Applied Statistics Using R

Christian Müller

Seminar Presentations in English – Arts and Social Sciences

April 1, 2010

Table of Contents

Why using R?
Some Reasons

- ▶ You want to use sophisticated quantitative methods
- ▶ You want to apply statistics in innovative ways
- ▶ You need a statistical package that can keep up with you
- ▶ R is your statistical package of choice

Why using R?
Some Reasons

- You want to use sophisticated quantitative methods
- You want to apply statistics in innovative ways
- You need a statistical package that can keep up with you
- R is your statistical package of choice

Why using R?
Some Reasons

- You want to use sophisticated quantitative methods
- You want to apply statistics in innovative ways
- You need a statistical package that can keep up with you
- R is your statistical package of choice

Why using R?
Some Reasons

- You want to use sophisticated quantitative methods
- You want to apply statistics in innovative ways
- You need a statistical package that can keep up with you
- R is your statistical package of choice

Why using R?

- ▶ R is maintained and constantly advanced by some of the most renowned statisticians worldwide
- ▶ Many additional packages and extension are available
- ▶ R is FREE! (GNU-GPL)
- ▶ R is a joint project of applied statisticians for applied statisticians

Why using R?

- ▶ R is maintained and constantly advanced by some of the most renowned statisticians worldwide
- ▶ Many additional packages and extension are available
- ▶ R is FREE! (GNU-GPL)
- ▶ R is a joint project of applied statisticians for applied statisticians

Why using R?

- ▶ R is maintained and constantly advanced by some of the most renowned statisticians worldwide
- ▶ Many additional packages and extension are available
- ▶ R is FREE! (GNU-GPL)
- ▶ R is a joint project of applied statisticians for applied statisticians

Why using R?

- R is maintained and constantly advanced by some of the most renowned statisticians worldwide
- Many additional packages and extension are available
- R is FREE! (GNU-GPL)
- R is a joint project of applied statisticians for applied statisticians

Why using R?
Advantages and Disadvantages

Advantages

- ► extremely flexible
- ► highly programmable
- ► you can do much more than standard, pre-canned analysis

Disadvantages

- ► harder to learn than other statistical packages
- ► scant documentation
- ► you should have some statistical background

Why using R?
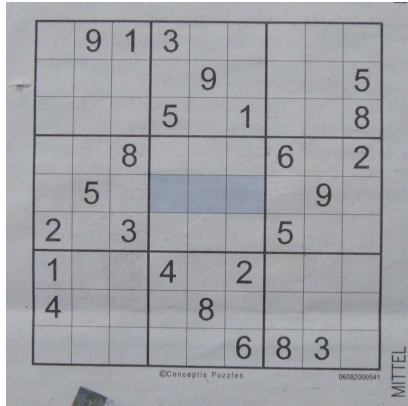## Advantages and Disadvantages

Advantages

- ▶ extremely flexible
- ▶ highly programmable
- ▶ you can do much more than standard, pre-canned analysis

Disadvantages

- ▶ harder to learn than other statistical packages
- ▶ scant documentation
- ▶ you should have some statistical background

Applied Example
Solving a sudoku

source: 20 Minuten, 30.3.2010.

Applied Example
Solving a sudoku

- Mathematically there are a number of different methods:
  - Cultural Genetic Algorithm (CGA)
  - Simulated Annealing (SA)
  - Quantum Simulated Annealing (QSA)
  - Hybrid method combining Genetic Algorithm with Simulated Annealing (HGASA)

Applied Example
Solving a sudoku

- Mathematically there are a number of different methods:
    - Cultural Genetic Algorithm (CGA)
    - Simulated Annealing (SA)
    - Quantum Simulated Annealing (QSA)
    - Hybrid method combining Genetic Algorithm with Simulated Annealing (HGASA)

Applied Example
Solving a sudoku

- Mathematically there are a number of different methods:
  - Cultural Genetic Algorithm (CGA)
  - **Simulated Annealing (SA)**
  - Quantum Simulated Annealing (QSA)
  - Hybrid method combining Genetic Algorithm with Simulated Annealing (HGASA)

Applied Example
Solving a sudoku

```
# target function to minimise
# that counts lines+rows+blocks
# missing the no duplicate constraint
target=function(s){
  tar=sum(apply(s,1,duplicated)+apply(s,2,duplicated))
   for (r in 1:9){
      bloa=(1:3)+3*(r-1)%%3
      blob=(1:3)+3*trunc((r-1)/3)
      tar=tar+sum(duplicated(as.vector(s[bloa,blob])))
     }
  return(tar)
  }
```

Applied Example
Solving a sudoku

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

 > target(cur)
[1] 0
```

Applied Example
Solving a sudoku

source: 20 Minuten, 30.3.2010.

Conclusion

- ▶ R is a high level statistical package
- ▶ R is versatile and can be adjusted for different needs
- ▶ You should use R for your statistical analysis

Conclusion

- ▶ R is a high level statistical package
- ▶ R is versatile and can be adjusted for different needs
- ▶ You should use R for your statistical analysis

Conclusion

- ▶ R is a high level statistical package
- ▶ R is versatile and can be adjusted for different needs
- ▶ You should use R for your statistical analysis

Bibliography

Meir Perez and Tshilidzi Marwala (2008)
Stochastic optimization approaches for solving sudoku
Int. Workshop on Stochastic and Applied Global Optimization: Skukuza, South Africa

R Core Developer Team (2009)
The R project for statistical computing
www.r-project.org

Ziheng Yang (2010)
Sudoku via simulated annealing
http://xianblog.wordpress.com/2010/02/23/sudoku-via-simulated-annealing/

Questions ?!