

Homework assignment

L8: Differential item functioning

Assignment date: 03.12.2019
Deadline: 16.12.2019 23:59
Slides: <http://www.cs.cas.cz/martinkova/NMST570>
Note: Send answers and R script to hladka@cs.cas.cz
Name:

1 Reading

Ex. 1.1 Read article available at

<https://doi.org/10.1187/cbe.16-10-0307>

and answer following questions:

1. Which methods were used for the DIF detection? [0.25]
2. What are their strengths and limitations? [0.25]
3. How many items and how many respondents are in the data in Case 1? [0.25]
4. Try to interpret significant gender gap in Case 1. [0.25]
5. How was the data set in Case 2 simulated? [0.25]
6. Which methods identify which items as DIF in Case 2? [0.25]

2 ShinyItemAnalysis

Use `ShinyItemAnalysis` (online or locally) and consider GMAT data described in previous paper. With DIF/Fairness tab answer following questions.

Ex. 2.1 Use delta plot method.

1. Which items are detected as DIF when using fixed threshold? [0.25]
2. Which items are detected as DIF when using threshold based on normal approximation? What is the value of the threshold now? [0.25]

Ex. 2.2 Use Mantel-Haenszel test.

1. Which items are detected as DIF? [0.25]
2. What are the values of DIF effect sizes for DIF items? [0.125]
3. What is the odds ratio for item 1. Interpret the result. What is the odds ratio considering only respondents with total score 12? [0.75]

4. What is the odds ratio for item 7. Interpret the result. What is the odds ratio considering only respondents with total score 12? [0.75]

Ex. 2.3 Use logistic regression

1. Which items are detected as functioning differently? [0.25]
2. Which items do favor males and which females? Is there any item performing non-uniform DIF? [0.75]
3. How do the results change when using Benjamini-Hochberg (BH) correction for multiple comparison? [0.125]

Ex. 2.4 Use Lord's test

1. Which items are detected as functioning differently for 2PL IRT model? [0.25]
2. With 2PL IRT model, how do the results change when using item purification? How many iterations were run until convergence? [0.5]

3 Real data analysis

Consider `verbal` data set from `difR` package. Follow selected R code in `ShinyItemAnalysis` and create R script to answer following questions.

HINT: Use `data(verbal, package = "difR")` to upload data.

Ex. 3.1 Explore data:

1. How many items and how many respondents are in the data? [0.125]
2. Explain names of items (e.g. what do `S1WantCurse` and `S3DoShout` mean?) [0.125]
HINT: Explore help for `verbal` dataset.

Ex. 3.2 Fit Mantel-Haenszel test with `difMH()` using item purification.

1. Which items are detected as functioning differently? [0.5]
2. Try to explain why some items perform positive effect size `deltaMH` and some items perform negative effect size. Try to interpret. [0.5]
3. Create table describing purification process. [0.5]
HINT: Check `difPur` value of the `difMH()` output.

Ex. 3.3 Fit logistic regression method with `difLogistic()`.

1. Which items are detected as functioning differently? [0.5]
2. Plot characteristic curves for DIF items using function `plot()` and argument `plot = "itemCurve"` [0.5]
3. Check plots with characteristic curves. Which items do favor males (reference) and which favor females (focal)? Is there any item showing non-uniform DIF? [0.5]

Ex. 3.4 Choose at least four methods for DIF detection offered by `difR` package. Use function `dichoDIF()` and create table displaying which items are detected as DIF by each function. Briefly describe your conclusion - which items are DIF, which items are unfair, and what would be the next step? [1]

4 Provide feedback

Here you can provide feedback on lecture, lab session and/or materials (slides, HW assignment, `ShinyItemAnalysis` manual) [1pt bonus] :)