

STT 465: Final Project

This is an **INDIVIDUAL** project towards your final exam grades. It counts for 15% of your final course grade and graded out of 100 points.

Since we are considering this project as a final exam, we should uphold all Michigan State University [plagiarism policies](#). That means all individual should complete and submit their own work independently.

Instructions:

Please read the following instructions carefully and make sure you use the correct data set to do your own project. If you use a wrong data set for the project, you will automatically lose 40 points.

1. There are 3 different data set and subsequent projects:
 - (i) If the first initial of your LAST name is **A through J**, you must do project #1
 - (ii) If you LAST name starts with **letters K – R**, you must do project #2
 - (iii) If you LAST name starts with **letters S – Z**, you must do project #3
2. The due date for your project is Monday December 9th @ 11:59 p.m. on D2L. You should upload two separate files for your project report:
 - (i) An R file with all your codes.
 - (ii) Please write a final project report and save it as a pdf file. An example of a project will be included in the project folder on D2L. This report includes your
 - (a) Introduction: Where you talk about the dataset and possible question, you wish to answer.
 - (b) Exploratory Data Analysis: Includes scattered diagrams and histogram with some description about the nature of the response and explanatory variables
 - (c) Methods: This includes all the analysis that you do.
 - (i) Frequentist regression model (`lm()` for normal OLS or `glm()` for Binary outcome – logistic regression in R) using all possible explanatory variables and the resulting interpretations.
 - ii. Fit a reduced model with explanatory variable that are significant in i. above
 - iii. Fit a normal Bayesian model or a generalized linear model in the Bayesian framework (in the case of logistic regression model).

Note that discussions about your likelihood function, prior and posterior distributions should be included under this subsection. Use flat or non-informative priors or IID normal prior where applicable.

- (d) Model diagnosis and Variable selection: Evaluate which explanatory variables are best for predicting the response, check residual plot and Q-Q plot.
- (e) Conclusion and discussion: What are your conclusion about the inference? Discuss differences between your best frequentist model and your best Bayesian model. What advantage does one have over the other?

3. The final project report (EXCLUDING THE R file) should be at least 4 pages long and no more than 8 pages.

Disclaimer: The information given here does not; in any way limit your ability to give a detail report of your findings. Please go over your notes and class examples to give you a good idea about what you need to complete this project successfully. There are many materials on the web about Bayesian inference and you are free to consult other notes.