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 15, 2013

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.

This course will cover

- Introduction to the R software and the basics of the S language.
- Importing, transforming and saving data.
- Graphics.
- Basic descriptive statistics and statistical analyses.
- Additional extension packages of R.

Lectures, computer class excercises, practical work

Lectures Tuesdays 14-16 and Wednesdays 16-18 from January 15 to February 20 in CK112.

Moodle Students must register to Moodle using **key**
which can be obtained from the lecturer by e-mail.
The key has been sent by e-mail to the registered students.

Computer class excercises

- Given by Christian Benner in room C128.
- Thursdays, 10-12, 12-14 and 14-16;
  Fridays 10-12, 12-14 and 14-16.
- The assignments and the data sets will be available in Moodle on Wednesdays.
- Students can use their own laptops.

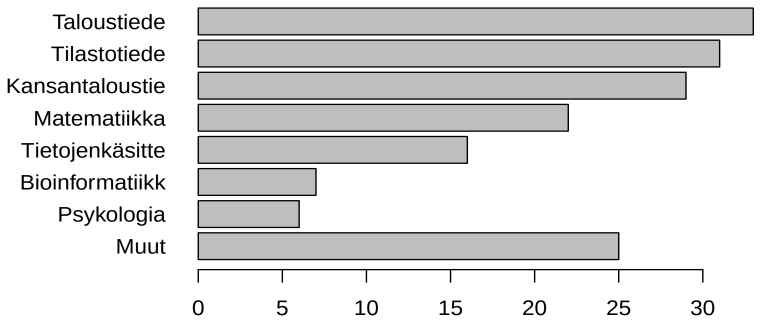
Practical work is used for the evaluation.

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): introductions, examples, new/updated R packages, . . .
- Examples on R graphics r-enthusiasts.com
- A blog in Finnish: r-ohjelmointi.org.
- An improved user interface for R is **RStudio**: rstudio.com Not covered by this course.

Documentation and help

- Documentation in pdf and html formats: see the Help menu.
- r-project.org:
  - Manuals
  - Mailing lists
  - Link to CRAN
- CRAN: The Comprehensive R Archive Network
  - **Download** R software for Windows, Mac OS X or Linux
  - Contributed **extension packages** (4,235 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 1st hour

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

Operations on vectors

```
x <- 1:9   ## Define variable x
x
z <- x * x   ## Define z as a square of x
z
z[c(3, 5)] <- c(-1, -2)  ## Change 3rd and 5th elements
z
z[x > 6] <- 0  ## Set elements of z to zero where x greater than 6
z
```

```
[1] 1 2 3 4 5 6 7 8 9

[1]  1  4  9 16 25 36 49 64 81

[1]  1  4 -1 16 -2 36 49 64 81

[1]  1  4 -1 16 -2 36  0  0  0
```

Summary of 2nd hour

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

Lists

```
m <- matrix(1:12, ncol=6) ## Define a 2 x 6 matrix
z <- list(m, v1="some text") ## Define list z
names(z) ## Get element names
names(z)[1] <- c("m") ## Rename 1st element of list 'z'
z  ## Print 'z'
```

```
[1] ""   "v1"

$m
     [,1] [,2] [,3] [,4] [,5] [,6]
[1,]    1    3    5    7    9   11
[2,]    2    4    6    8   10   12

$v1
[1] "some text"
```

Logical operations

```
x <- 1:5; x   ## Define and print variable x
x > 3
y <- c(3:1, 2:3); y   ## Define and print y
x > y
(x > 3) & (x > y)  ## AND: Both corresponding elements TRUE?
(x > 3) | (x > y)  ## OR:  Both corresponding elements TRUE?
```

```
[1] 1 2 3 4 5

[1] FALSE FALSE FALSE  TRUE  TRUE

[1] 3 2 1 2 3

[1] FALSE FALSE  TRUE  TRUE  TRUE

[1] FALSE FALSE FALSE  TRUE  TRUE

[1] FALSE FALSE  TRUE  TRUE  TRUE
```