APM 630 Regression Analysis
Project #8 – Multicollinearity

Data: PMSA.xls

The data on the following variables were obtained from the US Bureau of the Census State Metropolitan Area Data Book 1986. A set of 53 primary metropolitan area (PMSA) were selected. The description of the variables is:

(1) Dependent variable:

- CRIME – total number of serious crimes in 1980.

(2) Predictor variables:

- POP – total population in 1980, in thousands.
- AREA – land area in square miles.
- YOUNG – 1980 population ages 18-24 in thousands.
- DIV – total of divorces in 1980 in thousands.
- OLD – 1980 total number of SSN benefit recipients in thousands.
- EDUC – number of adults (>25 years old) having completed 12 or more years of school.
- POV – total number of persons below poverty level in 1980.
- UNEMP – total number of unemployed in 1980.

Assignment:

1. Compute descriptive statistics for all variables.
2. Compute correlations among all variables.
3. Draw a matrix (9x9) scatterplot for the relationships of CRIME and 8 predictors.
4. Apply OLS to estimate the regression coefficients for the full model
5. Conduct multicollinearity diagnostics on the above model and identify the problems of multicollinearity among the independent variables.
6. Conduct Ridge Regression for the model. What is your best "k" value and corresponding ridge regression coefficients. Why?
7. Conduct Principal Component Regression for the model using PCOMIT = 1 To 6 to obtain the transformed coefficients for the X variables. What is your “best” PC regression model. Why?
8. Compare the OLS model with your “best” ridge regression model and PC regression model (in terms of similarities of model coefficients). Show the three models.