BUAN 6356 (Johnston)
Homework 3A (20220911)
Due: 17 September 2022 (6:00PM)

Points available: 20

This assignment is about running linear models on data. HW 3A checks data import and a basic linear model (Ordinary Least Squares). The models required to answer the part 3B online questions are posted in a separate document. Code for part 3B does NOT need to be submitted and all 3B questions are graded automatically by eLearning.

The first commands of your code MUST include:

setwd("c:/data/BUAN6356/HW_3"); 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.

Also, if you submit more than 1 copy of your code before it is evaluated and scored there will be the assumption that the last copy is corrected and updated from the earlier copies and thus only the last copy will be run. The comment associated with each of the earlier copies will be "Multiple entries in queue".

The data for this assignment is "dobson_covariance_6_12_mod.txt" and is available through UTDbox>data. You will need to load this data yourself. This data will be available for your submitted code in the target directory specified by setwd().

No testing (validation) sample will be used with this assignment.

You can assume the "data.table" package is installed on my machine. Use either library() or require() to make it available. If you need to install it on your machine use the command 'install.packages("data.table")' but DO NOT include this statement in the code you submit.

You objective is to load and analyze the data. Your outcome variable (dependent) is "achievement" (a proficiency score). The independent variables (covariates) are "aptitude" (a skills score) and "method" (a nominal variable for different preparation/teaching methods). You will want to use as many models as are applicable from the Stat Notes handout 11 and scripts files in the pre-recorded materials section.

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 until you get full credit.

Deliverables (all names as listed):

1. raw the original data in data.table form (no factor variables)
2. m0Pred predicted values for the NULL model associated with this data

Notes/Hints:

- You CAN work with factor or indicator variables which you create.
- Refer to the "formulas" handout and stat_notes_11d_20220606.txt for additional syntax which may be useful
- An lm() model formula with only an intercept term would look like "Y~1".