PH240C Final Project Template
Title of Your Project

Your team members' name

April 3, 2025

Abstract

Please use up to 6 sentences to describe your project. Including te motivation of your scientific questions, how you have solved the questions with what methods, and your final conclusions.

1 Introduction

Please use this section to introduce the overarching goal of your project with the following components:

1. What dataset you have worked with? Please introduce the data source description used in this paper: Feng et al. (2021). If you are using the UK biobank data, please refer to their website for description.

2. What scientific questions do you want to investigate from the dataset? What are your motivations? Please include a comprehensive literature review.

3. What statistical methodology are you currently using? Why do you think this method is a good fit for answering the raised scientific question? Please explain and include a comprehensive literature review about the adopted statistical methodology.

4. A competitive final project will not only use existing methods, but also customize them to answer your raised question.

5. How is your report organized? What are the content in each section? Please include a paragraph talks about this.

Note:

1. Your final project will be graded based on whether you have discussed each points in different sections.

2. You should not change the font and spacing for this template when you submit your final proposal.

3. Maximum number of pages for this final proposal should be 15 pages, excluding reference.

4. If you plan to answer the bonus question and provide statistical ganrantees of your method, you are allowed to have maximal 25 pages. All proofs should be given in appendix, rather than the main paper.

2 Dataset description

Please introduce your data with basic descriptive statistics with the following components:

1. Did you further subseting the dataset? Why?

2. What are the demographics in the data? Sample size, number of attributes, and missingness need to be reported.

3. Based on the given data, please formulate your scientific question in mathematical/statistical terms. What are your objective functions?

3 Methodology

Based on the data description, please motivate the considered model or methodology with the following components:

1. What model have you proposed to answer your raised question? Why this model make sense to you?

2. What statistical methods have you chosen based on the proposed method? Why this method is a natural fit? How does it compare the model used in the existing literature?

3. Bonus question: What can you say about the statistical validity of the proposed method? Why do you think so? Can you derive statistical ganrantees for the proposed method?

4 Real data analysis

Based on your model and proposed statistical methods, please summarize the following points:

1. Have you done any pre-processing on your current data? For example, centering the variables, and standardizing covariates.

2. What packages have you used in your project? Or you have written your own code?

3. What tuning parameter have you chosen in your analysis? Have you used default ones in the package? If so, why do you think you can use the default tuning parameters?

4. If you have written your own code to implement the proposed method, please specify your rational to the adopted tuning parameter.

5. How do your data tell you about your question? What are your analysis procedures? What are your conclusions?

6. Bonus question: Do your result differ/align with existing literature?

5 Conclusion/Discussion

Please use one short paragraph to summarize your conclusion. Or you can use this space to discuss some additional thoughts.

References

Long Feng, Xuan Bi, and Heping Zhang. Brain regions identified as being associated with verbal reasoning through the use of imaging regression via internal variation. *Journal of the American Statistical Association*, 116(533):144–158, 2021.