Introduction to Econometrics

Summer Term 2017 - In-Class Exercise, Lecture 4

We start with analyzing a dataset which "... combines socio-economic data from the 1990 US Census, law enforcement data from the 1990 US LEMAS survey, and crime data from the 1995 FBI UCR". It can be downloaded from `http://archive.ics.uci.edu/ml/datasets/Communities+and+Crime`.

(i) Find the documentation for the dataset and take a quick look at which variables are present. Get a subset of the dataset which includes the following variables: household size, percentage of people between the ages of 16 to 24, percentage of people living in urban cities, percentage of people living with social security, per capita income, percent of unemployed, police budget per population and number of violent crimes per 100k population. Understand what each of these variables mean.
(ii) Regress number of violent crimes on all of the other variables at the same time. Also run a regression where police budget is excluded. For each, find which variables are significant at 5% level. Provide 95% confidence intervals for each variable's coefficients.
(iii) Provide the combined results of these regressions in a nice table using the R package *stargazer*. **(iv)** Try to come up with another variable from the original list, that becomes significantly different from zero on 1 percent level even after using all the controls at part ii except for police budget.