Statistical Methods for Data Science (SMDS)

(An Introduction)

L. Egidi, M. Stefanucci, V. Gioia

Fall 2021

University of Trieste

General information

Instructors

Leonardo Egidi - Marco Stefanucci - Vincenzo Gioia (lab)

Schedule and Organization

Lectures are on

Monday (9.00-11.00) [Aula 2, C5 bldg]

Tuesday (9.00-11.00), Wednesday (11.00-13.00) [Aula Morin, H2Bis bldg]

Lectures

	Egidi	Stefanucci
October	5, 6, 11, 12, 13, 19, 20, 25, 26, 27	
November	9	10, 15, 16, 22, 23, 24, 30
December	6, 7, 14, 15, 20	1

Lab (Gioia)

18 october; 8, 17, 29 november; 13, 21 december

Possible changes would be notified in advance.

Office hours

You may contact us:

- via MS Teams (or other similar tools), on demand.
- via mail at
    - legidi@units.it
    - marco.stefanucci@deams.units.it
    - gioia.vincenzo@spes.uniud.it

    if possible some days beforehand.
- in the office hours:
    - Egidi: tuesday, 11.30-13; thursday 12-13 [room 2.19, 2nd floor, D bldg (DEAMS)]
    - Stefanucci: monday, tuesday 14-16 [room 2.19, 2nd floor, D bldg (DEAMS)]

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.

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/).

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

Please, subscribe to the moodle's course here!
https://moodle2.units.it/course/view.php?id=8393

You will find there also some information about the Microsoft Teams group.

Also material from the pre-course on R is available

Information on the final exam

Final evaluation is based on

- **intermediate tests** (50%) 3 Intermediate tests with multiple choice selection will be administered by moodle in three occasions during the course:
    - 9 november
    - 30 november
    - 20 december
- **final project** (50%) 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 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)