Data analysis with R software

Data-analyysi R-ohjelmistolla

Tommi Härkänen

National Institute for Health and Welfare (THL), Helsinki
E-mail: tommi.harkanen@helsinki.fi

University of Helsinki, January 17, 2012

Contents

Practical issues

Introduction to R

User interface and basic syntax

General description of R

R is an integrated suite of software facilities for data manipulation, calculation and graphical display. It includes

- 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** for data analysis and display either on-screen or on hardcopy, and
- a well-developed, simple and effective **programming language** which includes conditionals, loops, user-defined recursive functions and input and output facilities.

Lectures, computer class excercises, practical work

Lectures Tuesdays 12-14 and Wednesdays 16-18 from January 17 to February 22 in B123.

Moodle Students must register to Moodle using **key** which has been sent to students by e-mail.

Computer class excercises Room C128. Thursdays, 10-12 and 12-14; Fridays 14-16 and 16-18; Mondays 10-12, 12-14, 14-16 and 16-18. The assignments and the data sets will be available in Moodle on Wednesdays.

Practical work is used for the evaluation.

Practical work

- The assignment of the practical work will be available via Moodle during week 8.
- Students write a report (in **pdf** format), which should contain
    - All R commands, which have been used,
    - their output and
    - short and clear verbal documentation, what has been done and what the results are.
- Students return the reports via Moodle.
- An assessment consists of a numerical grade (values 0 to 5).

Participants
Source: WebOodi
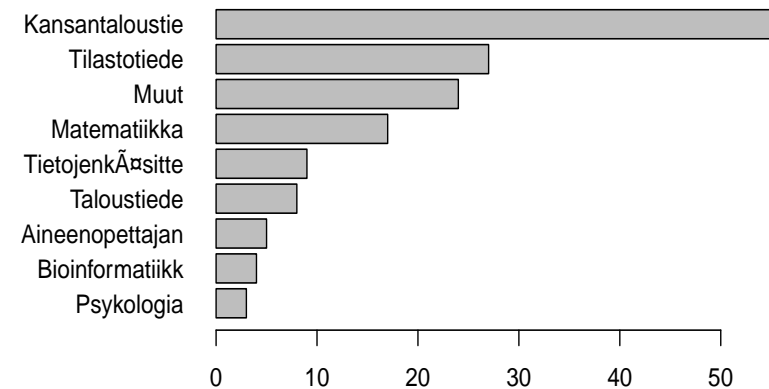
History

- The S language
    - Old S S is one of several statistical computing languages that were designed at Bell Laboratories, and first took form between 1975-1976.
    - New S In 1991, *Statistical Models in S* (White Book) was published, which introduced, e.g. data frame objects.
- The R software
    - Version 0.16 This is the last alpha version developed primarily by Ihaka and Gentleman. Much of the basic functionality from the White Book was implemented. The mailing lists commenced on April 1, 1997.
    - Version 1.0.0 (February 29, 2000) Considered by its developers stable enough for production use.

Web sites

- The main web site: r-project.org.
- wikipedia.org.
- r-bloggers.com is a central hub (e.g: A blog aggregator) of content collected from bloggers who write about R (in English).
- A blog in Finnish: r-ohjelmointi.org.

Documentation and help

- Documentation in pdf and html formats: see the Help menu.
- r-project.org:
  - Manuals
  - Mailing lists
  - Link to CRAN: The Comprehensive R Archive Network
    - **Download** R for Windows, Mac OS X or Linux
    - Contributed **extension packages** (3,523 of them at the moment)
    - CRAN **Task Views** allow you to browse **packages by topic** and provide tools to automatically install all packages for special areas of interest. Currently, 28 views are available.

Summary of 2nd hour

- Character strings
- Data frames and matrices
- Lists
- Logical operations

Summary of 1st hour

- Scalars and vectors
- Variables
- Assigning values to elements of a vector
- Functions
- Graphical output