Statistical Methods for Data Science (SMDS)

(An Introduction)

R. Bellio & N. Torelli

Spring 2018

University of Udine & University of Trieste

General information

Instructors

- RB Ruggero Bellio
- NT Nicola Torelli
- LE Leonardo Egidi (lab)

Timetable & Organization

Instructor	March	April	May	June
NT	8, 12	24	3, 8, 10, 14, 15, 17, 22, 28	5, 7
RB	13, 19, 22, 26, 27	5, 10, 12, 17, 19	29, 31	
LE	20	9, 23	7, 21	4

Office hours

Two possible ways to contact us

1. Right at the end of each class.
2. Via Skype, on demand. In that case, drop us an email at ruggero.bellio@uniud.it or nicola.torelli@deams.units.it, if possible some days beforehand. For Ruggero Bellio, the most suitable time is during his ordinary office hours, on Wednesdays (14-16), but other times are fine as well. For Nicola Torelli on Wednesday 9.30-11.00 is the ordinary office hour.

Aim of the course

From the *syllabus*

> *The course focuses on fundamental elements of statistical inference, along with some principles and statistical techniques useful for the analysis of complex data.*

The central theme of the course will be **statistical modelling** of data, yet the focus will be more on *ideas* and *principles* rather than on details of the statistical methodology.

Mathematical contents will be limited to a healthy minimum.

Role of R software

The *learning by doing* philosophy will be embodied by the constant usage of the R software throughtout the course.

R will be used in two ways:

- In the R laboratory sessions
- In the *R lab* slides used in classes, where R will be used to demonstrate some of the theoretical concepts *on the fly*.

Main textbooks

- S.N. Wood: **Core Statistics**, Cambridge University Press, 2016 (it can be freely downloaded from https://people.maths.bris.ac.uk/~sw15190/core-statistics.pdf)
- J. Maindonald, W.J. Braun. **Data Analysis and Graphics Using R** – An Example-Based Approach (Third Edition); Cambridge University Press, 2010.
- B. Efron, T. Hastie: **Computer Age Statistical Inference** – Algorithms, Evidence, and Data Science. Cambridge University Press, 2016 (available from the authors at https://web.stanford.edu/~hastie/CASI/).
- Additional material, slides presented and other information will be available on moodle.

Information on the final exam

Final evaluation is based on

- homeworks (40%)
- final project (60%).

Homeworks will be assigned each couple of weeks. Final project will be assigned well before the end of the course and will be presented by the students right after the end of the lectures.

Those who do not complete the homeworks or do not present the final project will be then admitted to an oral exam. Oral exams will be scheduled in each of the exam sessions (june-july, september, january-february)