

STAT 4640/7640 - Introduction to Bayesian Data Analysis

Final Project - Spring 2022

The final project will consist of a Bayesian data analysis and report. You will complete this project on your own. Your data set, question(s) of interest, and analysis must be unique. As part of this project, you will write a brief project proposal describing your planned analysis and post to the discussion board on Canvas. Don't begin working on your project until I have approved your proposal. You must post your proposal no later than April 27th by 5pm. I will provide any necessary feedback on your project proposal no later than Friday, April 29. The sooner you post to the discussion board, the sooner I will provide feedback and you can begin working on your analysis and report.

Final Project Proposal Find a dataset that you are going to analyze for your final project. Your dataset and analysis cannot be the same as a classmate. Make sure your dataset contains enough observations to draw reasonable conclusions from your model.

1. Describe the data you are going to model and the scientific question(s) of interest.
2. Propose at least one statistical model that you will fit in a Bayesian Framework. Identify possible prior distributions for all model parameters.
3. Describe the posterior inference that you will obtain, and how it will address the scientific question(s) of interest.

Final Project You will turn in a written report of your statistical analysis and results. Your report should be at least 2 typed pages, but cannot exceed 4 typed pages. Make sure to include the following:

- Clearly state the research question and the data used in the analysis.
- Describe the statistical model(s) you used, including all prior distributions and choices of hyperpriors.
 - Your analysis **MUST** include either a sensitivity analysis or a model comparison. Feel free to be creative. If you are doing a model comparison, make sure to include your metric(s) of comparison (e.g., MSE, LS, DIC). If you are doing a sensitivity analysis, include either or both DIC and WAIC in your comparison.
- Describe your method of inference (e.g., Gibbs sampler, Metropolis-Hastings algorithm). Provide a concise summary of your model fitting, including the number of posterior samples obtained, convergence diagnostics, etc.
 - DO NOT include trace plots of your parameters or tables of effective sample sizes. You are responsible for assessing convergence of your algorithm – say what you did and why you are comfortable with your results.
- Provide **ONLY** relevant posterior inference that addresses your scientific questions (e.g., figures, tables, posterior probabilities). This includes the results of your sensitivity analysis or the comparison between your models. Make sure to interpret your results in the context of the study. Your posterior inference **MUST** include a notion of statistical significance in a Bayesian context (e.g., interpretation of coefficient estimates in a regression model, hypothesis test, etc).
- Write a brief summary of the results of the analysis, providing an answer to the scientific question posed above. Include at least one idea for a future analysis that you could do as an extension to the one presented.
- Data and R-code should be available to me upon request.

You will also present your analysis in class. Presentations will occur during our final exam period, Tuesday, May 10th, 12:30p-2:30p. Presentations should be approximately 5 minutes, and summarize your analysis.

Due dates

- Project Proposal due by 5pm on Wednesday, April 27.
- Final report due at 12:30pm on Tuesday, May 10.
- Presentation occur during the final exam period, 12:30-2:30pm on Tuesday, May 10.