

Guidelines for STAT 7630 Project, Fall 2024

You'll be working on a dataset of your choice, ideally one that's well-suited for Bayesian regression analysis.

Select a dataset with real-world data that holds personal or academic interest to you, and which is compatible with one of the Bayesian regression methods we covered in STAT 7630. Your choice of Bayesian regression approach will hinge on the nature of your response (dependent) variable. If it is continuous and approximately normal (or can be transformed to approximate normality), you may employ normal Bayesian linear regression. For count data, Poisson or negative binomial regression could be fitting. For binary responses, logistic regression is recommended. As we will be covering normal linear regression first, starting with a continuous response variable allows you to begin analysis earlier.

The ideal dataset should include several candidate predictor (independent) variables, enabling you to apply Bayesian model selection techniques to determine the "best" model.

You may work individually or in pairs/groups of up to three. Given that team-based data analysis is common in professional environments, group work is encouraged. The project deliverables consist of two components:

1. **A concise typed report** (around three pages but no longer more than 5 pages, excluding plots, graphs, and R output) that summarizes your model choices, analysis, and findings. Include the following as relevant to your data and chosen model:

- Introduction and description of the dataset
- Model statement, specifying data and prior distributions
- Explanation of prior parameter selections
- Summary results for relevant point estimates, interval estimates, or predictions
- Overview of variable selection method (if applicable)
- Summary of model diagnostics, addressing model adequacy and, if relevant, MCMC convergence
- Any adjustments made to improve model fit, if necessary
- Final conclusions drawn from the data analysis

2. **A short in-person presentation** (approximately 10-20 minutes per presentation) in which you or your group discuss your dataset, Bayesian model, and key conclusions. This component aims to develop your ability to clearly explain technical data modeling while also allowing classmates to gain insights into each other's projects.

By end of November, please submit (1) the names of your group members (or indicate if working individually) and (2) a description of your selected dataset, including brief descriptions of the response variable, candidate predictors, and initial "prior beliefs." This submission will be the first graded component of the project.

The full project report is due **Sunday December 9, 9 pm**. Please submit your report as a pdf file (via email).

Project Presentations: The presentation will be in class during the scheduled final exam time (which is **Tuesday, December 10, 8-10 am**). Assessment will focus solely on content, so prioritize planning and clarity over production quality. If working in a group, each member should contribute to the presentation.

The length of the presentation will be adjusted for the size of the group. The allotted

presentation time will be 10, 15, and 20 minutes for one-person, two-person, and three-person projects, respectively. For a multi-person project, each person will be presenting some part of the project highlighting their contributions. The specifics of the presentation times for each group will be provided in a separate email.

Peer Evaluations: As part of the project, you are required to review and provide feedback on a report that I will send to you. The report you receive will be anonymized to ensure confidentiality. Your completed evaluation report is due by **Thursday, December 12, at 12:00 PM (midday)**. Please note that peer evaluation is an individual activity, meaning each person will be assigned a different project to review, and the assessments will be completed individually. More instructions will be provided on this later.

Project Evaluation Criteria

Dataset Selection (10 points)

- Appropriateness of the dataset for Bayesian regression analysis.
- Relevance of the dataset to personal or academic interests.
- Clear description of the response variable and candidate predictors.
- Justification for the choice of the Bayesian regression method based on the dataset.

Report Content and Analysis (50 points)

1. **Introduction and Dataset Description (10 points):**
 - Clarity and conciseness of the dataset description.
 - Explanation of the research question or objective.
 - Relevance of the data to the project goals.
2. **Model Statement and Priors (10 points):**
 - Clear and accurate specification of the Bayesian model and data structure.
 - Justification for the choice of prior distributions and parameters.
3. **Results and Interpretation (15 points):**
 - Presentation of point estimates, interval estimates, and/or predictions.
 - Clarity and depth of interpretation of the results.
 - Relevance of conclusions drawn to the data and research question.
4. **Variable Selection and Diagnostics (10 points):**
 - Explanation of the variable selection process (if applicable).
 - Inclusion and interpretation of model diagnostics, such as model adequacy or MCMC convergence.
 - Discussion of adjustments made to improve model fit (if necessary).
5. **Presentation of Data (5 points):**
 - Effective use of tables, graphs, and other visual aids to support analysis.
 - Inclusion of key outputs without overwhelming the narrative.

Presentation (20 points)

- Clarity and organization of the presentation.
- Effective communication of the dataset, methodology, results, and conclusions.
- Engagement with the audience and ability to answer questions effectively.
- Equal contribution and participation by all group members (for multi-person projects).

Peer Evaluation (10 points)

- Thoughtfulness and constructiveness of the feedback provided.
- Adherence to provided instructions for the peer review.
- Completeness of the evaluation report.

Timeliness and Adherence to Guidelines (10 points)

- Submission of group members and dataset description by the deadline.
- Adherence to format and length requirements for the report.
- Timely submission of the final report and presentation.

Bonus (Optional, up to 5 points)

- Innovative or particularly insightful application of Bayesian methods.
- Use of advanced techniques or creative adjustments to improve model fit or interpretation.
- Exceptional clarity or depth in presenting technical aspects of the project.

Total: 100 points (+ optional 5 bonus points)