495R Final Project

Winter Semester, 2025

Your final exam for this class will be a project consisting of a complete analysis of a real-world functional data set. This project will be done in groups of 2-3 people, and will be submitted in three parts:

I. A project proposal due February 28

II. A final report due April 16

III. A presentation during our scheduled final exam time (April 22 at 3pm-6pm)

Part I – Project Proposal (Due: February 28)

For the first report, you will need to form a group and, together, decide on the data set you will analyze. Several suitable data sets are available on Canvas for you to choose from. If you have an alternative data set you would like to use, you must get it approved by Prof. Petersen **no later than Tuesday, February 18**.

Once your group settles on a data set, you will draft a short report (3-4 pages) that includes the following:

- A summary of an appropriate exploratory functional data analysis. This should involve mean and covariance estimation, FPCA, and outlier detection/removal. Be sure to include appropriate plots and well-documented code.
- 3-5 scientific questions that you propose to answer using the data, at least two of which should involve other covariates (functional or non-functional) that were not the focus of your exploratory analysis. Be sure to explain why these questions might be important to the scientist who collected the data. You do not need to propose a way to answer these questions yet.

Workshop Days: On **February 25 and 27**, Dr. Petersen will meet with each group for 20-25 minutes to discuss their progress and ideas, as well as suggest a method for them to study and implement to answer 1 or 2 of their scientific questions.

The pre-approved data sets are:

Data Set	**R Package/File**
gasoline	refund
tecator	fda.usc
CandianWeather	fda
gait	fda
DTI	refund
content	refund
pbc	survival

Part II – Final Report (Due: April 16)

The final report should be written in either Markdown or Quarto. It should be **no more than 10 pages** and have the following sections:

- **Introduction:** Describe the dataset, state the 1-2 scientific questions to be answered, and provide a high-level overview of the methodology with appropriate references.
- **Methods:** Provide a summary of the method(s) you studied, with sufficient mathematical notation to explain population parameters, estimators, and necessary components.
- **Data Analysis:** Apply the methods to the dataset. This should include well-documented R code for running the methods, as well as tables and plots showing the results. Unrelated code (e.g., preprocessing) should instead be placed in the appendix.
- **Conclusion:** Summarize your findings. Discuss what your analysis revealed about the question(s) you asked, and how (un)certain you are of that answer. What follow-up questions would you ask as a result?

Your report will also have an appendix, which will **not** count towards your 10 pages. The appendix should include:

- Your project proposal (corrected for spelling/grammar and any parts of its analysis that were modified).
- Any additional plots related to your analysis that couldn't fit in the main report.
- All of your code, which should be well-documented. You can use the following code chunk to print all of your code:

```
```{r ref.label=knitr::all_labels(), echo =TRUE, eval = FALSE, include = TRUE}

```
```

Part III – Presentation (April 22, 2025, 3-6 PM)

Each group will be given 25 minutes to present the contents of their final report. The presentation should be structured roughly as follows:

- **Introduction (8-10 minutes)**
- **Methods (8-10 minutes)**
- **Analysis and Conclusion (5-7 minutes)**