

Statistical Methods (SM)

(Introduction to the course)

N. Torelli, G. Di Credico, V. Gioia

Fall 2024

University of Trieste

General information

Instructors

- Nicola Torelli
- Gioia Di Credico
- Vincenzo Gioia (lab)

Schedule and Organization

Lectures (or labs) are on

- Tuesday (8.00-11.00) (room B - Bldg C7)
- Wednesday (12.00-14.00) (room 1_B - Bldg D)
- Thursday (15.00-17.00) (room Z - Bldg G)
- Friday (15.00-17.00) (room 4_A - Bldg D)

Dates of the lectures Possible changes will be notified in advance.

month	Torelli - Di Credico	Gioia (lab)
September	26	
October	1, 2, 3, 8, 9, 10, 15, 16, 17, 22, 23, 24, 29, 30, 31	8, 11, 22, 25
November	5, 6, 7, 12, 13, 14, 19, 20, 21, 26, 27, 28, 29	15, 22
December	3, 4, 5, 10, 11, 12, 13, 17, 18, 20	6, 19

- Gioia di Credico - Office 2.19, Bldg D
 - Tuesday from 17.00 to 18.30
- Nicola Torelli - Office 2.10, Bldg D
 - Wednesday from 17.00 to 18.30
- Vincenzo Gioia - Office 2.13, Bldg D
 - Friday from 17.00 to 18.30
- or via MS Teams (or other similar tools), on demand. In that case, drop us an email at
 - nicola.torelli@deams.units.it
 - gioia.dicredico@deams.units.it
 - vincenzo.gioia@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*.

Textbook and recommended readings

- The **reference textbook** is
 - A. Agresti, M. Katery: **Foundations of Statistics for Data Scientists: With R and Python**, Chapman & Hall, 2021(Main Text)
- Other more advanced texts are
 - S.N. Wood: **Core Statistics**, Cambridge University Press, 2016 (it can be freely downloaded from the author webpage: [here](#))
 - B. Efron, T. Hastie: **Computer Age Statistical Inference – Algorithms, Evidence, and Data Science**. Cambridge University Press, 2016 (available from the authors webpage: [here](#)).
- A more practical introduction to some topics is in
 - J. Maindonald, W.J. Braun: **Data Analysis and Graphics Using R – An Example-Based Approach** (Third Edition); Cambridge University Press, 2010.

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

Moodle password: SM2024

If you are not automatically enrolled on the Team, you can access it using the code: szdisue

Information on the final exam

Final evaluation is based on

- **Homework** (10%) Homework 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**. Homework have strict deadlines.
- **Intermediate tests** (45%) Intermediate tests will be held in three occasions during the course
- **Final project** (45%) 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/four students, freely chosen. Each student has to take part at the presentation (it should take about 30 minutes).

Those students who do not complete all the homework *or* do not participate to the tests *or* do not present the final project will have to present a final project and to pass oral exam. Oral exams will be scheduled according to the rules for the Department (two possible dates for each session: January-February, June-July, September).

Students from the Master program in “Scientific and data intensive computing”

- Students enrolled in the Master program in “Scientific and data intensive computing” should attend only two third of the entire course (6 credits instead of 9)
- The aim of the course remains the same, but they are required to master the topics covered in the first two months
- Homeworks and intermediate tests will be reduced accordingly
- They will be also required to work on a final project (and possibly join groups with students from the Master program in Data Science and Artificial Intelligence)