MTH 325, Exam #2, At Home, Spring 2023 Name _____

Instructions: This exam is in two parts: Part I is to be completed partly at home using the materials posted in the course for the at-home portion and you will answer questions about that work during the in-class portion of the exam; Part II is to be completed entirely in class. You may not use cell phones, and you may only access internet resources you are specifically directed to use.

At home, prepare for questions in Part I using R. Open the data file entitled **325exam2data.xlsx** posted in Blackboard. There are multiple sheets in this file. Save them to separate dataframes. Complete the calculations noted below. You will be asked for additional analysis and interpretation of this data in the in-class portion of the test. Print out the results of your analysis and code, and bring the pages with you to the exam. You will submit all this work along with the in-class exam.

1. On Sheet 1 is data on boilers including boiler type, drum type, design pressure, capacity and worker hours. Use this data to predict worker hours using boiler type and drum type using a generalized linear model (ANOVA). Identify main effects and if any interaction term is significant. Be prepared to write the equation of the model and discuss diagnostics such as residual plots.

2. Using the same data on Sheet 1, (after eliminating the Boiler number column) create a logistic regression model that predicts Boiler Type from Worker Hours. Plot the graph. Create appropriate exploratory graphs. Create appropriate diagnostic plots, and a confusion matrix.

3. Create a graph of the data on Sheet 2 with Average Monthly Temperature on the horizontal axis, and Average Monthly Bill on the vertical axis. Create a nonlinear model for the data by transforming variables. Plot the resulting model. Create appropriate diagnostic plots. Bonus points for comparing your model to a LOESS model.

4. On Sheet 3 is employee data. Eliminate the Employee column. Gender is already encoded as a binary dummy variable. You'll need to encode the Department variable as separate dummy variables. The rest of the variables are numerical. Use LASSO regression on data to find a model of best fit. Compare the resulting model to a model using a linear model with the same variables. Prepare appropriate diagnostics and diagnostic graphs.