Chapter 1

Introduction

This document is designed to guide the students through the R course "R- a tool for statistical analysis" given at the University of Umeå. It is not a comprehensive introduction to R but focuses on introducing basic concepts. Examples are mostly of epidemiological nature, but the knowledge of how to use R, especially the broader context of scripting and data analysis, is not subject-specific and hence students from all disciplines should be readily able to adapt the concepts to their fields.

R is based on another programming language, S, and was first released in 1995. Nowadays it is a very popular choice for performing statistical evaluations and data analysis. One of the main advantages of R is, that a lot of extensions ("packages" in R terminology) have been developed for it that greatly enhance the functionality. R is also free and open source, meaning that everybody can look at the code of R itself, can modify it, and can adopt it for other operating systems and architectures.

RStudio is an integrated development environment for R. R itself basically just runs in a console. RStudio provides a user-friendly graphical interface for writing R code, running it, looking at plots data, or documentation, and much more.

R comes with some basic functionality inbuilt; commonly called 'base R'. However, a lot of people are working with R, and there is almost always multiple ways to do something in R. A lot of packages have been published that provide new functionality or that aim to replace the base R functionality with a more refined approach. One noteworthy example here is the tidyverse (`https://www.tidyverse.org/`), a whole collection of package that aim to make working with R more tidy.

Another advantage of R is related to reproducibility, an important topic in research. R makes it relatively easy to bundle R code, results, and documentation. For example, one could write R code and text in one file and that later is converted into a pdf report that contains the results and figures of a data analysis. One solution to achieve this is called RMarkdown.

During this course we will cover the basics of working with R, handling data with R, how to program in R, how to visualize data, how to calculate statistics, and how to work with RMarkdown.