Lab 3: Reducing Crime

w203 Summer 2018

Madeleine Bulkow, Kim Darnell, Alla Hale, Emily Rapport

July 15, 2018

1. Introduction

Crime presents a problem in our society, and it is up to local government to implement policies to reduce it. This report examines the available crime data to pick out the determininants of crime. Based on this analysis, we generate several policy suggestions applicable to local government in North Carolina for the late 1980s.

2. The Data

The data from 1987 were collected and combined by Cornwell and Trumball. HERE IS WHERE WE DESCRIBE OUR DATA.

3. Exploratory Data Analysis

First, we must evaluate the available data, clean it by removing anomolous values, and perhaps transform the data.

```
# Import the data
df = read.csv("crime_v2.csv")
#summary(df)
```

Clean up the apostrophe.

It appears that probconv is in percent, while the other two probability estimates (prbarr and prbpris) are fractions. To be able to compare coefficients more easily, let's get all percentage values in percent (0-100).

Remove the points where probabilities exceed 100 %.

```
# Clean the data

## NOTE FROM ALLA: This is just what I did to clean the data.  I am sure this can be done in a more eff
df_calc <- df
df_calc$prbconv <- as.numeric(as.numeric(df$prbconv))
df_calc$prbarr <- df$prbarr * 100
df_calc$prbpris <- df$prbpris * 100
df_calc$pctymle <- df$pctymle * 100
#summary(df_calc)
df_clean <-df_calc[with(df_calc, prbarr <= 100 & wser <= 2000),]
summary(df_clean)
```

```
##      county           year         crmrte            prbarr
##  Min.   :  1.0   Min.   :87   Min.   :0.01062   Min.   : 9.277
##  1st Qu.: 51.0   1st Qu.:87   1st Qu.:0.02216   1st Qu.:20.714
##  Median :101.0   Median :87   Median :0.03002   Median :27.095
##  Mean   :100.5   Mean   :87   Mean   :0.03397   Mean   :28.709
##  3rd Qu.:151.0   3rd Qu.:87   3rd Qu.:0.04086   3rd Qu.:34.339
##  Max.   :197.0   Max.   :87   Max.   :0.09897   Max.   :68.902
```

```
##  NA's   :6       NA's   :6    NA's   :6         NA's   :6
##     prbconv         prbpris          avgsen           polpc
##  Min.   : 3.00   Min.   :15.00   Min.   : 5.450   Min.   :0.000746
##  1st Qu.:25.00   1st Qu.:36.36   1st Qu.: 7.360   1st Qu.:0.001234
##  Median :47.00   Median :42.11   Median : 9.100   Median :0.001485
##  Mean   :46.78   Mean   :40.94   Mean   : 9.571   Mean   :0.001625
##  3rd Qu.:69.00   3rd Qu.:45.52   3rd Qu.:11.330   3rd Qu.:0.001859
##  Max.   :91.00   Max.   :60.00   Max.   :17.410   Max.   :0.004459
##  NA's   :6       NA's   :6       NA's   :6        NA's   :6
##     density            taxpc             west           central
##  Min.   :0.00002   Min.   : 25.69   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.56397   1st Qu.: 30.70   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.99623   Median : 34.87   Median :0.0000   Median :0.0000
##  Mean   :1.45224   Mean   : 38.13   Mean   :0.2472   Mean   :0.3708
##  3rd Qu.:1.57028   3rd Qu.: 41.07   3rd Qu.:0.0000   3rd Qu.:1.0000
##  Max.   :8.82765   Max.   :119.76   Max.   :1.0000   Max.   :1.0000
##  NA's   :6         NA's   :6        NA's   :6        NA's   :6
##      urban            pctmin80           wcon            wtuc
##  Min.   :0.00000   Min.   : 1.541   Min.   :193.6   Min.   :187.6
##  1st Qu.:0.00000   1st Qu.:10.005   1st Qu.:253.2   1st Qu.:375.2
##  Median :0.00000   Median :24.312   Median :283.7   Median :406.5
##  Mean   :0.08989   Mean   :25.331   Mean   :286.9   Mean   :411.5
##  3rd Qu.:0.00000   3rd Qu.:38.061   3rd Qu.:315.2   3rd Qu.:441.6
##  Max.   :1.00000   Max.   :61.942   Max.   :436.8   Max.   :613.2
##  NA's   :6         NA's   :6        NA's   :6       NA's   :6
##       wtrd             wfir             wser            wmfg
##  Min.   :154.2   Min.   :170.9   Min.   :133.0   Min.   :157.4
##  1st Qu.:191.2   1st Qu.:288.5   1st Qu.:229.0   1st Qu.:289.4
##  Median :203.0   Median :317.3   Median :253.2   Median :320.2
##  Mean   :212.0   Mean   :322.5   Mean   :254.5   Mean   :335.3
##  3rd Qu.:225.5   3rd Qu.:348.0   3rd Qu.:278.1   3rd Qu.:358.9
##  Max.   :354.7   Max.   :509.5   Max.   :391.3   Max.   :646.9
##  NA's   :6       NA's   :6       NA's   :6       NA's   :6
##       wfed             wsta             wloc            mix
##  Min.   :326.1   Min.   :258.3   Min.   :239.2   Min.   :0.01961
##  1st Qu.:403.1   1st Qu.:329.2   1st Qu.:297.2   1st Qu.:0.08101
##  Median :450.3   Median :357.7   Median :308.1   Median :0.10230
##  Mean   :443.6   Mean   :357.6   Mean   :312.0   Mean   :0.13006
##  3rd Qu.:478.5   3rd Qu.:383.7   3rd Qu.:329.2   3rd Qu.:0.15237
##  Max.   :598.0   Max.   :499.6   Max.   :388.1   Max.   :0.46512
##  NA's   :6       NA's   :6       NA's   :6       NA's   :6
##     pctymle
##  Min.   : 6.216
##  1st Qu.: 7.463
##  Median : 7.787
##  Mean   : 8.425
##  3rd Qu.: 8.354
##  Max.   :24.871
##  NA's   :6
```

4. The Models

4.1 Model 1

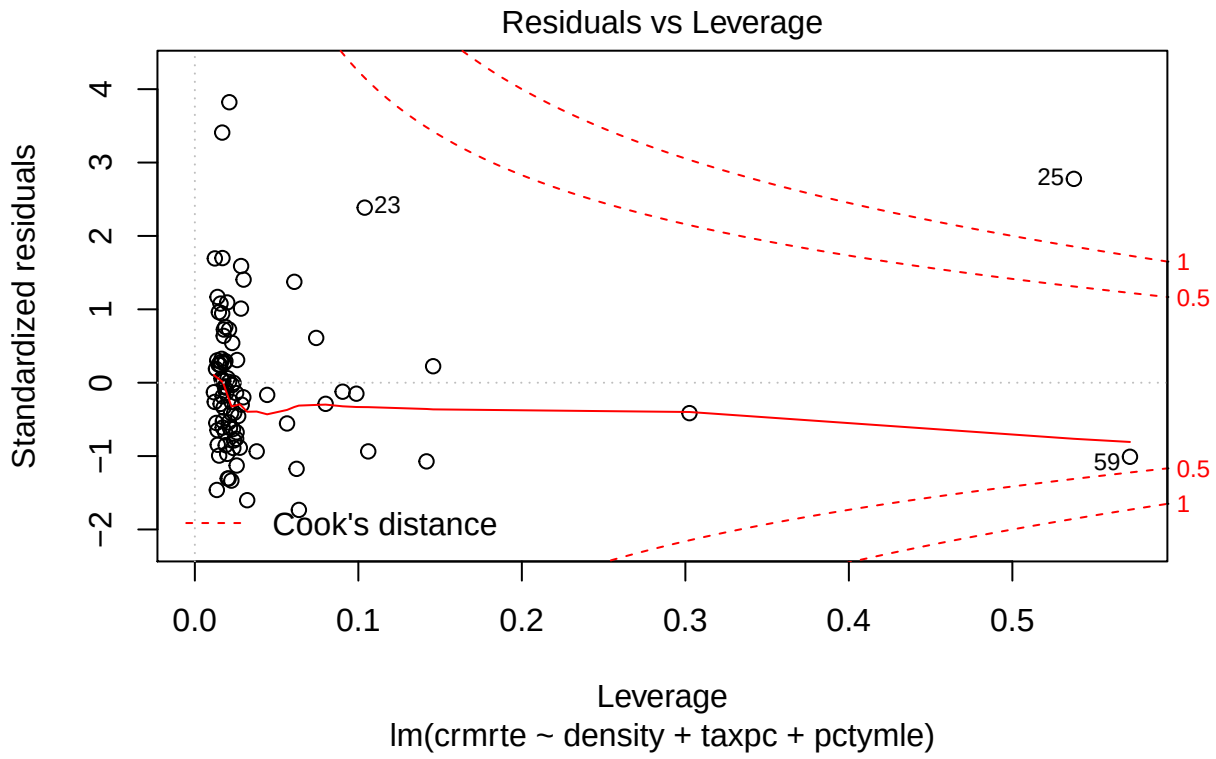
```
# Build Model 1
# model 1: things that totally make sense and have good r^2. density, taxpc, pctymle.
(model_1 = lm(crmrte ~ density + taxpc + pctymle, data = df_clean))
```

```
##
## Call:
## lm(formula = crmrte ~ density + taxpc + pctymle, data = df_clean)
##
## Coefficients:
## (Intercept)      density        taxpc      pctymle
##  -0.0079797    0.0074778    0.0003933    0.0019095
```

```
summary(model_1)$r.square
```

```
## [1] 0.638322
```

```
plot(model_1, which = 5)
```




```
#(model_1 = lm(crmrte ~ prbarr + log(prbconv) + log(polpc) + density + taxpc + pctmin80 + pctymle , dat
```

4.2 Model 2

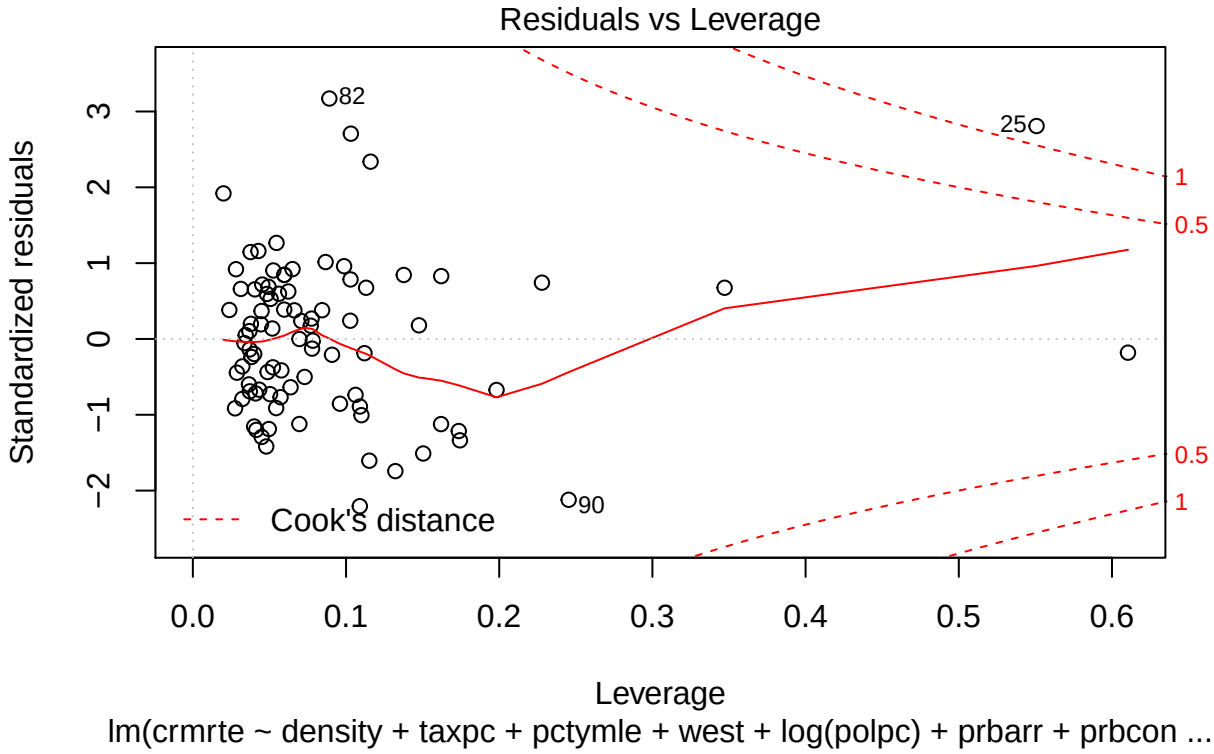
```
# Build Model 2
# model 2: other things that are explanatory but maybe questionable: west, polpc, arrest/conviction, pp
(model_2 = lm(crmrte ~ density + taxpc + pctymle + west + log(polpc) + prbarr + prbconv, data = df_clea
```

```
##
## Call:
## lm(formula = crmrte ~ density + taxpc + pctymle + west + log(polpc) +
##     prbarr + prbconv, data = df_clean)
##
## Coefficients:
## (Intercept)      density        taxpc      pctymle         west
##   0.1052374    0.0054699    0.0001426    0.0009244   -0.0105045
##  log(polpc)       prbarr      prbconv
##   0.0113736   -0.0003576   -0.0001260
```

```
summary(model_2)$r.square
```

```
## [1] 0.7423581
```

```
plot(model_2, which = 5)
```



4.3 Model 3

```
# Build Model 3
#model 3: not necessarily explanatory, but not problematic: central, avgsen, prison.
(model_3 = lm(crmrte ~ density + taxpc + pctymle + west + log(polpc) + prbarr + prbconv + central + avgs
```

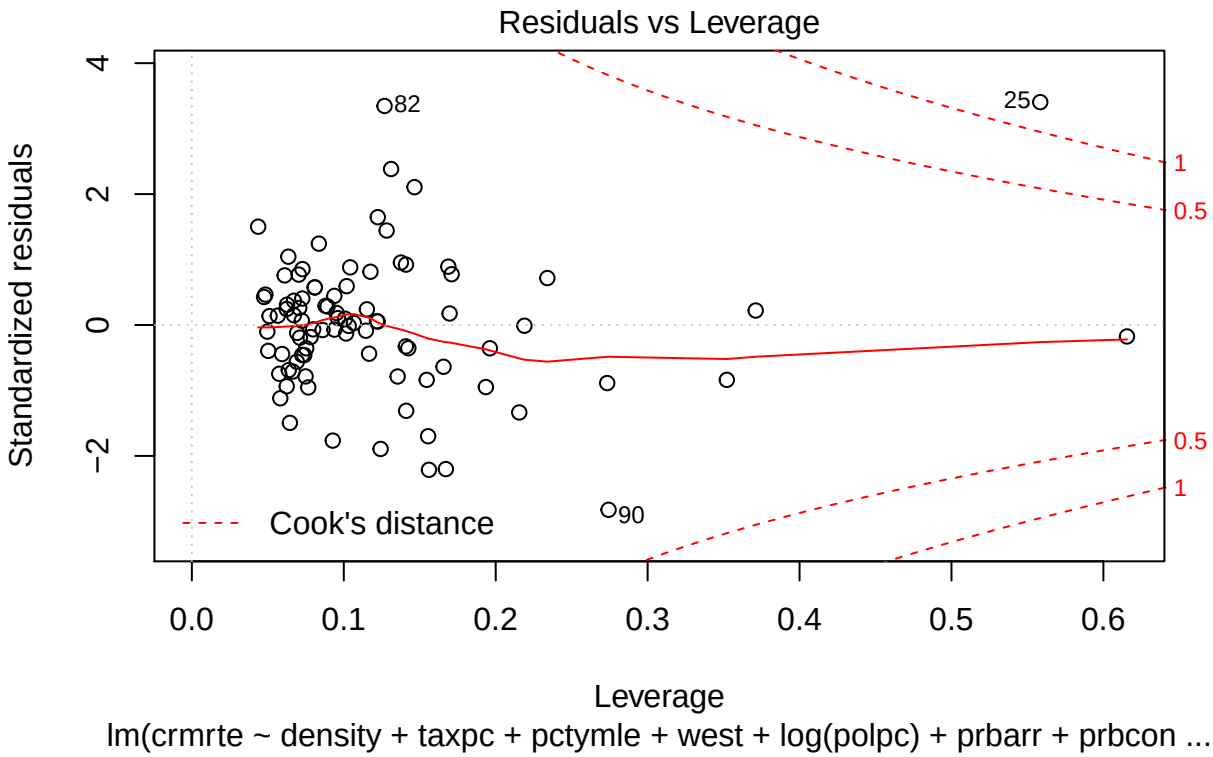
```
##
## Call:
## lm(formula = crmrte ~ density + taxpc + pctymle + west + log(polpc) +
##     prbarr + prbconv + central + avgsen + prbpris, data = df_clean)
##
## Coefficients:
## (Intercept)      density        taxpc      pctymle         west
```

```
##   1.336e-01    6.235e-03    8.738e-05    6.509e-04   -1.378e-02
##  log(polpc)       prbarr       prbconv      central       avgsen
##   1.402e-02   -3.917e-04   -1.307e-04   -7.990e-03   -7.654e-04
##     prbpris
##   1.075e-04
```

```
summary(model_3)$r.square
```

```
## [1] 0.7792778
```

```
plot(model_3, which = 5)
```



4.4 Model 4

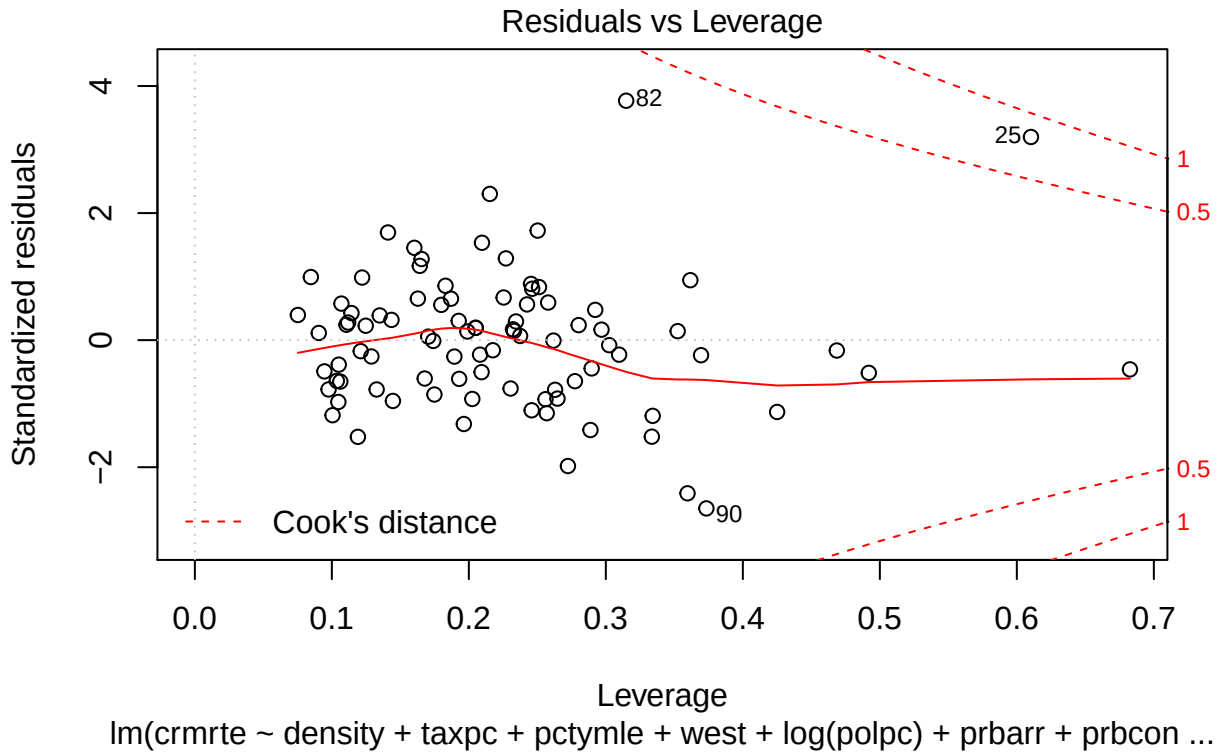
```
# Build Model 4
# model 4: kitchen sink. urban, wage.
(model_4 = lm(crmrte ~ density + taxpc + pctymle + west + log(polpc) + prbarr + prbconv + central + avgs
```

```
##
## Call:
## lm(formula = crmrte ~ density + taxpc + pctymle + west + log(polpc) +
##     prbarr + prbconv + central + avgsen + prbpris + wcon + wtuc +
##     wtrd + wfir + wser + wmfg + wfed + wsta + wloc, data = df_clean)
##
## Coefficients:
## (Intercept)      density        taxpc      pctymle         west
##   9.385e-02    5.512e-03    1.693e-04    1.229e-03   -1.124e-02
##  log(polpc)       prbarr       prbconv      central       avgsen
##   1.275e-02   -3.949e-04   -1.273e-04   -8.832e-03   -1.035e-03
##     prbpris         wcon         wtuc         wtrd         wfir
##   5.379e-05    2.458e-05    1.742e-05    3.242e-05   -2.449e-05
##        wser         wmfg         wfed         wsta         wloc
##  -1.095e-04    8.497e-07    6.575e-05   -3.460e-07    4.584e-05
```

```r
summary(model_4)$r.square
```

```
## [1] 0.8248468
```

```r
plot(model_4, which = 5)
```



4.5 Model 5

```r
# Build Model 5
# model 5: the model 1 version of a model for this dependent variable - crmrate*mix
```

4.2 Model Summary

This is where we put our model summary table.

```r
library(stargazer)
```

```
##
## Please cite as:
```

```
##  Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables.
```

```
##  R package version 5.2.2. https://CRAN.R-project.org/package=stargazer
```

```r
stargazer(model_1, model_2, model_3, model_4, type = "latex",
          report = "vc", # Don't report errors, since we haven't covered them
          title = "Linear Models Predicting Crime Rate",
          keep.stat = c("rsq", "n"),
          omit.table.layout = "n") # Omit more output related to errors
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Sun, Jul 15, 2018 - 19:42:12

Table 1: Linear Models Predicting Crime Rate

	Dependent variable:			
	crmrte			
	(1)	(2)	(3)	(4)
density	0.007	0.005	0.006	0.006
taxpc	0.0004	0.0001	0.0001	0.0002
pctymle	0.002	0.001	0.001	0.001
west		−0.011	−0.014	−0.011
log(polpc)		0.011	0.014	0.013
prbarr		−0.0004	−0.0004	−0.0004
prbconv		−0.0001	−0.0001	−0.0001
central			−0.008	−0.009
avgsen			−0.001	−0.001
prbpris			0.0001	0.0001
wcon				0.00002
wtuc				0.00002
wtrd				0.00003
wfir				−0.00002
wser				−0.0001
wmfg				0.00000
wfed				0.0001
wsta				−0.00000
wloc				0.00005
Constant	−0.008	0.105	0.134	0.094
Observations	89	89	89	89
R^2	0.638	0.742	0.779	0.825

5. Omitted Variables

6. Conclusion