

# Project

---

## Task

Choose an existing dataset or collect your own data, and carry out data analyses to explain the relationships among the variables involved. You can also compare performance of different statistical tools for prediction or classification, using the chosen dataset.

## Teams

For this project you are supposed to work with 1-2 persons and submit a joint report. If you cannot find a team member, they will be assigned to you. (This is not recommended. Try to find a student who is compatible with your work style.)

## Grading policies

Team members will receive the same grade for the project and it is up to you to make sure that the work is shared equitably. The total points of the project is 100 points, which can be divided into three parts:

1. Project Proposal (20 pts).
2. Presentation (40 pts): each team will give a 20 minutes presentation of the project; (dates to be assigned)
3. Final report (40 pts)

## Schedule

See syllabus.

## Methods

You can use methods we covered in class or other statistical learning methods, with preference for more advanced methods.

## Data

Find your own data set online, you will find plenty;

Popular collections of publicly-available datasets:

- UCI Machine Learning Repository
- Kaggle
- Academic Torrents (shares large datasets via bit torrent technology)

- StatLib (This is an older collection of data which is no longer updated.)

Some institutional collections of data:

- World Bank
- World Health Organization

## Guidelines

- The proposal should give information about the team, description of the data, potential research questions and possible methods to use. The proposal should not exceed one page. The proposal, preliminary and final reports should be uploaded via Google form: [Google Form for Project files](#).
- The preliminary report is a draft of the final report.
- The final report should not exceed 6 pages, including figures and tables, and must begin with an appropriate title highlighting your choice of topic and analysis.
- The final report should include:
  - Description of research questions / issues. The significance of the problems.
  - Description of the data.
  - Preliminary studies: data visualization, dimension reduction, feature extraction, feature selection etc.
  - Statistical analysis:
    1. Methods: what analyses were done and why. If there is any challenge in analysis, describe your approach to tackle the problem.
    2. Results: A small number of well-designed and tailored tables and graphics may be appropriate.
  - Discussion/Conclusion: Convey your findings to a broad audience.
  - Try to avoid including too much of the software output. Supporting code can be kept in appendix or a separate document.
- The final report will be evaluated on the basis of the following criteria
  1. How interesting is the dataset and the research idea.
  2. How well the dataset was prepared for analysis.
  3. Quality of the statistical analysis
  4. Quality of material presentation [The grammar, orthography and style matter.]
- The presentation in class can be done by a single team-member or by all team members, as you prefer it. It should include clear description of the data, research question(s) and findings. The evaluation criteria are similar to the criteria for the final report, with emphasis on the quality of presentation.

From:

<http://www2.math.binghamton.edu/> - **Department of Mathematics and Statistics, Binghamton University**

Permanent link:

[http://www2.math.binghamton.edu/p/people/kargin/math457\\_fall2023/project](http://www2.math.binghamton.edu/p/people/kargin/math457_fall2023/project)

Last update: **2024/06/28 16:07**

