`: Number of previous online exercises solved (out of 26).
- `nsolved`: Number of exam items solved (out of 13).
- `gender`, `study`, `attempt`, `semester`.

In R: Load package/data and exclude extreme scorers.

```
R> library("psychotools")
R> data("MathExam14W", package = "psychotools")
R> mex <- subset(MathExam14W, nsolved > 0 & nsolved < 13)
```

Large-scale exams

```
R> plot(mex$solved)
```



IRT with the Rasch model

Motivation: Item response theory (IRT) with Rasch model.

- Measure a single latent trait (here: ability in exam).
- Based on binary items $y_{ij} \in \{0, 1\}$ (here: solved correctly vs. not).
- Align person's ability θ_i ($i = 1, \ldots, n$) and item's difficulty β_j ($j = 1, \ldots, m$) on the same scale.

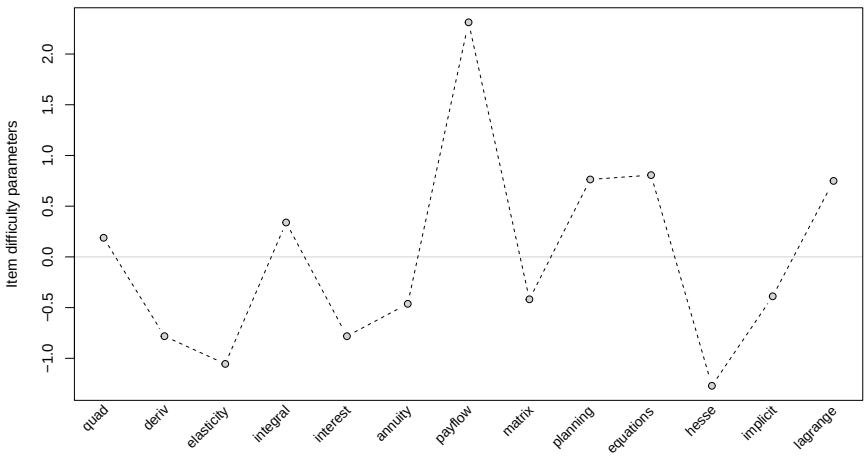
Model: Logistic model for probability that person *i* solves item *j*.

$$\begin{aligned} \pi_{ij} &= \Pr(y_{ij} = 1) \\ \text{logit}(\pi_{ij}) &= \theta_i - \beta_j \end{aligned}$$

- Consistent estimation via conditional maximum likelihood.
- Interval scale with arbitrary zero point.

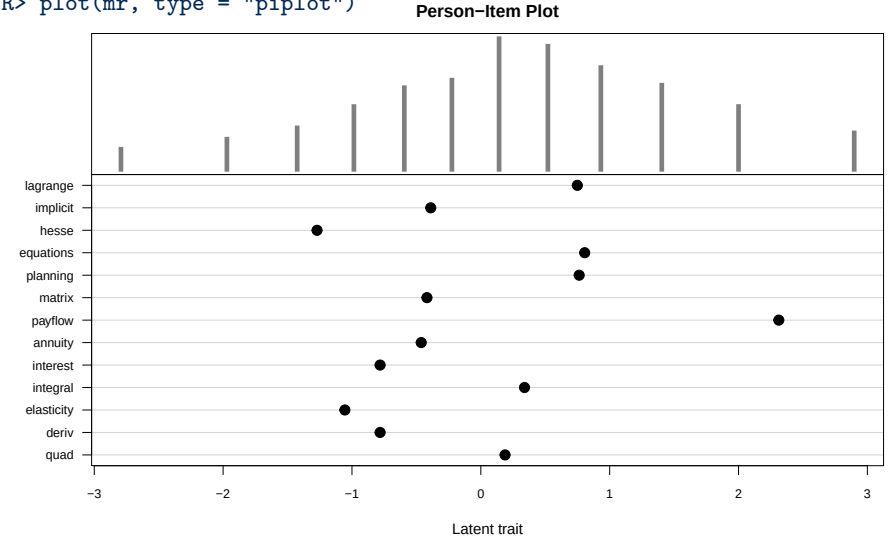
IRT with the Rasch model

```
R> mr <- raschmodel(mex$solved)
R> plot(mr, type = "profile")
```



IRT with the Rasch model

```
R> plot(mr, type = "piplot")
```



IRT with the Rasch model

Crucial assumption: Measurement invariance, corresponding to stability of item parameters across all possible subgroups.

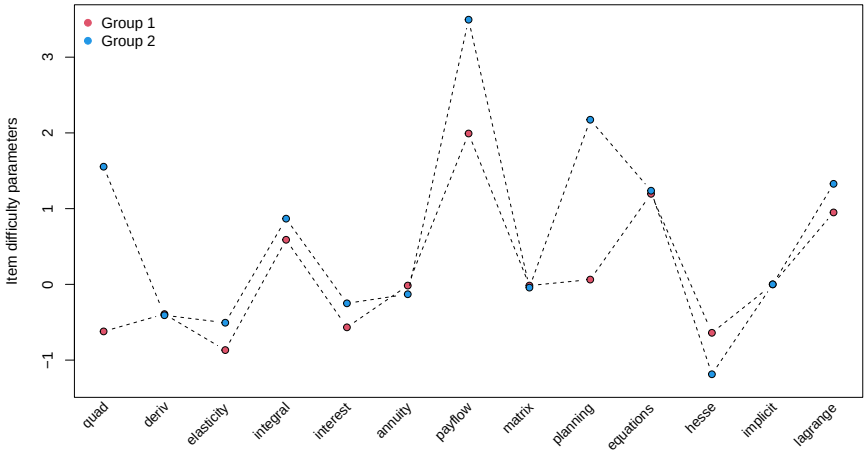
Assessment: Detect potential violations using covariates to form subgroups.

- Classical tests (likelihood ratio, Wald, score) for one binary covariate.
- Recursive partitioning along many covariates (continuous, ordinal, categorical).
- Finite mixture model without covariates.

Question: Are there differences in item difficulty between groups 1 and 2?

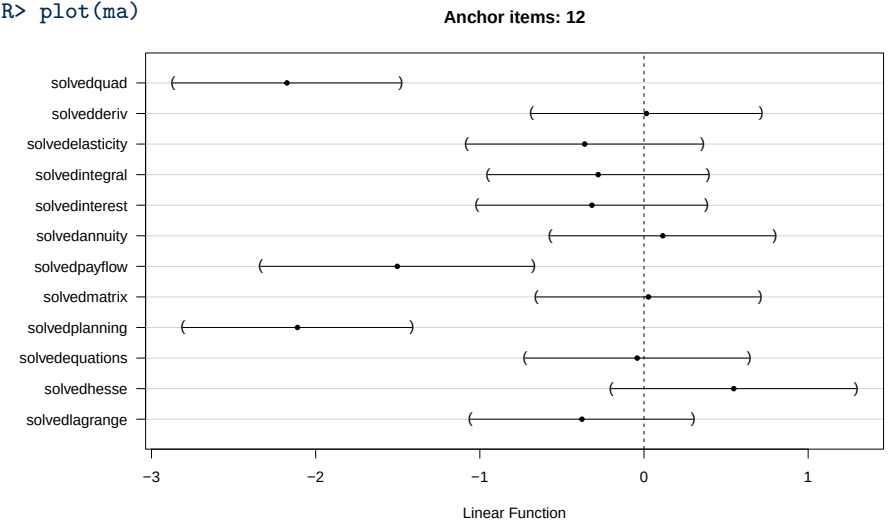
Classical tests

```
R> plot(mr1, parg = list(ref = 12), ...)
R> plot(mr2, parg = list(ref = 12), ...)
```



Classical tests

```
R> ma <- anchortest(solved ~ group, data = mex, adjust = "single-step")
R> plot(ma)
```



Rasch trees

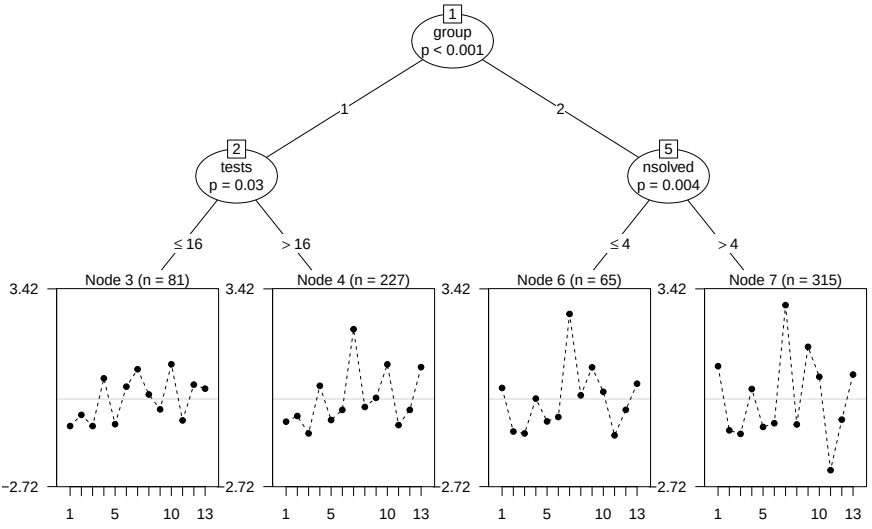
Questions:

- Are there further differences in the two exam groups?
- Especially with respect to mathematics ability (`tests` or `nsolved`)?

Here: Treat numeric variables with few levels as ordinal, simulate *p*-values.

```
R> library("psychotree")
R> mex <- transform(mex,
+    tests    = ordered(tests),
+    nsolved  = ordered(nsolved),
+    attempt  = ordered(attempt),
+    semester = ordered(semester)
+  )
R> mrt <- raschtree(solved ~ group + tests + nsolved + gender +
+    attempt + study + semester, data = mex,
+    vcov = "info", minsize = 50, ordinal = "L2", nrep = 1e5)
R> plot(mrt)
```

Rasch trees


1
group
p < 0.001
1
2
2
tests
p = 0.03
5
nsolved
p = 0.004
≤ 16
> 16
≤ 4
> 4
Node 3 (n = 81)
Node 4 (n = 227)
Node 6 (n = 65)
Node 7 (n = 315)
3.42
−2.72
1
5
10
13

Rasch mixture models

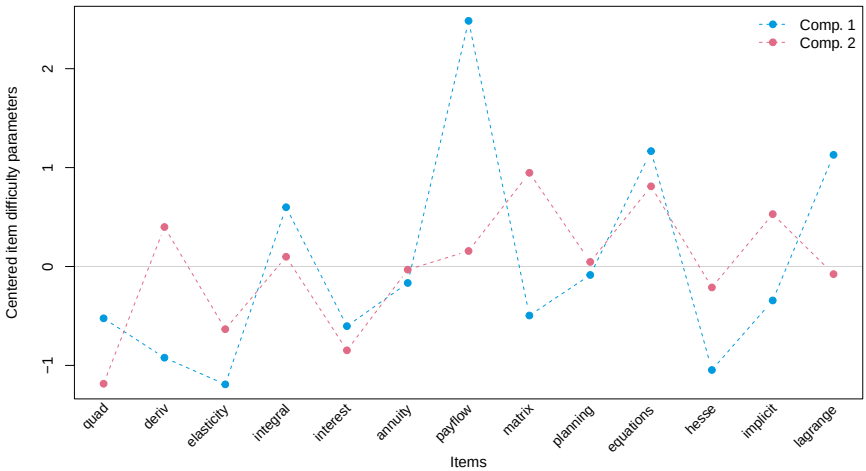
Question: How to detect differences without any covariates (e.g., in `group` 1)?

Here: Rasch mixture model with 2 components.

Result: The “soft” classification found by the mixture model is rather similar to the “hard” split by the tree.

```
R> library("psychomix")
R> mrm <- raschmix(mex1$solved, k = 2, scores = "meanvar")
R> plot(mrm)
```

Rasch mixture models



References

Zeileis A (2025). “Examining Exams Using Rasch Models and Assessment of Measurement Invariance.” *Austrian Journal of Statistics*. **54**(3), 9–26. doi:10.17713/ajs.v54i3.2055

Mastodon: @zeileis@fosstodon.org

Bluesky: @zeileis.org

Web: https://www.zeileis.org/