POLITECNICO
MILANO 1863

Nonparametric Statistics Deliverables Guidelines

Prof. Simone Vantini
Prof. Francesca Ieva
Dr. Alfredo Gimenez Zapiola

Nonparametric Statistics – Math Engineering
October 4th, 2024

Individual Final Exam

Part I. *Homework discussion.*

- A problem will be given to you by the end of the course (December) and it will remain the same for the entire academic year.
- At **each examination test** you will have to show and explain to Dr. G. Zapiola (using your laptop) part of the code that you will have previously prepared at home.

Part II. *Oral examination.*

- At **each examination test** you will be asked to answer one question by Prof Ieva pertaining to Block I and III.
- At **each examination test** you will be asked to answer one question by Prof Vantini pertaining to Block II and IV.
- You will be free to use a pen and a paper sheet to answer your questions.

The Project: Features

For the project you are expected to apply the techniques you have seen in the course to tackle a real-world problem.

- The project will be part of the evaluation! (1/3)
- To be conducted in teams of **3 ± 1** students
- The data should be:
  - Multivariate and/or functional
  - "Very interesting" for your group
  - Should come up with interesting and meaningful research questions possibly solved using the methods presented during the course.

The Projects: deliverables 1/2

- **End of October:** each team has to send an email to the class representative (matteo.cevolani@mail.polimi.it; ***https://chat.whatsapp.com/LegikDuYVnqJYy7yjvoD8O***) the following information:
  - List of team members
  - Data set (with description and link, if available)
  - When the project will be presented (Summer or Winter session)

The class representative will then collect and send the resulting file to alfredo.gimenez@polimi.it

- **Mid November:** each team will upload a 4-slide presentation containing the following information
  - Slide 1: Title and team members
  - Slide 2: Dataset description
  - Slide 3: Tentative goals of the project
  - Slide 4: Tentative analytical workflow (i.e., which tools you plan to use)

The Projects: deliverables 2/2

Mid December: WIP oral presentation (5 minutes) in which each group will have to briefly describe the chosen dataset and introduce the goals and the roadmap of the project (these could always be updated and modified while the project is in progress)

To be defined: final presentation (10/15 minutes) in which results should be presented by all the members of the group. Concurrently, you are required to hand over (dedicated webeep delivery folder will be created)

- Presentation
- Brief report (~20 pages long)
- Relevant code

The Project: Evaluation

The projects are evaluated looking at:

- How meaningful and clear the research question is
- How well you structure your «process» to answer the research question
- How well you decide the methods to use during your «process», and how well you implement them
- How compelling your presentation will be
- How tidy and well structured your code is
- To deem a project as EXCELLENT we expect your project to include
  - Impressive and useful visuals
  - Nonstandard methods
  - (A tiny bit of) methodological innovation
  - ...but don’t get carried away too much!
    We still want you to use methods we teach in NPS

The Project: Expected Outline

1. **Problem statement – Why this problem?** Why is this problem relevant? To whom is this problem relevant? What are the open themes in the area of knowledge you are exploring? How is the data you're using able to solve this kind of questions?
2. **Research questions** – Starting from 1., what are the questions are you asking to your dataset? What kind of methods do you want to use for these questions? Why Nonparametric methods?
3. **Analysis** – What are numerical results? Are they reliable? Are they robust? Can you trust them?
4. **Conclusions** – What are your answers to the questions in 2.? What are the real world implications of your results?

This is actually useful also for learning how to develop a proper statistical analysis of a dataset, either for academic or professional purposes

Datasets: some Socio-Economic Sources

POLITECNICO MILANO 1863

- https://ec.europa.eu/eurostat/home
- https://www.istat.it/
- https://dati.comune.milano.it/
- https://data.jrc.ec.europa.eu/
- https://data.oecd.org/
- https://www.ecb.europa.eu/stats/html/index.en.html
- https://data.un.org/
- https://iiasa.ac.at/web/home/research/researchPrograms/Energy/SSP_Scenario_Database.html
- https://ukdataservice.ac.uk/

Datasets: some Biosciences Sources

- https://biolincc.nhlbi.nih.gov/teaching
- http://featureselection.asu.edu/datasets.php
- https://mimic.physionet.org/
- https://higgi13425.github.io/medicaldata/
- https://hbiostat.org/data
- https://xenabrowser.net/datapages/

Datasets: general purpose websites

- http://mmds.org/
- https://www.kaggle.com/
- https://archive.ics.uci.edu/ml/datasets.php
- https://github.com/rfordatascience/tidytuesday

Clearly, you are free to propose anything you think could be interesting for your group, but always remember to ask/look for dissemination and confidentiality policies!

Project examples

POLITECNICO MILANO 1863

A very good project!

Project examples

POLITECNICO MILANO 1863

A not so good one...

POLITECNICO
MILANO 1863

Each team will receive a mark in a scale from 0 to 30

The project evaluation will be valid for the entire academic year

Questions?