Final project assignment

Final project comprises of validation and analysis of CERMAT measurement data or other data of your choice. Your final project is to be submitted electronically via e-mail in PDF form. As the Supplementary Materials, you should include .R (or .Rmd) file with commented R code, including data preparation and code for relevant analyses, dataset(s) needed to run the code, and voluntarily also a PDF or HTML ShinyItemAnalysis report.

Your project should contain following sections and content specified below:

Data (10 pts)

Describe your data: Which dataset did you chose? Provide the links for your dataset. Full item wording should be provided as Supplementary Material, or link/reference for document with item wording should be included. How many respondents were in your data? How many items and of what type are present in your data? Describe data preparation: How did you recode the data? Did you merge some items? What other changes to the data were necessary?

Provide descriptive statistics and histogram for the total score, descriptives for item scores, and for any additional variables present (grouping variables, criterion variables, etc.). Compute the Z-scores, T-scores, percentiles, and success rate for all respondents, interpret the results for the first respondent.

Test validity (10 pts)

What construct is your test measuring? What is intended use of test scores? Is there other use you can think of? How would you obtain evidence of validity of test scores based on test content? How would you obtain evidence of criterion validity, what data would you need? Use terms concurrent, predictive, incremental, convergent, discriminant. Provide proofs of criterion validity if criterion is available. Provide evidence of validity of test scores based on internal structure: Include corrplot, discuss which items correlate the most. Try cluster analysis, interpret dendrogram. Try exploratory factor analysis, discuss path diagram.

Test reliability (10 pts)

Provide reliability estimate(s) in terms of internal consistency, include confidence interval and interpret with respect to Cicchetti's cut-off values. Implement Spearman-Brown formula. What would be the estimated reliability if you doubled the number of items? What is the number of items needed for reliability of 0.9? What kind of data would you need to collect to provide other proofs of reliability? Discuss at least 2 other reliability estimates.

Item analysis (10 pts)

Include table of traditional item indices (difficulty, RIR, RIT, ULI, alphaDrop). Comment on items with highest and lowest difficulty, items of low discrimination or items with inappropriate distractor plot. Provide explanation taking into account wording of those items.

Choose one or two items and include empirical item characteristic curve these items, or distractor plots in case of multiple-choice items. Select and fit appropriate regression model with respect to standardized total score using intercept-slope and/or IRT parametrization. Discuss your model decision with respect to item type. Interpret item parameters.

Item response theory models (10 pts)

Select optimal IRT model: Provide reasoning for selected IRT model (based on data type or using comparison of more IRT models). Provide model equation(s) and interpretation of parameters. Which method was used for estimation of parameters?

Plot and discuss a Wright map (preferably on 1PL model or other Rasch-type model). Plot item characteristic curves, item information curves, test information curve. Provide table of model parameter estimates and their standard errors, or confidence intervals. Which item is the most informative for average ability level, and ability levels 1SD above and 1SD below the average?

Provide ability estimate for the first respondent, including the standard error or confidence interval. Plot relationship between ability estimates in IRT models and traditional ability estimates based on (standardized) total scores.

Differential item functioning (10 pts)

Explain why DIF analysis is important in test validation. Select one grouping variable, and provide analysis of DIF. Provide reasoning for selected DIF detection method, or try more methods and compare results. Discuss any items detected as DIF, provide wording or explanation for DIF: Which items favor one and which the other group?What other grouping variables should be considered for DIF analysis? If no grouping variable is present in your data, discuss what grouping variables should be collected and used for DIF detection, and which DIF detection methods could be used.

Discussion (10 pts)

Provide discussion to your results. What are the results and are they expectable? What recommendations do you have for the analyzed assessment instrument (in terms of removing/rewording items, increasing reliability or validity, data to be collected for further analyses, etc.). For what other situations/data is your analysis relevant? What are the limitations to your study/analysis? Could you think of a situation when range restriction may occur in your data, how would you adjust your analysis? What is your conclusion?

Supplement – Commented R code (10 pts)

Attach commented .R or .Rmd code for your analysis.

Supplement –ShinyItemAnalysis report and datasets (10 pts)

Attach automatically generated PDF or HTML report including available analyses. Attach datasets used to generate the report. In this section, describe the settings you chose and discuss the results. How do the models or results differ from results above? (Based on limitations of the ShinyItemAnalysis, or due to reduced dataset).

Report form (10 pts)

- Preferably use `RMarkdown` with `knitr`, see https://bookdown.org/yihui/rmarkdown-cookbook/
- Make the project one fluent text. Section instructions should not be repeated in your project.
- Hide code and output messages for this project, see options here https://yihui.org/knitr/options/
- Use your own words to describe methods. When citing longer text, citation needs to be given

Additional analyses (Bonus - up to 10 pts)

Provide any additional analysis you find relevant which is not included in the list above. You may include these analyses in relevant section above, in such case mention and discuss these analyses in this section. Which additional analyses did you consider, what were the results? Here are some possibilities: Testing item and person fit, estimating reliability with McDonnald's omega, running post-hoc CAT analysis. DIF: Provide table of item parameters and interpret for at least one DIF item; choose some DIF items and provide item characteristic curves; provide prediction for correct answer from members of one and of the other group.

Due date: Final project (PDF file and supplementary materials) is due at least 48 hours before the oral exam, and needs to be sent to martinkova@cs.cas.cz. One feedback will be provided to drafts sent at least two weeks before the oral exam. You may ask for help at dlouha@cs.cas.cz if you have any technical questions regarding `knitr` or `R`.