Introduction to UsingR

Wim Krijnen
Lector Analyse Technieken voor Praktijkonderzoek
Lectoraat Healthy Ageing, Allied Health Care and Nursing
Hanze University of Applied Sciences

July 1, 2015
Purpose of this course

- Open for anybody against minimal costs
- Give brief introduction in using R
- Getting you started and self sufficient
- Providing help, examples, tutorials, (free) literature
- Solving some of your problems
- Continuation on demand, once in 3 months
- Starting a community: share ideas and helping each other
How to do this course

- Bring in your own work and start from there
- Study and learn from examples
- Ability to write correct scripts determines usefulness
- Come up with ideas, problems, mistakes, and frustrations!
- Be active and explore!
- Step in the world of statistical programming
- Enjoy the power of programming
Outline

1. Purpose of this course
2. How to do this course
3. Situations for Using R
4. Reasons for Using R
5. Some advise before starting
6. Installing
7. Where to get Help
8. Reading and writing
9. Important Functions
10. Some Plots of data
11. Some Statistical Tests
Situations for Using R

- Repeated similar problems
- Programming of visualizations: publication ready plots
- Handle large data sets
- Desire flexibility in statistical programming
- Use of modern techniques (bootstrap, robust, Bayesian)
Outline

1. Purpose of this course
2. How to do this course
3. Situations for Using R
4. Reasons for Using R
5. Some advise before starting
6. Installing
7. Where to get Help
8. Reading and writing
9. Important Functions
10. Some Plots of data
11. Some Statistical Tests
Reasons for Using R

- Widely used in statistics and applied sciences
- Reliable free open source
- Versatile: SPSS, Matlab, MySQL, Perl, JAVA, C++, Fortran
- Extensive help
- Numerous libraries with modern methods
- High level language with many built-in-functions

Disadvantages:
- Steep learning curve (use it regularly)
- Command line, some GUI
- Not fastest (C++)
Outline

1. Purpose of this course
2. How to do this course
3. Situations for Using R
4. Reasons for Using R
5. Some advise before starting
6. Installing
7. Where to get Help
8. Reading and writing
9. Important Functions
10. Some Plots of data
11. Some Statistical Tests
Some advise on starting

- Analyze needs and whether it fits to R
- R for Beginners
- Simple R
- IcebreakR
- Work on your own problems
Outline

1. Purpose of this course
2. How to do this course
3. Situations for Using R
4. Reasons for Using R
5. Some advise before starting
6. Installing
7. Where to get Help
8. Reading and writing
9. Important Functions
10. Some Plots of data
11. Some Statistical Tests
Install R

URL: http://cran.r-project.org
choose operating system: Windows, Linux, Mac
choose base
html help
installing libraries

Install a library
```r
chooseCRANmirror()
install.packages(c("TeachingDemos"), repo="http://cran.r-project.org", dep=TRUE)
library(TeachingDemos)
plot(dice(12,1))
```

Install a bundle:
```r
install.packages("ctv")
library("ctv")
install.views("Robust")
install.views("Psychometrics")
install.views("Econometrics")
install.views("SocialSciences")
```
Outline

1. Purpose of this course
2. How to do this course
3. Situations for Using R
4. Reasons for Using R
5. Some advise before starting
6. Installing
7. Where to get Help
8. Reading and writing
9. Important Functions
10. Some Plots of data
11. Some Statistical Tests
An Introduction to R
The R Language Definition
R Installation and Administration
R Data Import/Export
Searching and Tutorials

From http://cran.r-project.org

- Search, Task Views, Manuals, FAQs, The R Journal, Wiki
- Contributed "R reference card",

Useful tutorial for beginners:
http://cran.r-project.org/doc/contrib/Verzani-SimpleR.pdf
http://www.r-project.org → choose Books
from commandline

- `help.start()`
- `library(), ls(package:stats), library(help="stats")`
- `help(t.test), ?sum, ??solve, apropos("if")`
- `methods(plot)` plotting functions
- `example(boxplot)` examples
- `demo()` demonstrations of code
- `mean.default` study code of function
Help on Programming

- help(Control) : “for” and “while” loops
- help(Syntax) : syntax of operators
- help(Logic) : logical operators AND, OR, negation
- help(Arith) : on arithmetic, relational, logical operators, mathematical functions
- help(Special) : gamma function
### Reading and writing

<table>
<thead>
<tr>
<th>Reading</th>
<th>Writing</th>
</tr>
</thead>
<tbody>
<tr>
<td>function</td>
<td>function</td>
</tr>
<tr>
<td>library</td>
<td>library</td>
</tr>
<tr>
<td>base</td>
<td>base</td>
</tr>
<tr>
<td>utils</td>
<td>utils</td>
</tr>
<tr>
<td>read.table</td>
<td>write.table</td>
</tr>
<tr>
<td>read.csv</td>
<td>write.csv</td>
</tr>
<tr>
<td>source</td>
<td>base</td>
</tr>
<tr>
<td>read.spss</td>
<td>save</td>
</tr>
<tr>
<td></td>
<td>x11</td>
</tr>
<tr>
<td></td>
<td>postscript</td>
</tr>
<tr>
<td></td>
<td>xtable</td>
</tr>
<tr>
<td></td>
<td>grDevices</td>
</tr>
<tr>
<td></td>
<td>grDevices</td>
</tr>
<tr>
<td></td>
<td>xtable (LaTeXusers)</td>
</tr>
</tbody>
</table>
Outline

1. Purpose of this course
2. How to do this course
3. Situations for Using R
4. Reasons for Using R
5. Some advise before starting
6. Installing
7. Where to get Help
8. Reading and writing
9. **Important Functions**
10. Some Plots of data
11. Some Statistical Tests
Important Functions

scalar, vector, matrix, list, function, environment;
Everything is an object, belonging to a class!
Once defined extract information from it, use functions on it.
q(), history; quit, previous commands
rm(), rm(list=ls()); remove objects
ls(), objects(); listing of objects
class(x), str(x); class or structure of object x
getwd, setwd, dir; get set working directory, interaction with OS
numeric, character, matrix, data.frame, list; construct object
factor, gl; construct factor
summary, residuals, coef; generic functions for lm, glm etc.
mean, median, sd, IQR, quantile; descriptive statistics
plot, matplot; plotting
rownames, colnames, rowcolnames
function, apply, lapply; apply function on row/columns of matrix
grep, regexpr; regular expressions
I. Purpose of this course

II. How to do this course

III. Situations for Using R

IV. Reasons for Using R

V. Some advise before starting

VI. Installing

VII. Where to get Help

VIII. Reading and writing

IX. Important Functions

X. Some Plots of data

XI. Some Statistical Tests
## Some Plots of data

<table>
<thead>
<tr>
<th>Graphical representation</th>
<th>R function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Box-and-Wiskers</td>
<td>boxplot</td>
</tr>
<tr>
<td></td>
<td>hist</td>
</tr>
<tr>
<td></td>
<td>pie</td>
</tr>
<tr>
<td></td>
<td>barplot</td>
</tr>
<tr>
<td></td>
<td>plot(density())</td>
</tr>
<tr>
<td></td>
<td>stripchart</td>
</tr>
<tr>
<td>Histogram</td>
<td></td>
</tr>
<tr>
<td>Pie</td>
<td></td>
</tr>
<tr>
<td>Bar</td>
<td></td>
</tr>
<tr>
<td>Density</td>
<td></td>
</tr>
<tr>
<td>Dot Plot</td>
<td></td>
</tr>
</tbody>
</table>
Outline

1. Purpose of this course
2. How to do this course
3. Situations for Using R
4. Reasons for Using R
5. Some advise before starting
6. Installing
7. Where to get Help
8. Reading and writing
9. Important Functions
10. Some Plots of data
11. Some Statistical Tests
### Some Statistical Tests

<table>
<thead>
<tr>
<th>Test</th>
<th>$H_0$</th>
<th>function</th>
</tr>
</thead>
<tbody>
<tr>
<td>t-test</td>
<td>$\mu_1 = \mu, \mu_1 = \mu_2$</td>
<td>t.test</td>
</tr>
<tr>
<td>Wilcoxon; Mann-Whitney</td>
<td>$F = G$</td>
<td>wilcox.test</td>
</tr>
<tr>
<td>ANOVA</td>
<td>$\mu_1 = \mu_2 = \mu_3$</td>
<td>anova</td>
</tr>
<tr>
<td>Kruskal-Wallis</td>
<td>$F_1 = F_2 = F_3$</td>
<td>kruskal.test</td>
</tr>
<tr>
<td>association</td>
<td>$\tau = \tau_0$</td>
<td>cor.test</td>
</tr>
<tr>
<td>probability of success</td>
<td>$\rho = \rho_0$</td>
<td>binom.test</td>
</tr>
<tr>
<td>normality</td>
<td>$X$ norm. distrib.</td>
<td>shapiro.test</td>
</tr>
<tr>
<td>Kolmogorov-Smirnov</td>
<td>$F = G$</td>
<td>ks.test</td>
</tr>
<tr>
<td>F-test</td>
<td>$\sigma_1 = \sigma_2$</td>
<td>var.test</td>
</tr>
<tr>
<td>Fisher’s exact test</td>
<td>independence</td>
<td>fisher.test</td>
</tr>
</tbody>
</table>
apply(matrix, margin, fun)
col.means <- apply(x, 2, mean)
col.means <- numeric
for (j in 1:ncol(x))
col.means[j] <- mean(x[, j])
Example: Daily energy intake

Daily energy intake (Altman, 1991, p.183) of group of woman; recommended intake 7725 kJ

$H_0 : \mu = 7725\text{kJ}, \; H_0 : \mu \neq 7725\text{kJ}$

> x <- c(5260, 5470, 5640, 6180, 6390, 6515, 6805, 7515, 7515, 8230, 8770)
> t.test(x, mu=7725)

One Sample t-test

data:  x

 t = -2.8208, df = 10, p-value = 0.01814
alternative hypothesis: true mean is not equal to 7725
95 percent confidence interval:
  5986.348 7520.925
sample estimates:
mean of x
  6753.636

Conclusion: $H_0$ not rejected.