You can submit the assignment in a pdf file, Jupyter notebook, or you can write down your discussion and explanation in Excel or R Markdown, etc. You are welcome to use any software. Please make sure to submit the file that contains your analyses, too (Excel file, R codes, etc.).

1   Comparison of weekdays

For each day of the week, plot the histogram of the actual number of surgeries. Are the average and standard deviation for the number of surgeries on each day of the week (Monday to Friday) the same? Perform appropriate hypothesis tests and discuss the results. Is there a specific day of the week with a relatively higher average than the others? What could be the reason for the higher average? (*25%*)

2   Longer prediction time and precision trade-off

Ajay Bose would like to predict the final number of surgeries on a specific day, using the data of scheduled surgeries. However, he doesn't know how many days before the desired date, he could predict the demand. Note that the sooner he predicts, the more error in prediction he will probably observe. Divide the data into 80 percent training and 20 percent testing. Consider the first model discussed in the class and calculate MSE (mean square error) on the test set for $T-5, T-6, T-7, T-8, T-9$ as predictors. Visualize the data and discuss the trade-off between sooner prediction and an increase in error. Do the same steps for R^2 values. Which day do you suggest as the predictor? (*35%*)

3   Time-Series vs. Regression

The provided data includes the number of surgeries scheduled to be performed on a specific date prior to the surgery (actual) date. As discussed in the lecture, there is a strong correlation between the predictor variables (columns in the data).

1. To reduce the correlation, consider add-on surgeries (the difference between two columns) as new predictors and develop a new regression model. Implement the following models and compare them with the models discussed in the lecture. (*15%*)

   - Model 1: Does not stratify by the day of the week.
   - Model 2: Stratified by the day of the week.

2. Consider the surgery (actual) date as a time series with September 4th to September 14th as the testing set and the rest as the training set. Fit a Moving Average (MA) model to the time series and visualize it. (*15%*)

3. Compare the result of the regression model with the MA model visually and based on MSE. Which model provides a better prediction? What could be the potential reason? (*10%*)