BUAN 6356 (Johnston) 20220830
Homework 2A
Due: 10 September 2022 (6PM)
Part B Due: 11 September 2022 (11:59PM)

Points available: 80

This assignment is about running linear models on data. HW 2B (the interpretation and analysis of the HW 2A results) will become available after your correctly completion of this assignment or the submit window closes.

The first commands of your code MUST include:

 setwd("c:/data/BUAN6356/HW_2"); source("prep.txt", echo=T)

and the last command of your code MUST include:

source("validate.txt", echo=T)

The required code CAN be set up for conditional execution. (E.g.: set a Boolean variable and then use it in an if() to execute these statements.)

Be careful with the quote characters as they must ALL be the same at the beginning and end of a string. (Use the single or double quote character from the key next to "Enter".) Inclusion of these lines is required BEFORE your code will be tested.

The data for this assignment is Boston from the R package MASS. You will not need to install or use any other R packages for this assignment.

Use a 10% testing (validation) sample with **722538843** as a random number seed.

You objective is to load and analyze the Boston Housing data. Your outcome variable (dependent) is "medv" (median value). Use all remaining variables as covariates (independent) initially and then refine the model via step() .

Submit the code to eLearning as an ASCII file which can be copied directly into R.

You may submit this assignment as many times as needed before the Part A Due Date to get full credit.

Deliverables (all names as presented):

1. Boston   Boston (dataframe from MASS)
2. wk   Boston dataframe with no missing data
3. seed   (vector) RNG seed value
4. tNum   (vector) number of index values in testing sample
5. tst   (vector) index values of 10% testing sample
6. predBase   (vector) predicted values from baseline model (training set)
7. predStep   (vector) predicted values from stepwise model (training set)
8. tstPredS   (vector) predicted values for stepwise model (testing sample)