

Introduction to R

Klaus Schliep

Max Planck Institute of Psychiatry

16.02.2004

About R

- R is a statistical computing environment with graphics capabilities
- It is fully scriptable and contains elements of programming languages
- A very high number of statistical procedures have been made available for R
- R is completely free, open source, and available for different platforms (Windows, Unix, Linux, MacOS etc.)
Downloads at: www.cran.r-project.org

History of R and S

- The S language has been developed since the late 1970s by John Chambers and his collaborators at Bell Labs
- Several versions of this language have been issued, the latest being S4
- *S-PLUS* – a commercialised version of Chambers' work which is marketed by *Insightful*
- *R* – an free-software implementation started from 1997 by R. Gentleman and R. Ihaka at the University of Auckland
- Though not fully compatible to S-Plus, most of the code runs also in R without alterations and vice versa

Pros

- an effective data handling and storage facility
- a suite of operators for calculations on arrays, in particular matrices
- a large, coherent, integrated collection of intermediate tools for data analysis, graphical facilities and display either directly at the computer or on hardcopy
- a well developed, simple and effective programming language which includes conditionals, loops, user defined recursive functions and input and output facilities

Cons

- R is not a database, but it can connect to DBMS (Postgres, Oracle)
- R has not a graphical user interface (but interfaces to JAVA and Tcl/Tk exist)
- no spreadsheet view of data (connection to Excel possible)
- as an interpreted language, some procedures may be slow; however, compiled C, C++ or FORTRAN code can be called from within R

References

- Dalgaard, P. (2002): *Introductory Statistics with R*, Springer
- Venables, W. and Ripley, B. (2002): *Modern Applied Statistics with S-PLUS*, Springer
- Venables, W. and Ripley, B. (2000): *S Programming*, Springer
- Chambers, J. (1998): *Programming With Data*, Springer (also called the “Green” book, very sophisticated)

Let's start

