MTH 325, Final Exam, At Home, Spring 2024 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 990 purchases from a store. Sheet 2 has 10 additional purchases from the same store.

1. Import the data in the file into R and remove the Person column (it is not a variable). Separate your data into two dataframes. One for the Sheet 1 (training) and one for Sheet 2 (test data).

2. Conduct a two-way ANOVA test (with interaction) for the variables Age (category) and Salary (category) to predict Amount Spent. Create appropriate diagnostic plots for your model. Be prepared to describe your hypothesis tests and their outcomes.

3. Recreate the same model with the glm() function. Be prepared to discuss how the general linear model differs from the ANOVA model.

4. Convert all your categorical variables to dummy variables. Let your defaults be Young (Age), Female (Gender), Rent (Home), No (Married), and Low Salary (Salary Category).

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

6. Create a multiple variable model of Amount Spent 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.

7. 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.

8. Using the data on Sheet 2, predict the Amount Spent for the remaining customers. Compare the results to the provided Amount Spent values. Calculate your error.

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

9. Create a new column in the dataset (or a separate vector) that represents the ratio of front-end crashes to rear-end crashes. Construct appropriate one-variable numerical plots to describe the overall data set.

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

11. Create an ACF and PACF graph for the time series.

12. Construct an ARIMA model. Plot the model against the original time series. You may need to experiment with settings to select the best combination of p, d, and q.