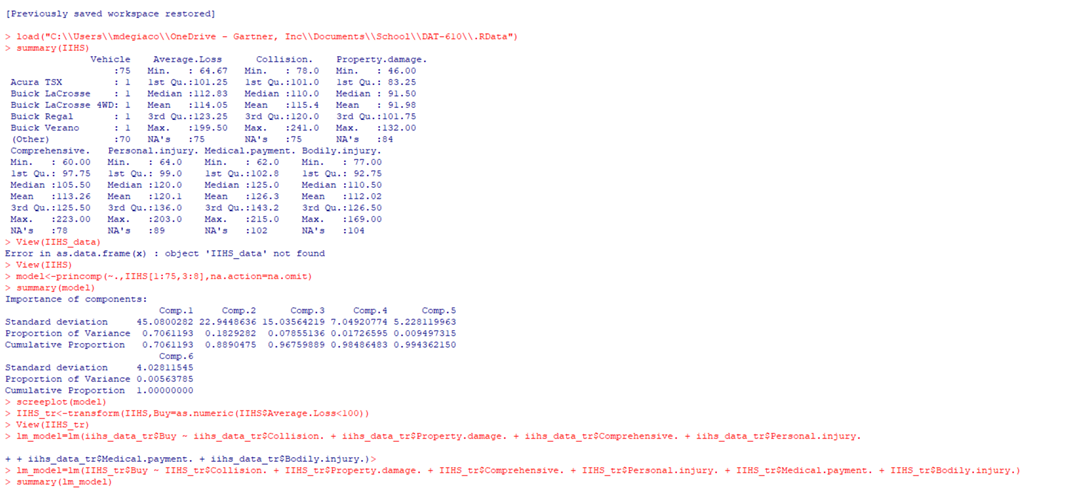
**Module 6 R Activity**

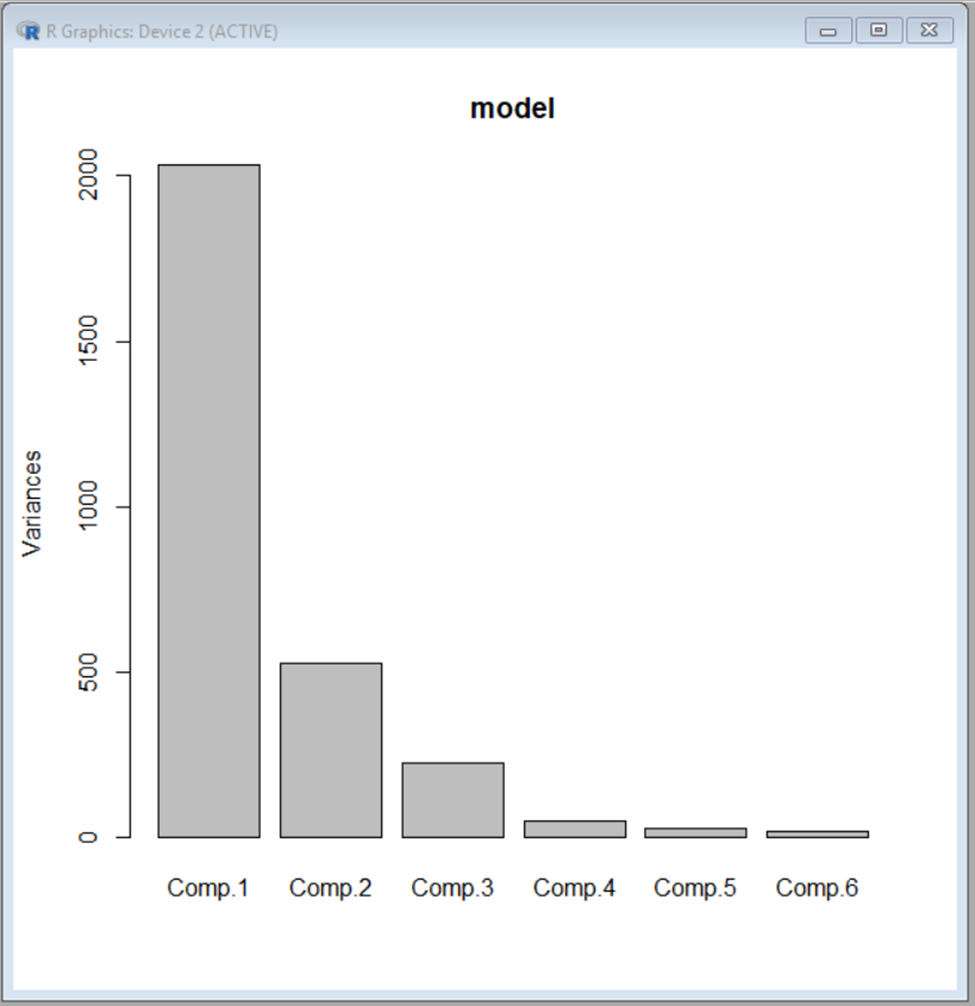
**Determining Key Risk Indicators**

**Southern New Hampshire University**

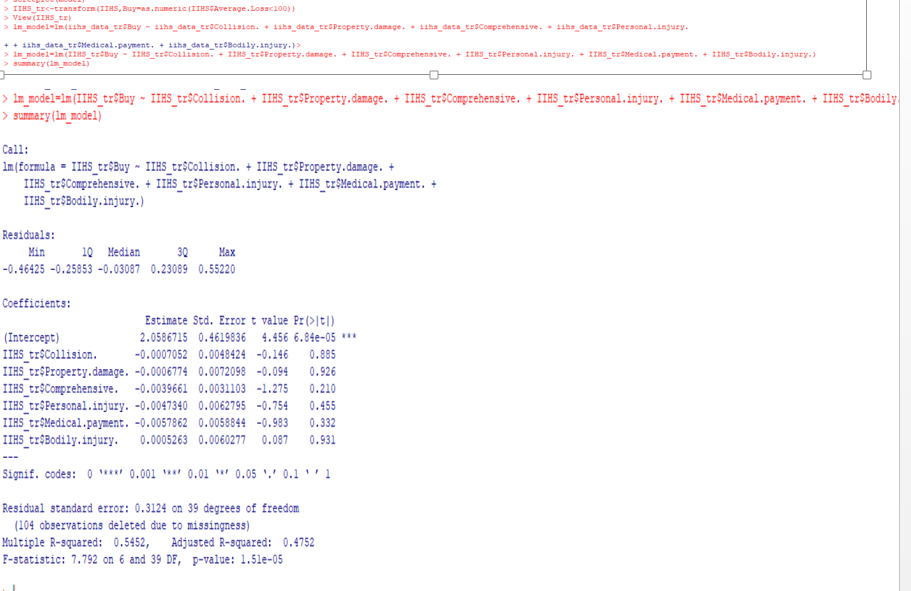
**4/21/2019**

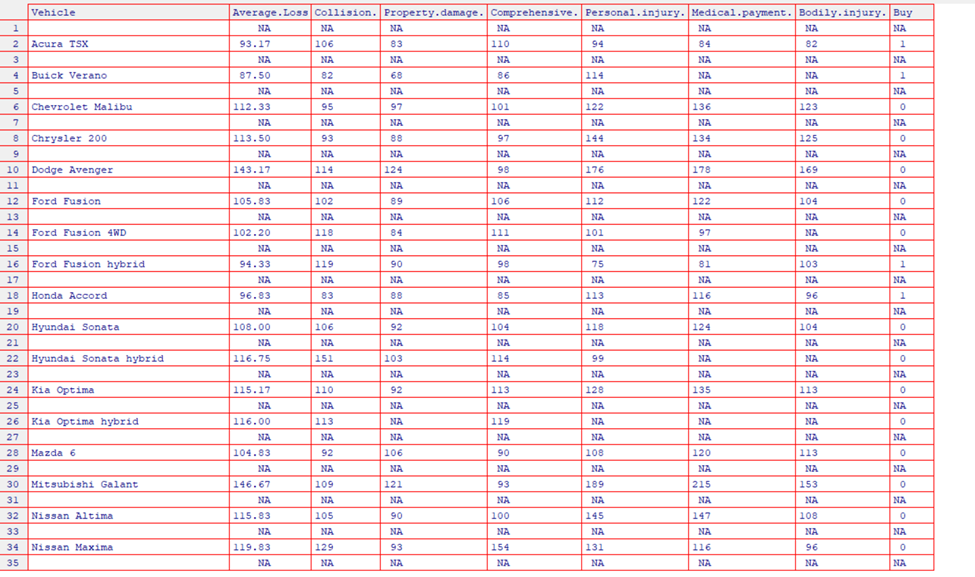
This week we used R to help us determine what the Key Risk Indicators of our data set that had the strongest correlation to the end result. A KRI (key risk indicator) is a set of factors used to monitor an organization’s operational risk performance. There are three types of KPI’s that include: exception indicators which provide immediate notice of risk, lagging indicators which provide a reflective view of the risk related trends, and lastly, leading indicators which allow an operational risk management program to anticipate upcoming risk events. We are using R this week to determine this data set’s KPI’s using the PCA analysis, linear regressions, and logistic regression techniques. First I read the new data set into R, the next string got us to create a PCA model, then we did a summary of that model with a screeplot to visually display the results. 

This produced 6 components in which component 1 has the highest standard deviation by far and drops by half from comp 1 to comp 2 and then keeps taking a major drop off as we from 1-6. The proportion of variance does the same thing, as we go from component to component to proportion of variance dips with each one as well. Component 6 has the only cumulative proportion of 1.0 with the rest of them all building up to the 1.0 score in component 6. This screeplot (below) visually shows the difference that these components aren’t as close as we think with component 1 being by far the most useful to help summarize the data. We can use this to determine KRI’s by using these components that are high on the plot like component one to give insight into variables that have a high correlation to hypothesis of the data.

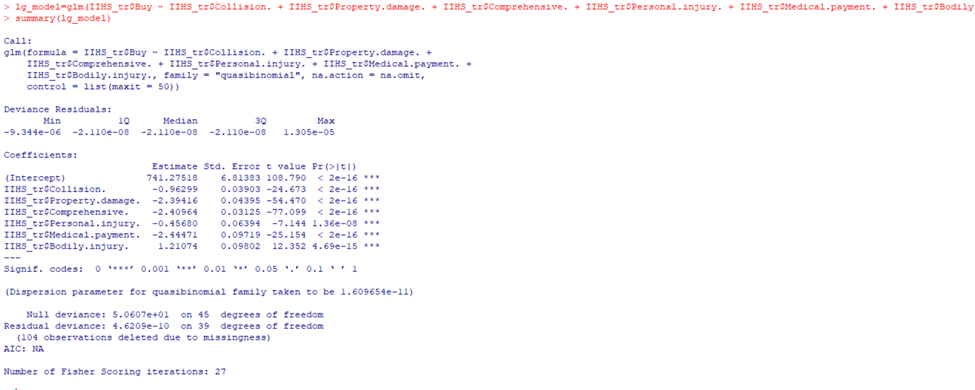


To start off the linear regression we added the buy column to the data and then we viewed our new modified set to do the linear regression, then used to lm function string of code to perform it.





Linear regression can help us derive KPI’s by helping us show which variables are significant or not, meaning that if its significant there a higher likelihood that the null hypothesis is false and no significance means we accept the null. Knowing this helps us show the relationship between that variable and the question we are trying to answer. It also shows in the estimate column which shows the variable effects the main factor, in our example we see the variable collision as the top variable meaning that for every collision claim that happens, the likelihood of recommending that car decreases -0.0007052.

Lastly is logistic regression, we use this unlike the linear regression, it’s used to produce a dual outcome. We used the lg\_model string of code in R to produce an output that looks similar to the linear regression but has different outputs for those same variables. 

Now we can see, that each variable in the estimate column looks different from the linear regression one and we have some new insight into new possible KRI’s. I think something may have gone wrong because bodily injury is positive but also has a high error rate so we know that recommendation isn’t accurate all the time, so when a bodily injury claim comes in the likelihood of recommending that increases but disregarding that one you can see how using the logistic gives us some different looks at the variables.

Overall after looking at the PCA, linear regression, and logisitic regression we have come up with a subset of KRI’s that have a strong correlation to the “buy” factor. I would start with the variable collision. It scored as the highest component of our PCA in terms of Standard deviation, in our linear regression it had the second lowest error rate, and the same in the logistic regression. The next variable I would recommend would be comprehensible, that scores well in both regressions analysis’s by having the lowest error rate and having a high impact of effecting a buy recommendation or not. After this preliminary look the last variable I would consider as a KRI is personal injury, the more times a medical payment happens when driving that kind of car the less likely someone would be to recommend to buy it. This has one of the highest estimate numbers in our logistic regression, with a score over -2 and one of the highest scores in the linear regression as well.

REFERENCES:

* Girling, P. (2013). *Operational risk management: A complete guide to a successful operational risk framework*. Hoboken: Wiley
* DAT 610. (2016a). *DAT 610: Final Project Guidelines and Rubric Module 6*. Manchester, NH: Southern New Hampshire University.
* DAT 610. (2016a). *DAT 610: Module 6 Guidelines and Rubric*. Manchester, NH: Southern New Hampshire University.
* DAT 610. (2016a). *DAT 610: Module 6 additional materials*. Manchester, NH: Southern New Hampshire University.