STAT 6012: Linear Models for Data Science
Class Activity 5

Due date: Tuesday, July 23 by 10:50 am Via Canvas.

Complete the following questions an R Markdown file and submit your compiled HTML file. If you are working in a group, list the names (last, first) of the group members in alphabetical order of last names.

The attached dataset contains information compiled by the World Health Organization and the United Nations to track factors that affect life expectancy in 2015.

Our goal is to build a multiple linear regression model to predict `life.Expectancy` using:

`Adult.Mortality,infant.deaths,HIV.AIDS,BMI, GDP,Schooling`.

To make calculations easier, the following sample R code subsets the data on the two variables, as well as removes all missing values.

```
life_data3<-select(life_data,Life.expectancy, Adult.Mortality,
                infant.deaths,HIV.AIDS,BMI, GDP,Schooling)%>%
            na.omit()
```

We will estimate the parameters of the multiple regression model using the least square method.

1. [5] We also deduced in class that the general least square estimates for y-intercept and slope(s) is given by:
$$(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$$
where $\mathbf{X}$ is the design matrix, and $\mathbf{Y}$ is vector of values of the response variable.
   (a) [2] Use R to generate the design matrix, $\mathbf{X}$ for this simple linear regression mondel.
   (b) [3] Use R to find the model coefficients using $(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$.
2. [2] Find the estimates of the multiple linear regression using the `lm()` function in R.
3. [3] Use the Bootstrap method with 10000 replicates to find the bootstrap estimates for the multiple regression model.