Write-up

**Summary**

The current status of the project is as follows. The rest of tomorrow (Thursday) I plan to try simple forward and backwards regression and report my findings. I think the most interesting thing to note so far aside from some PELD differences is that the platelet count seems to be significant, contrary to previous findings. I plan to reconfigure my fisher and chi-square tests tomorrow as well, and report back on those.

**Page 1**

The first chunk of R code is the function I wrote to manually calculate the Growth Factor failure. I am pretty sure it’s works accurately, and it’s actually a tool that can probably be distributed to anyone publicly who runs across the problems we encountered early on.

I can explain how it works on our call.

**Page 2**

I define all the datasets that are to be used in the analyses. I merge the biliary and ekg datasets, and I make sure all the values for PELD indicators are rounded up to their minimum values.

**Page 3**

Most of page 3 is old code that I have commented out, but at the end is where I use my newly written function to calculate the GF factor for every patient.

**Page 4 & 5**

I use the GF factor to calculate the PELD score for each patient. Using the logistic regression model, I identify influential points in the data and remove them and store the remaining data points as a new dataset. The influential points are based on Cooks Distances

**Page 6 – 10**

I ran all the t-tests again using the method we discussed. I created a new dataset for each factor with only the observations that had values, instead of removing all observations with any missing values. The tests that indicated significant p-values were INR, bilirubin, listing age, and **platelets.**

**Page 11**

Commented t-tests to be updated.

**Page 12 & 13**

I produced four classification trees. In order: 1) all factors, 2) the factors that showed to be significant In the t-tests, 3) all non-PELD factors, 4) only PELD factors.