# BUAN 6357 (Johnston)
# Homework 5A (20230407)
# Code Due: 15 April 2023 (6PM)

Points available: 60.

This assignment is bootstrapping with Ordinary-Least-Squares. Use "data(airquality)" for this analysis. The outcome variable (dependent) is Ozone. Seed the RNG with 337337168 for each analysis section which requires the use of random numbers. Your deliverables will require performing both "Classic" and "Balanced" bootstraps, each with 500 bootstrap samples. You should perform both "Classic" and "Balanced" via data.table() 3-parameter, "by group" style code (see the data.table() example at the end of the how2 file). Both method/style combinations will require use of the tidy() function from the "broom" package. Reading the "broom article" from UTDbox>supplemental_materials is strongly recommended although the article code uses only the "tidyverse" programming approach. See the "Notes" and "Hint/Comments" sections in this document.

For this assignment you will need the package "tidyverse", "broom", and "data.table". You should not use any additional packages. You should use only the "require()" or "library()" statement in your code. Any use of the install.packages() function in submitted code will result in a score of 0 for that submission. Multiple instances of concurrently submitted code will result in only the most recent code being evaluated and scored on the assumption that you realized the earlier submitted code contained a problem which was fixed and then re-submitted.

The first commands of your code submitted for grading to eLearning MUST be:

**setwd("c:/data/BUAN6357/HW_5"); source("prep.txt", echo=T)**

and the last command of your code (which actually provides the scores) MUST be:

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

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 close to "Enter".) Inclusion of these lines is required BEFORE your code will be tested.

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

You may submit HW 5A as many times as needed until you get full credit. (Then you should stop.)

Deliverables:

1. seed (vector) random number generator seed
2. raw (data.table) see Notes
3. n (vector) number of observations in "raw"
4. b (vector) number of bootstrap iterations
5. cl.dt (data.table) DT "classic" bootstrap result; see Notes
6. bal.dt (data.table) DT "balanced" bootstrap result; see Notes

HW 5B will direct you to explore both the intermediate results, the deliverables from Part A and answer questions about each of them. You may submit answers to HW 5B as many times as you wish but only the score for the last submitted code will be retained.

Notes:

- raw: retain only complete case observations; drop variable "Day"; convert variable "Month" to factor
- follow the example codes found in the bootstrap how2 and demo 06_bootstrap
- use all information returned by the tidy() function; see the "broom article" in "recommended_reading" for more information about this function
- Use tidy(<lm>,conf.int=T) to get 95% CI for coefficient estimates on each linear model.
- Re-start the RNG for each Bootstrap strategy