**eHealth 705 Term Projects, Winter 2021 term**

These projects have been provided since they include good datasets that have been obtained from important medical research. Some of these datasets are accompanied by journal articles. You do not need to replicate the analysis contained in the journal articles; that would be at a technical level much higher than covered in eH705 in most cases. You are expected to only use what you have learned in this course.

I recommend that you use the journal articles as background and for comparison of your data to the findings in the article if they are at an understandable level.

Your analysis should follow the steps taken in this course:

1. Read the material provided. Sometimes very little is mentioned about the data. In all cases, I’ve provided everything I have about the dataset that you could use. You may find more information through online searches. (Good datasets are very difficult to obtain. )
2. Set up your R Markdown file and read the dataset. This may be done in RStudio or in R Commander. Feel free to use either R-code directly in RStudio (R Markdown) or in R Commander (producing an R Markdown document).
3. Exploratory Data Analysis. Produce basic tables and graphs so that you understand the data and can explain the basics in your report. You may have to google the literature to get background on some data and medical terms.
4. You may have to convert some variables to factors and will likely need to recode to provide value labels for factor variables.
5. Check on missing values and determine the best way of handling missing values.
6. There should be one variable that qualifies as a response or dependent variable. Your focus for much of your work will be to investigate how to explain and/or predict values of the response variable by using predictor variables.
7. Your work in point 6 will likely include some form of regression and, perhaps ANOVA and correlation. The regression may be linear regression or logistic regression. That will depend on the nature of the response variable. Sometimes the analysis in the journal article was conducted using survival analysis. While that topic is covered in this course, you may choose to use survival analysis or logistic regression.
8. Explaining your work and the findings is extremely important. Naturally, this is a statistics course and you may use statistical terminology and notation in your reports. However, you must show that you understand the material and can explain it in plain language to others who have limited statistical backgrounds. Simply producing a mass of statistical output with no or very little explanation will not attract a very high mark.
9. Outstanding term projects may receive recognition through bonus marks.
10. Ideally you should start now. We are just beginning regression and you can use your project data to help learn regression and move ahead with your project.
11. Good luck!