# MAT453 - Assignment 7

Spring 2025

Assignment Due (by 11:59 P.M.): Sunday, March 23

**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) Similar to Assignment 5, you will first simulate the dataset:

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

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

**In this assignment:**

1. Fit a random forest model using entire dataset. Get the variable importance measures of the 100 variables.
2. Fit a Extreme Gradient Boosting (XGBoost) model using entire dataset. Get the variable importance measures of the 100 variables.
3. Fit a LASSO regression using the *glmnet* package. Use 10-fold cross-validation to tune the parameters. See *?cv.glmnet()*. Which predictors have non-zero coefficients?
4. Fit a Elastic Net regression using the *glmnet* package. Use 10-fold cross-validation to tune the parameters. Compare the coefficients with those of LASSO.