Homework 6 (66 pts)

The learning objectives for this week are as follows: Perform model diagnostics for MLR to identify possible violations of model assumptions and utilize advanced regression models to deal with challenges such as non-constant variance, influential outliers, and multicollinearity; Implement bootstrapping methods to enhance the alignment of results with data that does not meet underlying assumptions.

Problems 1-5 use the life expectancy dataset.

Problem 1 (24 pts)

For the life expectancy data, Consider the MLR model given by $Y \sim X_1 + X_2 + X_3$.

(a) [6 pts] Obtain a partial regression plot for X_1, X_2, and X_3 and discuss whether the regression relationships in the fitted regression function are inappropriate for any of the predictor variables.

(b) [10 pts] Check for influential points by calculating DFFITS, DFBETAS and Cook's distance values.

(c) [8 pts] Determine whether there is multicollinearity present based on VIF.

Problem 2 (3 pts)

Use R to summarize the OLS model. Then compute the confidence interval for the linear impact of X_1, X_2 and X_3.

Problem 3 (9 pts)

Use R to complete the WLS analysis on the model (note: just do one iteration). Discuss the difference (3 pts). Then compute the confidence interval for the linear impact of X_1, X_2 and X_3 (6 pts).

Problem 4 (6 pts)

Use R to complete the Robust analysis on the model. Then compute the confidence interval for the linear impact of X_1, X_2 and X_3.

Problem 5 (15 pts)

Apply the bootstrapping method to compute the confidence interval on the parameters for WLS (5 pts), OLS (5 pts) and Robust (5 pts).

The following problem uses the body fat dataset.

Problem 6 (9 pts)

Note that the lm.ridge() algorithm turns out to be indeterminate when VIF is close to 1, resulting in a flat ridge trace line. On the other hand, the lmridge() function only reports $\lambda = 0$ (i.e., OLS). For practicing the lm.ridge() and lmridge() functions, this question will be to use the body fat example (in Lecture 21) and complete the confidence interval for X_1, X_2 and X_3 in the ridge model. Data is available from the course website.