Data Brief

Purpose:

The Vanderbilt University Medical Center surgical operations leadership team has asked for a Data Brief to help them better understand and use predictive modeling for surgery planning. The goal of this brief is to provide an overview of the benefits of predictive modeling in surgery planning, as well as a short analysis of the accompanying data.

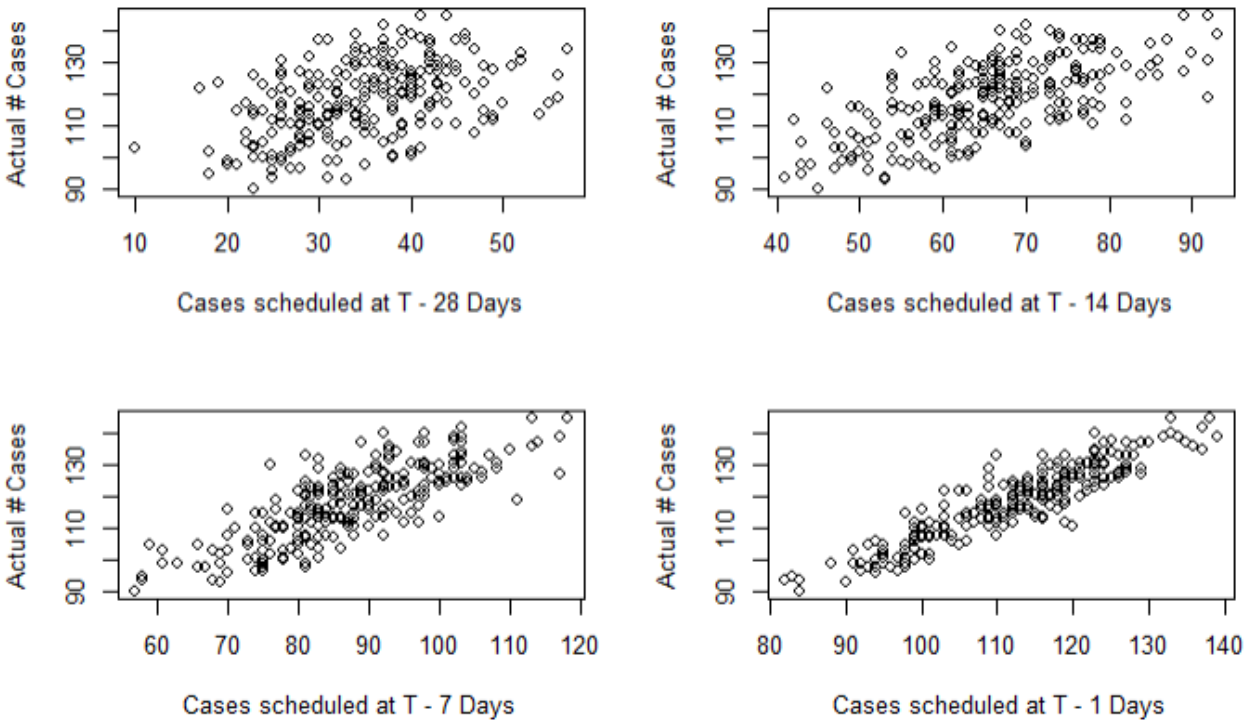
Background:

Vanderbilt University Medical Center (VUMC) in Nashville Tennessee faces a significant challenge in managing daily surgical case volume. Although the surgical staff schedules are made weeks in advance, the final number of surgeries to be performed is only know with precision the day of. The inability to accurately predict the number of elective surgeries causes over and understaffing of the surgical teams and inefficiencies in the use of operating room (OR) resources. To address this, VUMC is looking for ways to better predict the number of elective surgeries. This would allow OR managers adjust staffing levels to better match demand on any given day, saving VUMC time and capital.

Looking at the Data:

After analyzing the provided data, there are a few key takeaways.

A) Number of cases is predictive of actual final case volume. This is evident when looking at the plot below:



Each scatterplot shows a positive relationship between the two variables. In each of these scatterplots the cases scheduled are positively related to the number of cases on the actual day of surgery. This results in the variables holding some predictive power of the actual number of cases.

B) We can also see from the scatterplot above that predictive power (linearity) also changes as the number of days until the actual date decreases. As we get closer and closer to the specified date, the relationship's variability decreases significantly around the line Actual = Scheduled.

C) After investigating the influence that the day of the week has on the actual number of surgeries performed, it is clear that there is a discernible and statistically significant impact. The relationships are as follows:

- Monday: 5.48 more surgeries than Friday on average
- Tuesday: 7.24 more surgeries than Friday on average
- Wednesday: 5.46 more surgeries than Friday on average
- Thursday: 11.19 more surgeries than Friday on average
- Friday: Relatively low amount of surgeries on average

Predictive Model:

I believe the linear regression model below will be most helpful:

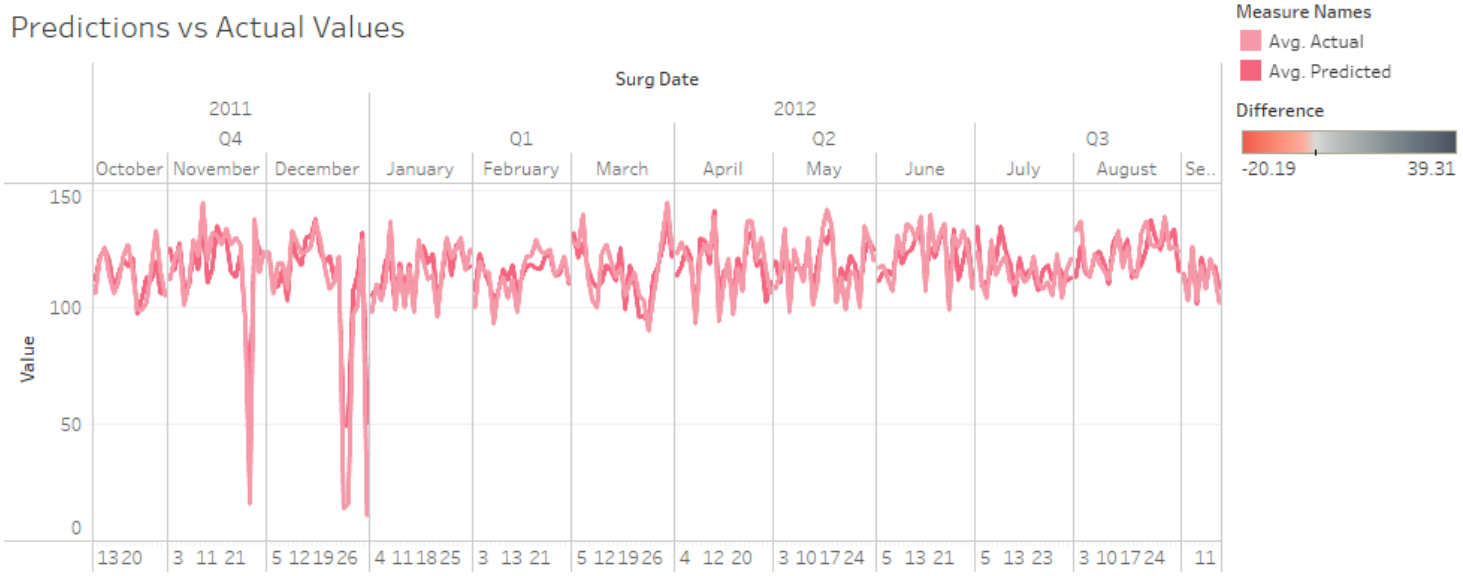
```
Call:
lm(formula = Actual ~ T28 + T21 + T14 + T13 + T12 + T11 + T10 +
    T9 + T8 + T7, data = train)

Residuals:
     Min       1Q   Median       3Q      Max
-15.7231  -5.1683  -0.5313   4.3629  20.1902

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 49.52737    3.96013  12.506  < 2e-16 ***
T28         -0.04285    0.12370  -0.346 0.729458
T21          0.13126    0.12934   1.015 0.311583
T14         -0.08323    0.19433  -0.428 0.668969
T13         -0.51773    0.30296  -1.709 0.089210 .
T12          0.68439    0.34249   1.998 0.047208 *
T11         -0.51174    0.31612  -1.619 0.107256
T10          0.32394    0.24389   1.328 0.185813
T9           0.17745    0.22835   0.777 0.438136
T8          -0.13504    0.25229  -0.535 0.593132
T7           0.75269    0.20131   3.739 0.000249 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.058 on 178 degrees of freedom
Multiple R-squared:  0.6435,    Adjusted R-squared:  0.6235
F-statistic: 32.13 on 10 and 178 DF,  p-value: < 2.2e-16
```

As mentioned above, I suggest using a linear regression model to predict the number of actual cases taking place on a given day. As input the model takes the number of cases scheduled from T – 28 up to T – 7. I chose not to include from T – 6 to T – 1 because I want to give VUMC a one week notice regarding actual number of surgeries. This will give the managers time to breathe and adjust surgical staff schedules as needed. Do not trust this model blindly, it will not (along with almost any machine learning algorithm) be able to predict perfectly. It will give management a good idea of the number of actual cases on a given day.

Dashboard:

I have created a dashboard that helps VUMC monitor the performance of the model by displaying the actual and predicted daily cases on the top sheet. The top sheet displays the actual daily cases on top of the predicted daily cases, which allows the hospital to easily compare the two and identify any discrepancies. The bottom sheet is displaying the difference between the two lines on the top chart, which gives the hospital an at-a-glance view of how well the model is performing. The hospital will be able to keep an eye on this dashboard to see how well the model is performing and make adjustments as needed to improve its accuracy.

Predictions vs Actual Values



Difference in Predictions and Actual