**Statistical/Hypothetical Question**

**DSC 530**

**Abed Tabbalat**

After attempting an exploratory data analysis exercise on insurance data, the exercise is whether to determine there are relationships between variables such as premiums, ceded premiums, and incurred losses. The initial expectation was that all components have relationships, and the outcome was not as expected. It turned out that losses have no relationships to the premiums, but ceded premiums do.

I believe if more variables were factored in between losses and premiums such as credit score history, age of home, etc. It would give a better conclusion to the relationship between policy holders and their losses. However, with the time constraints, and the work needed to clean all 110 columns to figure out the definitions and what is needed to get the data clean, a decision was made to not compute this exercise into the project.

I think adding net margin and using it as its own variable can be a mistake in this analysis, since the variable is a computation of all the numbered variables that exist within the analysis. Hence, Net Margin will have relationships between all the variables that were used.

My biggest challenge in this project was running PDF scenarios. I struggled mostly with the Thinkplot package and was getting errors most of the time. In addition, PDFs and the meaning behind them has been the most challenging part in this course and till now I am not in full shape of how and when to use them and what the output would mean in an analysis. I will be working on understanding them more as I work closely with the Actuary department in our company.

All in all, this course has opened my eyes on various functions that I never knew about and how I can utilize these functions in analysis exercises that can give the decision makers better understanding on what is going on.