***BUS 347.01 Introduction to Business Analytics***

***Homework 4 (60 Points)***

**Homework Description:**

The attached file “HW4 Data.csv” contains 3000 observations from a simulated data table. In this simulated dataset, Y is the target variable that has been generated by X1-X5.

You need to perform a stepwise selection method to identify the best model structure and perform a model diagnosis on the optimal model structure. Your submission should be an analytics report in the word format knitted from RMarkdown code.

In your analytics report, you need clearly label the following items:

* Question Number
* R Code
* R Output
* Conclusion, if applicable

You need to submit the knitted analytics report on Blackboard. Please carefully check your work before the submission, as you can only submit your work once. Late submissions will not be accepted.

**Homework Questions (Each Question has 10 point)**

**Q1.** Treat the whole sample as training dataset. Consider only the main effects of variables in your model specification, use the step() function in R to find the “optimal” model (name it as Optimal\_1). Report this optimal model fit and the corresponding adjusted r-squared.

**Q2.** Treat the whole sample as training dataset. Consider both the main effects and the interactions of variables in your model specification, use the step() function in R to find the “optimal” model (name it as Optimal\_2). Report this optimal model fit and the corresponding adjusted r-squared

**Q3.** Use the first 2000 observations as training dataset and the remaining 1000 as validation dataset. Repeat Q1 to find the “optimal” model with only main effects (name it as Optimal\_3).

**Q4.** Calculate the RMSE for Optimal\_3 for both training and validation datasets. Discuss whether “Optimal\_3” model structure has the over-fitting issue.

**Q5**. Examine whether the residuals in “Optimal\_3” follow a normal distribution. Explicitly draw your conclusion.

**Q6.** Examine whether the residuals in “Optimal\_3” have constant variance. Explicitly draw your conclusion.