MAT453 - Assignment 6

Spring 2025

Assignment Due (by 11:59 P.M.): Friday, Feb 21

Directions: You may discuss the exercises with other students and with the instructor, but the work you turn in must be your own. You will need the to submit your R code and answers to the questions below in **one** word or pdf file.

Exercises: (10 points total) Again, you will try to perform simulated data analysis in academic research. You have finished the following tasks in Assignment 5:

1. Simulate ten variables (i.e., $X_1, X_2, ..., X_{10}$) from standard normal distributions. Each variables has a sample size of $n = 100$.
2. Calculate the mean parameter μ_i as

$$\mu_i = 1 + 2X_1 + X_2 + 0.5X_5 + 1.5X_{10},$$

   where $i = 1, ..., 100$ is the sample index.
3. Generate the count reponse Y_i from

$$Y_i \sim \text{Poisson}(\mu_i).$$

4. Randomly split your Y and X into training dataset with 80 samples and test dataset with 20 samples.
5. Fit a Poisson regression using your training dataset. Check if the estimated coefficients match the actual one. Make a prediction on your test dataset.

In this assignment:

1. Fit a negative binomial regression using your training dataset. You can Google the R commands and still use the *glm()* function. Make a prediction on your test dataset.
2. Fit a random forest model using your training dataset. You can Google the R commands. Make a prediction on your test dataset.
3. Fit a Extreme Gradient Boosting (XGBoost) model using your training dataset. You can Google the R commands. Make a prediction on your test dataset.