



## DEPARTMENT OF MEDICINE, HUDDINGE

### **H7F5300, Get started with R – Programming Basics, Data Analysis and Visualisation, 3 credits (hec)**

Kom igång med R – programmeringsgrunder, dataanalys och visualisering, 3 högskolepoäng

*Third-cycle level / Forskarnivå*

---

#### **Approval**

This syllabus was approved by the The Committee for Doctoral Education on 2024-02-05, and is valid from autumn semester 2024.

#### ***Responsible department***

Department of Medicine, Huddinge, Faculty of Medicine

#### **Prerequisite courses, or equivalent**

No prerequisite courses, or equivalent, demanded for this course.

#### **Purpose & Intended learning outcomes**

##### **Purpose**

The course is practical and aims at teaching students how to:

Use the programming environment R and RStudio, which includes installation, how to handle errors, problem solve and access helper documents.

Use basic concepts of programming, such as data types, logical and arithmetic operators, if else conditions, loops and functions.

Use common R packages to perform basic statistical analysis (e.g., t-test, chi2-test, correlation) and visual presentation (e.g., boxplot, histogram and heat-map) of data in R.

##### **Intended learning outcomes**

After attending the course the student should know:

- How to download, install and navigate R and RStudio
- How to solve common problems arising from data formatting and handling
- Common programming concepts and how to employ them in R

- How to import data and packages in R
- How to use R for basic statistical analysis and visual presentation of data

## Course content

Course participants start the course by installing and familiarising with the R and RStudio environment. This includes version control, as well as structuring and documenting code for publication. Next, basic concepts shared between all programming languages are introduced, such as data types and operators. Students will also learn how to use recommended naming conventions, syntax and how to comment code. Methods for importing packages and data is then introduced and students will learn how to search for help and get examples of common problems that may arise. Finally, students will practice using packages for data management, statistical analysis and visual presentation. Methods include distribution tests, power-analysis, t-test, chi2-test, correlation, boxplot, scatterplot and bar plot. Visual presentation will mainly use the ggplot2 package, providing a good example of object-oriented programming in R.

Throughout all lectures focus will be on application and understanding of the methods used, not statistical assumptions or interpretation of the results. Examples will primarily be taken from experimental research and tasks will use dataframes available upon installation of R. However, when possible students are encouraged to use their own data.

The last day of the course can either be used to continue to apply R on own data or to learn procedures that can be performed with R which most other statistical software's cannot. Such as, managing folders and files, querying databases and importing codes and algorithms.

## Forms of teaching and learning

Distance learning with online interactive lectures. Group and individual exercises where a teacher will be available to help. Assignments and Canvas quizzes that the student completes on their own. Reviewing other students' code and interaction with other students. Individual project work. Four days each week will consist of lectures in the morning introducing concepts and tasks in the afternoon, where these concepts are put to practice. The last day of each week will be a larger exercise where the student is required to combine introduced concepts into a whole. This exercise will be reviewed by a fellow student who will have the opportunity to comment on ways to improve the work. The 11th (last) day is optional and described in the previous paragraph.

### *Language of instruction*

The course is given in English

## Grading scale

Pass (G) / Fail (U)

## Compulsory components & forms of assessment

### Compulsory components

Canvas quizzes and tasks. Individual projects and reviews of other students' project.

Participation during project presentation and review.

Students who miss obligatory elements will complete extra tasks associated with the specific element. Course participants unable to participate during the project presentation will have the presentation for the course administrator but will miss the opportunity to get their work reviewed by other participants.

### **Forms of assessment**

Project presentation and review.

## **Course literature**

Literature recommendations (not required):

Michael J. Crawley, "Statistics: An Introduction Using R", Wiley, 2014, ISBN 978-1118941096

For students used to Excel another alternative is (although it deals primarily with data manipulation):

John L Taveras, "R for Excel Users: Introduction to R for Excel Analysts", CreateSpace Independent Publishing Platform, 2016, ISBN 978-1500566357