MTH 325, Final Exam, 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 **325final_data.xlsx** posted in Blackboard. 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.

Use the data on motels to complete the following tasks. Sheet 1 has data on existing motels and information about their location, relationship to competitors, specific amenities and operating margin. Sheet 2 has data on possible locations for building a new motel.

1. Import the data in the file into R and remove the Motel and Site columns (they are not a variable). Separate your data into two dataframes. One for the existing locations and one for the possible building sites (to be used later).

2. Convert the Indoor Pool Variable to a binary dummy variable.

3. Create a correlation table of the variables. Make a correlation plot (type is of your choice), or a pairplot.

4. Which variable has the strongest relationship with whether or not the motel has an indoor pool? Create a logistic regression model to predict the indoor pool variable. Create a graph of the model in ggplot. Create appropriate graphs for diagnostic testing of assumptions, and identify potential outliers. Create a confusion matrix for your model.

5. Create a multiple variable model of operating margins using all remaining available variables. Use appropriate automated selection techniques. Compare the result to manual backward selection. In your backward selection, stop only when all the coefficients are significant at the 0.05 level.

6. Construct diagnostic plots for your machine selected model and your manually selected model (these may be the same). Identify any potential problems with model assumptions, outliers and influential points.

7. Using the data on Sheet 2, predict the operating margins for all locations if a) the motel does not have an indoor pool, b) if it does.

To complete the calculations below, use the time series Seatbelts.

8. Create a new column in the dataset (or a separate vector) that represents the ratio of drivers killed to total drivers. Construct appropriate one-variable numerical plots to describe the overall data set.

9. Create a plot of the new time series. Perform seasonal decomposition and plot the resulting graph.

10. Create an ACF graph for the time series.

11. Construct an ARIMA model. Plot the model against the original time series.