BTRY 6020 Lab III

February 13, 2017

The goals of the following lab are:

1) To illustrate the source of multicollinearity in multiple regression;

2) To illustrate the effects of multicollinearity in multiple regression;

3) To illustrate the usefulness of polynomial regression in approximating curves.

##Question 1: Introductory Multiple Linear Regression

With health care costs spiraling out of control, administrators look for inefficiencies everywhere. One particular study examined the number of nurses used to staff a hospital (NumNurses) based on the number of beds (NumBeds) and the average number of patients in the hospital (NumPats).

A) Regress the number of nurses (NumNurses) on the number of patients (NumPats). What is the coefficient of NumPats? Its estimated standard error?

B) Regress the number of nurses NumNurses on both the number of patients (NumPats) and the number of beds (NumBeds). What is the estimated regression equation? Give a point estimate for the number of nurses that would staff a 280 bed hospital that averaged 228 patients per day.

C) Now with Numbeds in the regression, what is the coefficient of NumPats? Its standard error? How has this changed from the simple linear regression in part A above?

D) Plot the Numbeds vs NumPats; does your answer in part C appear to be correct?

E) Get the correlation coefficient between NumBeds and NumPats; does this confirm what you answered for part C.

F) From part F above, compute the VIF between the two predictors. Confirm this by getting the VIFs from R.

G) Interpret the coefficient of NumPats you got in the simple linear regression in Part A; then interpret the coefficient of NumPats you got in Part B. Is your interpretation of the coefficient of NumPats in the multiple regression setting valid? Why?

H) What is the inferential model for this multiple linear regression? What assumptions does this include?

I) Get the standardized (or studentized) residual plot and qqPlot for the residuals in this multiple regression. Comment on the usefulness of this regression for inference.

J) Plot NumNurses against Numbeds and NumNurses against NumPats. Based on these plots and the plots obtained in part J above, how would you suggest proceeding before doing inference?

##Question 2: Polynomial Regression

The effectiveness of a new experimental overdrive gear in reducing gasoline consumption was evaluated in 12 trials with a light truck equipped with this gear. A study was undertaken to determine the effectiveness of this in increasing gas mileage at different speeds.

A) Plot the data (of course!).

B) Based on what you observe in the plot, what order polynomial appears appropriate here?

C) Run a polynomial regression with the order (highest power of a polynomial) one greater than what you answered in part b above. Is the highest power statistically significant?

D) Based on your answer above, make any changes necessary to your regression and rerun this. Remember, the objective of polynomial regression is to run an order one higher than necessary (that is non-significant), then drop back to an appropriate level (so the highest power is statistically significant).

E) Without giving thought to being correct, use the definition of slope in multiple regression to interpret the coefficient of speed 1 . Can you see why this makes no sense whatsoever, and so interpreting these coefficients is incorrect?

F) Currently used transmissions get 31 mpg at 52 miles per hour. Can we get a greater average than this with the new overdrive gear? State hypotheses, test statistic, p-value, and conclusions, assuming all assumptions for statistical inference have been met.