

## APM 630 Regression Analysis

### Project #7 – Influence Diagnostics

**Data:** Influence.xls

The data were obtained from the 1988 Statistical Abstract of USA to determine the factors related to the state expenditures on criminal activities (court, police, etc.). The variables are:

(1) ID variable

- State – the name of the States.

(2) Dependent variable:

- EXPEND – state expenditures on criminal activities (\$1000)

(3) Predictor variables:

- BAD – the number of people under criminal supervision.
- CRIME – the crime rate per 100,000.
- LAWYERS – the number of lawyers in the state.
- EMPLOY – the number of people employed in the state.
- POP – the population of the state (1000).

#### Assignment:

1. Compute descriptive statistics for all variables.
2. Compute correlations among all variables.
3. Draw a matrix scatterplot of EXPEND and five predictor variables.
4. Fit the **full model (including the 5 predictor variables)** to the data.
5. Conduct an influence diagnostics to identify possible outliers and high influential points, and evaluate their effects on model fitting and performance.
  - Residual and StudentResidual plots
  - RStudent and  $h_{ii}$
  - Influence diagnostics
  - **Summary table** to show outlier(s) and high influence point(s)