Statistics for data science
Project A.Y. 2022/23

Motivation

❑ As a data scientist, you must be able to **understand and reproduce** innovations in the field, which are typically described in scientific papers

❑ **Understanding** means being able to comprehend and communicate to others theoretical and methodological innovation, *using appropriate formal statistical language*

❑ **Reproducing** means being able to re-implement portions of methods and algorithms, or to re-implement portions of the tests and experimental validation, *using the R programming language*

❑ The project of Statistics for Data Science will focus on practicing with those skills, applied to one of several topics reported in topic papers

    ❑ Topic papers available in Teams under the Files/Project tab

    ❑ The list of topic papers may be expanded up to the end of the course

Groups and topic paper

Students autonomously cluster in groups of up to 3 members

❑ They select *at least* 3 topic papers and send an email to salvatore.ruggieri@unipi.it

❑ The teacher assigns to the group <u>one</u> of the selected topic paper

❑ Deadline: **31 May 2023 at the latest, but the earlier the better**

The group studies the assigned topic paper and deliver to salvatore.ruggieri@unipi.it

❑ A presentation of up to 15 slides

❑ An R script with some re-implementation of methods and/or experiments

❑ Both presentation and R script may focus on a portion of the topic paper – *no need to cover everything*

❑ Each member of the group register to an exam date (appello), only to ensure filling the student's questionnaire

❑ Deadline **31 August 2023 at the latest , but the earlier the better**

Discussion and grades

Project discussion and oral discussion

- ❏ The group gives a 20 min talk with Q&A using the presentation slides
  - ❏ Dates of project presentation: **31 May, 21 June, 11 July, 4 September**
- ❏ Each member of the group takes an individual oral discussion with questions on the project and on the course theory/R programming
  - ❏ Dates of individual oral discussion: **shortly after the project presentation**, ideally *on the same day* (depend on the number of project discussions)

Grade composition

- ❏ up to 10 points: slides and R script quality
- ❏ up to 10 points: project presentation and Q&A clarity
- ❏ up to 10 points: individual oral discussion

Q&A

Q: What is the language of slides/presentation?

A: Slides must be in English, presentation can be in English or Italian

Q: When do the slides+R script has to be delivered?

A: They must be sent by email at least 48 hours prior to the date of project discussion.

Q: Can the project be discussed after September 2023?

A: No, students with assigned project that do not discuss within September 2023 will have to do the normal written+oral test.