Introduction

Your team has been hired to provide research for a political campaign. They have obtained a dataset of crime statistics for a selection of counties in North Carolina.

Your task is to examine the data to help the campaign understand the determinants of crime and to generate policy suggestions that are applicable to local government.

The Data

The data is provided in a file, crime_v2.csv. It was first used in a study by Cornwell and Trumball, researchers from the University of Georgia and West Virginia University (C. Cornwell and W. Trumball (1994), "Estimating the Economic Model of Crime with Panel Data," Review of Economics and Statistics 76, 360-366.) While we are only providing you with a single cross-section of data, the original study was based on a multi-year panel. The authors used panel data methods and instrumental variables to control for some types of omitted variables. Since you are restricted to ordinary least squares regression, omitted variables will be a major obstacle to your estimates. You should aim for causal estimates, while clearly explaining how you think omitted variables may affect your conclusions.

While you are free to look at the Cornwell and Trumball study (or other papers in the vast literature on crime), that's not necessary and may even harm your grade. We want you to focus on learning from the data, which shouldn't require specialized knowledge beyond what's in this document.

The usual disclaimer applies: the data may have been modified by your instructors to test your abilities.

You are given the following codebook:

variable	label
1 county	county identifier
2 year	1987
3 crmrte	crimes committed per person
4 prbarr	'probability' of arrest
5 prbconv	'probability' of conviction
6 prbpris	'probability' of prison sentence
7 avgsen	avg. sentence, days
8 polpc	police per capita
9 density	people per sq. mile
10 taxpc	tax revenue per capita
11 west	=1 if in western N.C.
12 central	=1 if in central N.C.
13 urban	=1 if in SMSA
14 pctmin80	perc. minority, 1980
15 wcon	weekly wage, construction
16 wtuc	wkly wge, trns, util, commun
17 wtrd	wkly wge, whlesle, retail trade
18 wfir	wkly wge, fin, ins, real est
19 wser	wkly wge, service industry
20 wmfg	wkly wge, manufacturing
21 wfed	wkly wge, fed employees
22 wsta	wkly wge, state employees
23 wloc	wkly wge, local gov emps
24 mix	offense mix: face-to-face/other
25 pctymle	percent young male

In the literature on crime, researchers often distinguish between the certainty of punishment (do criminals expect to get caught and face punishment) and the severity of punishment (for ex- ample, how long prison sentences are). The former concept is the motivation for the 'probability' variables. The probability of arrest is proxied by the ratio of arrests to offenses, measures drawn from the FBI's Uniform Crime Reports. The probability of conviction is proxied by the

ratio of convictions to arrests, and the probability of prison sentence is proxied by the convictions resulting in a prison sentence to total convictions. The data on convictions is taken from the prison and probation files of the North Carolina Department of Correction.

The percent young male variable records the proportion of the population that is male and between the ages of 15 and 24. This variable, as well as percent minority, was drawn from census data.

The number of police per capita was computed from the FBI's police agency employee counts.

The variables for wages in different sectors were provided by the North Carolina Employment Security Commission.