PART I

Copy the file harvard to your own directory. This is a 14 column dataset consisting of 307 measurements of leaf photosynthesis and a set of environmental variables sampled from July through September 2001 at the Harvard Forest Long Term Ecological Research site.

- 1 Month
- 2 Tower
- 3 Canopy Level
- 4 Hour
- 5 Photosynthetic Rate
- 6 Leaf Conductance
- 7 Photosynthetically Available Radiation
- 8 Carbon Dioxide Concentration
- 9 Leaf Temperature
- 10 Air Temperature
- 11 Relative Humidity
- 12 Saturation Vapor Pressure of Leaf at T
- 13 Vapor Pressure of Air
- 14 Vapor Pressure Deficit

The goal of the assigned exercise is to evaluate the role of Leaf Temperature, PAR, Leaf Conductance, and Vapor Pressure Deficit in controlling Photosynthetic Rate. Add these variables into the model in order of their explanatory power, given the other variables already in the model. Calculate the PRESS statistic for this model and interpret it (see below first).

Qualitatively evaluate the sensitivity of the model coefficients to the month of sampling. Does month make a difference? If so, demonstrate this difference graphically on one plot. Does the month or months considered influence the strength of the model? If so, demonstrate this difference graphically on one plot.

Are there any strong collinearities among the explanatory variables in your model? Do the model coefficients appear sensitive to the presence/absence of these variables in the model?

Examine the relationship between Carbon Dioxide Concentration and Photosynthetic Rate. Does this relationships correspond to your expectations? What explanation can you provide for the nature of this relationship?

PART II

Write a function that produces standardized, studentized, and jacknife residuals and produces a boxplot that shows each of these residual types as well as the normal residuals. Apply the function to your full model from above. Make your function as generic as possible so that it can be used in the future with other datasets.