BUAN 6357 (Johnston)
Homework 5b
Due: 6 April 2019 (11:59PM)

The number of points available for this homework assignment is large.

This assignment builds on the results from HW 5A by analyzing the various residual values. If you did not get full credit for HW5A you may use the posted R code found in eLearning. I suggest you generate the results from HW 5A and use save.image() to store them all for use in answering HW 5B. You will not need to turn in code for this assignment but will need to do some programming.

Basically, you generated these results – now what can we do with them?

For this assignment you will need the package "data.table". You will not need any additional packages but may use any package(s) you choose.

You will need to generate several measures: parametric and non-parametric CI (90%, 75%, and 50%), the proportion of designated residuals which fall inside a particular CI, and the RMSE of designated residuals. You may want to write simple functions to implement these procedures (this is not a requirement). You may also wish to explore different ways of displaying designated residuals such as scatter plots with lines delimiting CI intervals, color coding of residual categories (which is the largest CI in which they are to be found), and other graphics including but not limited to boxplots. This is not a requirement for this assignment.

General guideline: treat the residuals from each deliverable as a fundamental group and be prepared to answer questions about each. You should also treat the k-fold groups as an additional collection of groups and be prepared to answer questions about those groups.

Deliverables (data tables) from HW 5A:

1. raw original data, as read, with addition of “idx”
2. base_m baseline (non-CV) model residuals: resid_base, idx_base
3. simple_m simple CV (10%) training results: resid_simple, idx_simple
4. simple_cv simple CV (10%) test results: resid_cv, idx_cv
5. jk_m LOOCV training results: resid_jk, idx_jk, iter_jk
6. jk_cv LOOCV test results: resid_jk, idx_jk, iter_jk
7. k_m k=10 training results: resid_k, idx_k, grp_k
8. k_cv k=10 test results: resid_cv, idx_cv, grp_cv