DEPARTMENT OF POLITICAL SCIENCE
AND
INTERNATIONAL RELATIONS
Posc/Uapp 816

Assignment 12
DATA ANALYSIS

Name_____
(Printed)
Student Number_____
(Social Security Number)
E-mail_____

First retrieve the data set "Boston Cities" from the class web site. Some of you may remember that these are the data used in 815. Anyway, I found them in the StatLib where you can obtain the background information.[1] The data on the web page come in two versions: One contains 14 variables for 506 towns in the Boston area. The other contains a subset of 4 of these variables. The smaller one will "fit" in the Student Version of MINITAB for Windows. (The full version will, of course, accommodate either set.)

The information was apparently collected to investigate air quality on housing prices, but I want you to use them to write a report on crime in the Boston area.

You should retrieve the set that your "system" can handle. The stripped down or minimal batch is easier to work with but doesn't allow you to investigate as many hypotheses. In any event you should save the files on your diskette or hard drive as you have been doing so that you can analyze them as needed.

Here are the variables in the full set:
c1 per capita crime rate by town
c2 proportion of residential land zoned for lots over 25,000 sq.ft.
c3 proportion of non-retail business acres per town
c4 Charles River dummy variable (= 1 if tract bounds river; 0 otherwise)
c5 nitric oxides concentration (parts per 10 million)
c6 average number of rooms per dwelling
c7 proportion of owner-occupied units built prior to 1940

---

[1]Harrison, D. and Rubinfeld, D.L. 'Hedonic Prices and the Demand for Clean Air', J. Environ. Economics & Management, vol.5, 81-102, 1978.   They were used in Belsley, Kuh & Welsch, ***Regression diagnostics*** , Wiley, 1980.   N.B. Various transformations are used in the table on pages 244-261 of the latter.

c8 weighted distances to five Boston employment centers
c9 index of accessibility to radial highways
c10 full-value property-tax rate per $10,000
c11 pupil-teacher ratio by town
c12 $1000(Black - 0.63)^2$ where Black is the proportion of blacks by town
c13 % lower status of the population
c14 Median value of owner-occupied homes in $1000's

Here are the variables for the **smaller** version:
c1 Per capita crime rate by town
c2 Proportion of owner-occupied units built prior to 1940
c3 % lower status of the population
c4 Median value of owner-occupied homes in $1000's

Your task is to explain variation in crime rates by reference to the independent variables. Not all of them will be "important," so as a guideline select a set of two to five, say, and try to build a model that explains as much of the variation in the crime rate (c1) as possible. It would be nice if you used the large data set and included about half a dozen variables. But if you don't have access to the large version of MINITAB, by all means use the abbreviated data.

Your analysis should attempt to use as many of the analytic and diagnostic methods that we have covered as you can. Obviously time series and logistic regression will not be part of the analysis. But, in addition to the one provided, you could construct a "dummy" variable by dividing the data in half: those cities above the median on some factor and those below it. Use your judgement and imagination. There really isn't a completely right and wrong way to do analyze these data.
         Your report should include:
a.   A summary of the **substantive** findings.
b.   Tables and graphs that support these results.
      i.   Make sure everything is properly labeled. Remember: this is a report to the city commission or whatever.
c.   An appendix that shows the steps you followed in developing the model and conclusions.
      i.   It could contain, for instance, plots of residuals, some variables discarded because of colinearity problems, correlation matrices, descriptive statistics, and the like..

It's VERY IMPORTANT that you be neat and organized. I don't want to have to reconstruct your analysis. I would suggest preparing a first draft, fine tune it, and then submit a "clean" copy.

I've left the assignment open ended to see how well you can use statistics to throw light on a public problem. So I can't give more specific guidance.