Statistical Computing
Week 1: 4. October 2019
Course Introduction and Using RStudio **for the First Time**

Lecture 1: Contents

- Course Introduction
- A brief Introduction to R
- Using R for the first time: Workshop
- At home: Read through the further reading sheet in Moodle.

Course Admin

The important details are given in the course information sheet in Moodle. This is just a quick overview.

Lecturer: Prof. Tim Downie (FB II, Raum A126)

Email: tim.downie@beuth-hochschule.de

Office Hour: I have no formal office hour. Please email me if you want to arrange a meeting.

Moodle: http://lms.beuth-hochschule.de/moodle/course/view.php?id=17295
All course material and news will be distributed using Moodle. You should check your Beuth email regularly to see if there are any messages.

Course Aims: To learn the basics of statistical methods, graphics and data analysis.
Particular focus is put on using R and R-Studio to analyse data by applying these methods.

The skills learnt in this course are directly relevant to the following Data Science courses

- Regression
- Data Visualization
- Machine Learning I & II

Timetable

Officially there are weekly blocks of *Seminarische Unterricht* (lectures) with *Übungsgruppen* (workshops) fortnightly.

In practice each week will start with a lecture, which will be followed (most weeks) by a computing workshop. The exact time allocation to lecture/workshop will vary from week to week. This gives you more hands-on computing time than if we stick rigidly to the official lecture/workshop allocation.

All lectures and workshops take place in Room A 129L.

On odd calendar weeks we have one teaching block from 12:15 until 13:45.

On even calendar weeks [like this week], we have two teaching blocks from **12:15** until **15:30**. There will be a fifteen minute break at a convenient time.

Semester Timetable

Week	Date	No. blocks	Topics covered	
1	4/10/2019	2	Using RStudio for the First Time	
2	11/10/2019	1	Population, samples, frequencies	
3	18/10/2019	2	Measures of location, quantiles, basic graphics	
4	25/10/2019	1	Measures of dispersion, covariance	
5	1/11/2019	2	Ecdf, boxplots and exploratory data analysis	
...	...	...	...	...

The complete timetable is in the info sheet but it is only meant as a guide. The topics might shift week or change as the semester progresses. I will keep you informed via Moodle.

Assessment

Practical Test There will be one sixty minute practical test using R and R-Studio. This test contributes 20% to your final course mark.

The planned date for the test is **20th December 2019**.

The practical test is not compulsory but absentees without a medical certificate will get 0 points. There will be no “Nachklausur” (second test) for the practical test.

Exam There are two ninety-minute written exams contributing 80% to your final course mark. **You only need to pass one of the two exams**. The proposed dates for the exam are **24th January 2020** and **26th March 2020**.

A minimum of 50% (practical test + exam) is required to pass the course.

The week before the exam there will a revision class.

Course Notes

You should have the course notes and the worksheets (for the computer workshops) available during the lessons.

The course notes provided are fairly comprehensive, BUT they contain gaps left for you to fill in. You should make sure that you can edit the script during the lectures. Either print out a copy before the lecture or use a tablet and e-pen.

It is not necessary to print out the worksheets, you can view these directly on the Lab-PCs.

What is R?

- R is a computer statistics system based on the programming language S (technically a “function language”).
- R has many “inbuilt” statistical commands, e.g. to fit a linear regression model. It also provides an environment for writing your own functions.
- Advantages over other statistics packages include flexibility, simplicity, and the quality of graphical display.
- R is the software most widely used by professional statisticians and along with *Python* has become one of the most widely used interactive software for data analysis and machine learning.
- R is available free of charge. You can download your own version from http://www.R-project.org.

A quick History of S-Plus and R

See the Further Reading document

R and RStudio

The statistics courses in this Data Science Program will use R extensively. You will be using RStudio as the front end, which makes writing source code, handling data objects and graphics a little easier.

All the computing is done by R itself. To use RStudio, R has to be installed first.

Other lecturers may use R without RStudio; there is no difference in the code and output, so you will be able to use RStudio in other courses, even if the the lecturer uses R directly.

If you use Linux or a Mac at home, again there will be no difference in the code and output.

In the exam all questions will be independent of operating system and front end software. In the R-Test the exercise sheet will be written assuming that you will be using RStudio.

Workshop 1

- You will be given an account to login to the Lab-Computers.
- When you log in change your password immediately. Bring your log in details with you every Friday!
- Go to the Moodle website for this course and find the PDF file *Worksheet 1*.
- I will talk you through section 2.1 and 2.2 (Getting Started).
- Work through the rest of the worksheet at your own pace, entering the R-Commands and answering the questions (usually marked ►).
- Ask me if you have questions, if something isn't working or don't know an English word that I use. ... but be prepared to wait as there is only one of me.
- **Let's go!**