

Proposal for a workshop at the IMPS 2011 in Hong Kong Monday July 18th 2011

Psychometrics in R

Thomas Rusch and Patrick Mair
Institute for Statistics and Mathematics
Vienna University of Economics and Business
thomas.rusch@wu.ac.at; patrick.mair@wu.ac.at

Introduction

Over the last few years the R Open Source environment for statistical computing has become one of the most popular analysis tool in the field of computational statistics. Recently, numerous packages in the area of Psychometrics have been implemented.

The striking features of R are the following: R is Open Source which means that it is freely available and the user has full insight into the source code. R consists of a base distribution that includes the vector and matrix oriented R language and contains basic statistical techniques. In addition to this base distribution more than 2700 add-on packages are available freely for download via the Comprehensive R Archive Network (CRAN). These packages are implemented in a standardized manner which means that once the user has gained understanding of basic R concepts, he or she can use all of these packages in the same way. Another striking feature is that everything in R is stored as an object. Thus, results are not static text output but can be post-processed by means of additional data manipulation or used for subsequent analysis. Finally, it has to be mentioned that R has a powerful plot engine. Graphical outputs are highly customizable and can be plotted on a level suitable for journal publication.

The aim of this workshop is to teach participants the basic concepts of the R language and the R environment, respectively, and show them how to use R for basic statistical and advanced psychometric modeling. No prior R experience is needed in order to follow the workshop.

Outline of the Workshop

The workshop consists of three modules: The first module, “R Introduction and R programming” deals with basic concepts and the organization of R. It introduces the R language by means of psychometric examples. We start with a “tour” of the CRAN (Comprehensive R Archive Network) repository, show important on-line resources such as manuals, the R-help repository, and subject-specific task views. We explain how to import external datasets and how to install and load packages. Furthermore, we examine which data types exist in R and how to sub-select or extract elements from these data types. A core emphasis of this module is the use of R functions which we introduce in a conceptual way (help files, arguments, values). Finally, since R is matrix oriented, we show some basic matrix computations which should help the participants to implement their newly developed methods in R. Throughout the modules we will also show possibilities for plotting data and corresponding results using the R plot engine.

The second module introduces basic statistical modeling where we focus on linear and generalized linear models. The corresponding “glm” function is very powerful and versatile

and works with a simple formula syntax for model specification which will be explained in detail. As a slightly advanced topic the computation of (generalized) linear mixed-effects models (e.g., multilevel models) using the “lme4” package will be demonstrated. The corresponding “lmer” function can be used in a similar fashion as the “glm” function.

The third module is all about psychometric methods in R. We start with some descriptive multivariate techniques for categorical data such as correspondence analysis and Gifi methods (“homals”). As an additional exploratory approach we will demonstrate the “smacof” package for multidimensional scaling. Briefly, it will be shown how classical test theory can be performed in R. A strong emphasis of this module is on item response models. We will focus on the “eRm” package for extended Rasch models and the “ltm” package for higher parameterized IRT models. Within this context we will also point out the computation of non-parametric IRT (“mokken” package) and multidimensional IRT models using “MCMCpack”. Finally, we will demonstrate the computation of classical latent variable methods such as simple factor analysis and structural equation models using the newly developed package “lavaan”.

Throughout the workshop we use real-life datasets from the psychometric area. All datasets, the full R code file and additional materials will be available at

<http://statmath.wu.ac.at/~mair/IMPS2011/>

prior to the workshop such that the participants can fully reproduce our demonstrations. At the end of the workshop the participants will have a conceptual and technical understanding of R such that they can use this environment for their own analyses and hopefully have a starting point for implementing their own methods.

Program

9:30- 9.45	Welcome and Introduction to the Workshop
9:45-11.00	Introduction to R and R Programming
11.00-11.30	Break
11:30-12.30	Statistical Modeling in R
12.30-14.00	Break
14:00-15:30	Psychometric Modeling in R
15.30-16.00	Beyond the Workshop and Discussion

Links:

Main R website: <http://R-project.org>

CRAN download repository: <http://CRAN.R-project.org>

Psychometrics task view: <http://cran.r-project.org/web/views/Psychometrics.html>

Suggested code editors:

Tinn-R under Windows OS: <http://www.sciviews.org/Tinn-R/>

Aquamacs for Mac OS: <http://aquamacs.org/>

Emacs including ESS for Linux: <http://ess.r-project.org/>

Selected Books:

- Venables, W. N., & Smith, D. M. (2002). *An Introduction to R*. Bristol, UK: Network Theory. Freely available at <http://cran.r-project.org/doc/manuals/R-intro.pdf>.
- Dalgaard, P. (2008). *Introductory Statistics with R*. New York: Springer.
- Chambers, J. (2008). *Software for Data Analysis: Programming with R*. New York: Springer.
- Adler, J. (2010). *R in a Nutshell*. Sebastopol, CA: O'Reilly.
- Crawley, M. J. (2008). *The R Book*. Chichester, UK: Wiley.

Selected Articles: (freely available at <http://www.jstatsoft.org>)

- Mair, P., & Hatzinger, R. (2007). Extended Rasch modeling: The eRm package for the application of IRT models in R. *Journal of Statistical Software*, *20(9)*, 1-20.
- de Leeuw, J., & Mair, P. (2009b). Multidimensional scaling using majorization: SMACOF in R. *Journal of Statistical Software*, *31(3)*, 1-30.
- de Leeuw, J., & Mair, P. (2009a). Gifi methods for optimal scaling in R: The package homals. *Journal of Statistical Software*, *31(4)*, 1-21.
- Rizopoulos, D. (2006). ltm: An R Package for Latent Variable Modeling and Item Response Analysis, *Journal of Statistical Software*, *17(5)*, 1-25.
- van der Ark, L. A. (2007). Mokken Scale Analysis in R. *Journal of Statistical Software*, *20(11)*, 1-19.