

Statistics with R – Bootcamp

Course instructor

Prof. Dr. Christian Hildebrand
Full Professor of Marketing Analytics
Institute of Behavioral Science & Technology (IBT)
University of St. Gallen
Torstrasse 25
CH-9000 St. Gallen
Email: christian.hildebrand@unige.ch
Web: <https://www.ibt.unisg.ch/team/christian-hildebrand/>

Location

Campus Munich (1 day) and virtual (2 days), see “Preliminary schedule” below

Rooms: 2566

Application and registration procedure

Goal and target audience

This course is an introduction to the statistical computing environment R. The course is aimed at doctoral candidates.

Application/Registration process

Please register for the course by **Tuesday, August 22** by sending an email to contact.dm@mgt.tum.de.

Course aims

What this course is

This course is an introduction to the statistical computing environment R. In this course, you will learn how to interact with, process, analyze and visualize data in R. The main objective of this course is to familiarize participants with R and to develop competencies to effectively work on their own research projects with R. This course is also helpful as a primer for other summer program courses such as Advanced Regression or Data Mining.

Course objectives

These workshop lectures are designed to introduce participants to the statistical computing environment R. In this course you will learn how to use R for effective data analysis. We will cover a selected range of both basic topics (e.g., reading data into R, data structures (i.e., data frames, lists, matrices), data manipulation, statistical graphics) to more advanced topics (e.g., writing functions, control statements,

loops, reshaping data, string manipulations, and statistical models in R).

This course is also helpful as a primer for other summer program courses that will use R, such as the courses on Computational Statistics, Data Mining, or Advanced Regression Modeling, among other courses. No prerequisites are required for this course.

The focus of this course is on learning the specifics of the statistical computing environment R. This course is not designed to learn various statistical methods as in other summer program courses, although we cover some implementation details.

After taking this course you should be able to analyze any data set with R – from (1) data processing, (2) visualization, to (3) the statistical analysis of the data.

Preliminary schedule

Class Number	Date	Topic
1	Thursday, August 31, 2023 (MUNICH)	Part 1 – From Syntax to Data Structures <ul style="list-style-type: none"> • Getting R • R Environment • Syntax & Mathematical Operations • Data Types, Data Structures & Data Type Conversion • Reading Data from Various Sources • Speeding up your Learning Curve – Tipps and Tricks to Learn R
2	Thursday, September 7, 2023 (VIRTUAL)	Part 2 – From Data Manipulation to Summary Statistics <ul style="list-style-type: none"> • Reading and Manipulating Data • Conditional Execution – Control Statements • Repetitive Execution – Loops • String Manipulations • Summarizing Data • Descriptive Statistics
3	Friday, September 8, 2023 (VIRTUAL)	Part 3 – Data Visualization & Statistical Model Building <ul style="list-style-type: none"> • High-level Plotting Commands • Graphics Parameters • Plotting Multivariate Relationships • Correlation & Covariance • T-Tests, ANOVA, Regression • Statistical Model Building & Model Selection

Core readings

Lander (2017): R for Everyone: Advanced Analytics and Graphics, 2nd edition, Addison-Wesley.

Long & Teetor (2019): R Cookbook, 1st edition, O'Reilly.

James / Witten / Hastie / Tibshirani (2021): An Introduction to Statistical Learning with Applications in R, 2nd edition, Springer.

Grolemund & Wickham (2017): R for Data Science, O'Reilly.

Course procedures

Static Resources To Foster Your Learning Success

We will talk more extensively about the bibliography in the first class and I will highlight different paths for different people based in varying interests and prior knowledge.

Some of the books are recommended as a general purpose summary of R commands and examples (e.g., "R Cookbook") while others are excellent references for a variety of statistical concepts and how they are implemented in R (e.g., "Statistical Learning with Applications in R"), or even focusing on advanced aspects of programming (e.g., "Art of R Programming").

- Lander (2017): R for Everyone: Advanced Analytics and Graphics, 2nd edition, Addison-Wesley.
- Long & Teetor (2019): R Cookbook, 1st edition, O'Reilly.
- James / Witten / Hastie / Tibshirani (2021): An Introduction to Statistical Learning with Applications in R, 2nd edition, Springer.
- Grolemund & Wickham (2017): R for Data Science, O'Reilly.

Dynamic Resources To Foster Your Learning Success

Additional resources I can recommend:

Google Developer Playlist on R (21 Videos; a bit dated but still well explained): <https://goo.gl/7jCOSV>

Researchers at the biostatistics department at Johns Hopkins University have developed a highly recommended R package called "swirl". It allows you to learn topics within the R console interactively at your own pace. I will walk you through at the end of day 1 or day 2.

<http://swirlstats.com/>

Software & Editor Options

I generally recommend using RStudio in class to understand the operational details of R better (object orientation, workspace management etc.) and to have a nice interface including helpful features such as syntax highlighting, code completion.

Participants with a background in programming may already use notepad++ and there is a plugin for R as well. However, I recommend using R Studio as the to-go-to environment writing code in R (code completion, package referencing, debugging etc.).

<http://www.rstudio.com/>

<http://sourceforge.net/projects/npptor/>

Assessment

There will be no final exam or assessment. The course only requires active participation.

Workload

3 ECTS (21 hours lectures, 90 hours total workload)