Nonparametric Statistics Project Guidelines

Prof. Francesca Ieva (francesca.ieva@polimi.it)

Dr. Andrea Cappozzo (andrea.cappozzo@polimi.it)

Nonparametric Statistics – Math Engineering

Milano, 29/09/2022

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 that can be possibly solved using the methods presented during the course.

The Projects: deliverables

- **End of October:** each team has to send an email to the class leader (lorenzo.angiolini@mail.polimi.it, or telegram link) containing the following information

    a) name and email address of the team leader

    b) list of the team members

    c) the preliminary title of the project

  The class leader will then collect and send the resulting file to andrea.cappozzo@polimi.it

- **Mid December:** WIP oral presentation (5 minutes) in which one or two members of each group will have to present all the team members, 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)

- **Second half of January/ first half of February:** 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
    - Relevant code.

The Project: Evaluation

The projects are normally 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 those

- 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? 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 you're 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

- 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!

A very good project!

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?