Lab 4: Reducing Crime

w203 Instructional Team

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, contained in the file crime.csv.

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.

You have been given the following codebook:

variable	label
1 county	county identifier
2 year	1988
3 crime	crimes committed per person
4 probarr	'probability' of arrest
5 probconv	'probability' of conviction
6 probsen	'probability' of prison sentence
7 avgsen	avg. sentence, days
8 police	police per capita
9 density	people per sq. mile
10 tax	tax revenue per capita
11 west	=1 if in western part of the state
12 central	=1 if in central part of the state
13 urban	=1 if in Standard Metropolitan Statistical Area
14 pctmin	proportion that is minority or nonwhite
15 wagecon	weekly wage, construction
16 wagetuc	weekly wage, transportation, utilities, communications
17 wagetrd	weekly wage, wholesle, retail trade
18 wagefir	weekly wage, finance, insurance and real estate
19 wageser	weekly wage, service industry
20 wagemfg	weekly wage, manufacturing
21 wagefed	weekly wage, federal employees
22 wagesta	weekly wage, state employees
23 wageloc	weekly wage, local government employees
24 mix	ratio of face to face/all other crimes
25 ymale	proportion of county males between the ages of 15 and 24

As this is a policy exercise, you should do your best to address the campaign's questions from a causal perspective. At the same time, you should clearly explain the limitations of your analysis, and provide discussion around whether your estimates suffer from endogeneity bias.

Assignment

You may work in a team of up to 3 students. This is not a requirement, but we strongly encourage you to form a group and believe it will add considerable value to the exercise.

When working in a group, do not use a "division-of-labor" approach to complete the lab. All students should participate in all aspects of the final report.

Prepare a report investigating the determinants of crime and addressing the concerns of the political campaign.

A successful submission will include

1. A brief introduction
2. An initial exploratory analysis. Detect any anomalies, including missing values, top-coded or bottom-coded variables, etc.
3. A model building process, supported by exploratory analysis. Your EDA should be interspersed with, and support, your modeling decisions. In particular, you should use exploratory techniques to address

- What transformations to apply to variables and what new variables should be created.
- What variables should be included in each model
- Whether model assumptions are met

4. A minimum of three model specifications. In particular, you should include

- One model with only the explanatory variables of key interest (possibly transformed, as determined by your EDA), and no other covariates.
- One model that includes key explanatory variables and only covariates that you believe increase the accuracy of your results without introducing bias (for example, you should not include outcome variables that will absorb some of the causal effect you are interested in). This model should strike a balance between accuracy and parsimony and reflect your best understanding of the determinants of crime.
- One model that includes the previous covariates, and most, if not all, other covariates. A key purpose of this model is to demonstrate the robustness of your results to model specification.

5. For your first model, a detailed assessment of the 6 CLM assumptions. For additional models, you should check all assumptions, but only highlight major differences from your first model in your report.
6. A well-formatted regression table summarizing your model results. Make sure that standard errors presented in this table are valid. Also, be sure to comment on both statistical and practical significance.
7. A detailed discussion of causality. In particular, include a discussion of what variables are not included in your analysis and the likely direction of omitted variable bias. Highlight any coefficients you find that appear to have the wrong sign from a causal perspective, and explain why this is the case.
8. A brief conclusion with a few high-level takeaways.

You should only use R libraries and statistical techniques covered in this course.

Please limit all submissions to 30 pages.

Submission

Submit your lab via ISVC; please do not submit via email.

Submit 2 files:

1. A pdf file including the summary, the details of your analysis, and all the R codes used to produce the analysis. **Please do not suppress the code in your pdf file.**
2. The Rmd source file used to produce the pdf file.

Each group only needs to submit one set of files. Use the following naming convention for your files:

SectionNumber_lab4_Student1FirstNameLastName_Student2FirstNameLastName.fileExtension

For example, if you are in Section 1 and have two students named John Smith and Jane Doe, you should name your file as follows:

Section1_lab4_JohnSmith_JaneDoe.Rmd

Section1_lab4_JohnSmith_JaneDoe.pdf