

Preliminary Instructions for the STAT 535 Project, Spring 2024

You will analyze a data set of your choosing that is suitable for a Bayesian regression analysis.

The data set could contain real data that is of personal or academic interest to you and that fits one of the methods of Bayesian regression analysis studied in STAT 535. Which Bayesian regression method you choose will depend on the nature of the response (dependent) variable: If the response is continuous and approximately normal (or can be transformed to be approximately normal), then you can use normal Bayesian linear regression. If the response variable is a count, then Poisson regression (or negative binomial regression) may be appropriate. If the response variable is binary, then logistic regression could be used. Note that we will cover normal linear regression before covering those other methods, so if you pick a continuous response variable, you could get started a little earlier.

Ideally, your data set should have multiple candidate predictor (independent) variables, so that you can employ some form of Bayesian model selection to choose the "best" model.

You may work individually, or in groups of two or three people. Since it is very common in the working world to do data analysis in teams, I strongly encourage you to work in a group! The end product will be two things: (1) A typed report (no longer than 5 typed pages, not counting plots/graphs or R output), summarizing your model choices, analysis, and conclusions.

Some things to include (depending on the data set and choice of model) might be:

- An introduction and discussion of the data set itself
- A statement of the model, including choice of data distribution and choice of prior distribution
- Discussion and justification of prior parameter choices
- Summary results for any relevant point estimates, interval estimates, or predictions
- Discussion of variable selection approach (if appropriate)
- A summary of model checking, both in terms of checking for model adequacy and (if appropriate) diagnosing MCMC convergence issues
- Any remedial action you took to fix an inadequate model, if applicable
- Your overall conclusions about the data, based on your analysis

(2) I also want you to turn in, for the rest of the class to view, either:

-a two-page typed summary of your project's data set, modeling choices, and major findings

OR

-a video file of yourself or your group making a very brief oral presentation about your data set, Bayesian model, and conclusions. This is to (1) give you practice in speaking clearly and understandably about technical data modeling, and (2) allow others in the class to learn about the projects of their classmates.

The videos are meant to be short (like around 10 minutes at the most), and they do NOT have to be high-tech. For the video, I will NOT grade you on how high-tech it is; I will only consider the content of what you say. So you should plan out what you'd like to talk about and practice it; that will be more important than making a fancy presentation. And if you work in a group, each group member should speak at some point during the video presentation.

Part of the project grade will require you to watch and comment on some of your classmates' written summaries or videos (they will be available on a shared site in Blackboard that I will set up). I'll decide on the details of this once I know how many projects there are, but you'll be required to comment on a certain number of presentations, and you can comment on more for a little bit of extra points.

I would like you to send me **by March 18** (1) the names of your group members (or whether you will be working as an individual), and **by March 25** (2) a description of the type of data set you will be studying (with brief descriptions of the response variable and the candidate predictor variables, along with some initial "prior beliefs"). Sending these things will be the first part of the project grade.

The project itself (full report, and video or written summary) will be due in **mid-April (a bit before the end of the semester, so that there will be time for you to comment on others' videos or summaries)**. An exact due date and specifics about grading will be posted later.