

Missing data project description

February 2025

1 Due date, submission instructions, grade breakdown

Report	Max Length (w/o refs)	Due date	Grade
Project proposal	1 page	3/6 11:59pm	10% of final grade
Final project report	8 pages	3/20 6:00pm	40% of final grade

The **final report** will be due on 3/20 at 6pm PT, and will be worth **40% of your final grade**. The **project proposal** will be due on 3/6 at 11:59pm PT, and will be worth **10% of your final grade**.

Submission of both will be through Gradescope, where I'll create new assignments for the proposal and the final project.

2 Project overview

The final project for this course is a 4 to 8 page report (not including the reference page(s)) focusing on one of several areas:

1. Data analysis
2. Methodology investigation
3. Pedagogical material

All projects should have some implementation in R Code. That makes RMarkdown, Quarto, or Knitr useful tools for this report.

2.1 Data analysis

Define a research question that can be answered with a dataset that you have access to. The dataset will likely have missing, or coarsened values; any dataset that has no missing data is suspect. Use the dataset to answer your research question, and use techniques from the class to handle the missingness/coarseness.

Run sensitivity analyses to determine if changing assumptions from MCAR, to MAR to MNAR affects your conclusions. Write up your conclusions in a report, and detail the methods you used, and the decisions you made about how to model your data and the missingness process. Argue why/why not ignorable missingness is a good assumption. You can approach this project in two ways: you could find an interesting dataset, and then think of a question that could be answered by analyzing the dataset, or you could think of a question you're interested in and look for data to answer the question. The first route is easier, but the second route is possible too.

2.2 Methodology

We've studied and will study different methods for dealing with missing data: Joint modeling, multiple imputation, latent variable modeling, EM, Bayes. If you like, you could do a deep dive on a methodology in one of these areas from a theoretical or simulation-study standpoint. If you go with this sort of project and you're doing a simulation study, make sure that you design your simulation study to have a high-enough power to detect whatever interesting thing you're investigating. This may require using the computing cluster. If you're doing a theoretical deep dive on the material, you'll also want to run a simulation study to ensure that your theoretical results hold in different scenarios.

2.3 Pedagogical material

If you're interested in learning more about a certain topic we covered in class or a topic you wished we had covered in class, you can develop pedagogical material related to this topic. The material should have some coding aspect to it because missing data is an applied discipline.

3 Project proposal instructions

Describe the project you plan to do using a maximum of one page. Be specific about what dataset you plan to use, what questions you want to ask, and the methodologies you think you might use. Please come talk to me about your idea if you have any questions.