

Statistical Methods for Data Science (SMDS)

(An Introduction)

N. Torelli, L. Egidi, G. Di Credico

Fall 2020

University of Trieste

General information

Instructors

- Nicola Torelli - Leonardo Egidi - Gioia Di Credico (lab)

Schedule and Organization

Lectures are on - Tuesday (9.00-11.00) and Wednesday (11.00-13.00)
(only online lectures) - Friday (9.00-11.00) and Friday (16.00-18.00 Lab)
(room F, Bldg A)

Lectures

	Torelli	Egidi
October	6, 7, 9, 13, 14, 20, 21, 23	27, 28, 30
November	25	4, 6, 10, 11, 17, 18, 20, 21, 24
December	1, 11, 15, 16, 18	

Lab (Di Credico)

16, 30 october, 13, 27 november, 11, 18 december

Possible changes, which would be notified in advance.

Given the present situation there is only a possible way to contact us:

Via MS Teams (or other similar tools), on demand. In that case, drop us an email at

- nicola.torelli@deams.units.it
- legidi@units.it or
- gioia.dicredico@deams.units.it

if possible some days beforehand.

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.

This will give you a deeper understanding of many tools used in AI and ML and more awareness on properties of methods used.

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.

The *learning by doing* philosophy will be embodied by the constant usage of the R software throughout 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/>).

The slides of the lectures, the text of the homeworks plus any announcement related to the course organization will be posted on the UniTS Moodle repository.

Some slides with some basic elements of probability, already covered in the first semester, will be also placed there. They are a **highly recommended preliminary reading**.

Information on the final exam

Final evaluation is based on

- homeworks (20%) Homeworks will be assigned each couple of weeks to groups of about three/four students. The groups will be **formed by randomly chosen students, possibly changing across occasion**. Homeworks have strict deadlines.
- intermediate tests (40%) Intermediate tests will be administered by moodle in two or three occasions during the course
- final project (40%) 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. Here the groups will be formed by three students, freely chosen. Each student has to make 1/3 of the presentation (30 minutes in total).

Those students who do not complete all the homeworks or do not participate to the tests or do not present the final project will have to pass a final test and an oral exam. Oral exams will be scheduled in each of the exam sessions (june-july, september, january-february)