Statistics for data science
Project A.Y. 2023/24
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 topics reported in selected papers.
  - Selected papers available in Teams under the Files/Project tab.
  - The list of topic papers may expand up to the end of the course.
Group formation and paper assignment

Students autonomously cluster in groups of up to 3 members

- They select \textit{at least} 5 papers and send a ranked list to salvatore.ruggieri@unipi.it
- The teacher assigns to the group one of the selected paper
- Deadline: \textbf{24 May 2024 at the latest, but the earlier the better}
The group studies the assigned paper and deliver to salvatore.ruggieri@unipi.it

- A presentation of up to 15 slides
- An R script with re-implementation of some methods and/or experiments
  - Both presentation and R script may focus on a portion of the paper – no need to cover everything

At the time of delivery, each student must have registered to an exam date (appello), only to ensure filling the student’s questionnaire

Deadlines (no extension will be granted!):

<table>
<thead>
<tr>
<th>Batch</th>
<th>Delivery deadline</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>27 May h. 23:59</td>
</tr>
<tr>
<td>2nd</td>
<td>24 June h. 23:59</td>
</tr>
<tr>
<td>3rd</td>
<td>22 July h. 23:59</td>
</tr>
<tr>
<td>4th</td>
<td>9 Sept h. 23:59</td>
</tr>
</tbody>
</table>
Project and oral discussions

The group gives a 20 min talk with Q&A using the presentation slides and R script

- All members of the group must be present, no split discussions
- Q&A will regard the project only

Each member of the group takes the individual oral discussion, which may include Q&A on the project as well as Q&A on the course theory/R programming

- Students must demonstrate to be able to summarize both the theory and the software related to any of the lessons using the slides and R scripts of the lessons.

<table>
<thead>
<tr>
<th>Batch</th>
<th>Delivery deadline</th>
<th>Group presentation</th>
<th>Individual oral</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>27 May h. 23:59</td>
<td>29 May h. 9:00</td>
<td>29 May h. 14:00</td>
</tr>
<tr>
<td>2nd</td>
<td>24 June h. 23:59</td>
<td>27 June h. 9:00</td>
<td>27 June h. 14:00</td>
</tr>
<tr>
<td>3rd</td>
<td>22 July h. 23:59</td>
<td>24 July h. 9:00</td>
<td>24 July h. 14:00</td>
</tr>
<tr>
<td>4th</td>
<td>9 Sept h. 23:59</td>
<td>11 Sept h. 9:00</td>
<td>11 Sept h. 14:00</td>
</tr>
</tbody>
</table>
Grade composition

Grade composition
- up to 5 points: slides writing, organization, clarity
- up to 5 points: R script quality, organization, documentation
- up to 5 points: project presentation (overall)
- up to 5 points: project discussion (individual)
- up to 10 points: individual oral discussion

This means
- 15 points assigned collectively to the group + 15 points assigned individually to each member of the group
- 10 points assigned to the offline work + 10 point assigned to the reporting of such work + 10 points assigned to the knowledge of the topics of the course
- 20 points assigned to the project + 10 points assigned to the oral discussion
Q&A

Q: What is the language of slides/presentation?
A: Slides must be in English, presentation can be in English or in Italian

Q: Can the project be discussed after September 2024?
A: No, students with assigned project that do not discuss within September 2024 will have to do the normal written+oral test.

Q: Can the project be discussed remotely?
A: No, only in presence. The room for the discussion will be communicated after the delivery of the project.