

Course Group Project

In this project, each team will select a research paper in applied statistics, focusing on papers where the application is central and provides the primary motivation for the work. The team is expected to understand the application problem, the statistical methods proposed and used, and the analysis conducted in the paper. The team will summarize their understanding in both a written report and a presentation. While implementing the methods from the paper is encouraged, it is not mandatory.

Project Stages

Stage 1: Team Formation and Paper Selection (Week 1-2)

- Form a group of three students. A fourth member can be included, provided they are either in the Masters in Applied Statistics program or a master's or PhD student from another department.
- Select a paper from the past 20 years with an “application focus”, where applications are central to the paper and involve relevant datasets. Suitable journals include:
 - Annals of Applied Statistics (AOAS): <https://projecteuclid.org/journals/annals-of-applied-statistics/issues/>
 - Journal of the Royal Statistical Society Series C (JRSS-C): <https://academic.oup.com/jrsssc/issue>
- Papers should have publicly available datasets or be reproducible using common statistical software. You will be required to download the dataset.

Stage 2: Reading and Planning (Week 2-3)

- Each group member will read the paper, and the group will jointly submit a one-page proposal. The proposal should contain brief descriptions (two or three lines) of the following:
 1. Problem Description and Modeling Objective
 2. Data Description and Availability of Dataset
 3. Model and Methods Description

(see description below for what is expected for each item)

Stage 3: Report Writing (Week 3–5)

The group will compile their findings into a detailed report. This report should be structured as follows:

- Problem Description and Modeling Objective: Provide a clear statement of the problem being investigated in the paper. Clearly state the specific objective of the statistical modeling.
- Data Description: Provide a detailed overview of the dataset used (from the main application), including information about the variables in the dataset. If the paper involves a regression model, identify the response variable and predictor variables. Include the first few rows of the datasets in the report as a table. This section should also include essential exploratory data analysis (EDA) which **you perform**, including univariate, bivariate, multivariate (e.g., visualizing potential interaction effects) analyses. The EDA should include visualizations (with informative titles and labels) and summary statistics with descriptions of the main results from the visualizations.

- Model and Methods Description : Clearly describe the proposed model in the analysis. Outline the statistical estimation methods. For the proposed inference methods, provide an algorithm (e.g., Algorithm 1) with sufficient detail to implement the methods, using the input described in the Data Description section.

Algorithm 1: An example algorithm

Data: $x \in \mathbb{R}, n \in \mathbb{N}$
Result: $y = x^n$

```

 $y \leftarrow 1;$ 
 $X \leftarrow x;$ 
 $N \leftarrow n;$ 
while  $N \neq 0$  do
    if  $N$  is even then
         $X \leftarrow X \times X;$ 
         $N \leftarrow \frac{N}{2}$ 
    else
        if  $N$  is odd then
             $y \leftarrow y \times X;$ 
             $N \leftarrow N - 1;$ 
        end
    end
end

```

- Simulation Studies (or Reproducing Results): Choose one of the following:
 - (Simulation study) Describe the simulation setup, including any competing methods, and summarize the findings.
 - (Reproducing results) Pick a simulation setup (which can be simpler than the one used in the paper) and apply the method you implemented to the simulated data.
- Results: Summarize the main findings from the statistical analysis, and explain how the fitted model was used to address the scientific questions.
- Conclusion: Provide a brief summary of your report. Discuss the strengths and weaknesses of the paper, as well as any challenges you faced, including issues related to data access, differences in interpretation of methods, difficulties reproducing the analysis, or any other technical issues.
- Author Contribution Statement: Include the statement which clarifies each group member's contribution to the project (e.g., <https://www.elsevier.com/authors/policies-and-guidelines/credit-author-statement>).

Stage 4: Presentation (Week 6)

Each team is required to submit presentation slides, with a maximum of 8 slides, and prepare a presentation. Each member should roughly equally contribute to the presentation. More details about the presentation will be provided later.

Grading Criteria

Component	Weight
Written Report	60%
Presentation	10%
Peer Evaluation	20%
Contribution Evaluation	10%

- **Bonus Points:** Up to 10% extra credit will be awarded for groups that choose to reproduce the methods from the paper.

- **Peer Evaluation:** Each student will be required to read another group's report and submit grades based on the provided rubric. The anonymized grades from other teams, together with mine, will be shared to the group.
- **Contribution:** All team members are expected to contribute equally to the completion of this assignment. **There will be an anonymous survey which surveys each group member's participation.** While different team members may have different backgrounds, every team member is expected to participate in developing analyses, and understand how the codes and approaches work.