ANALYSIS OF PANEL DATA

An Introduction

datascience@berkeley

Using OLS Regression Model on Panel Data

OLS Regression

- The pros and cons of using OLS regression:
- **Pros**
  - Easy
  - Can apply a model that we've already learned
- **Cons:**
  - Estimates are unreliable.
  - Statistics are invalid as the key underlying assumptions of OLS models are violated.
  - Statistical inferences are incorrect.

OLS Regression – Let's Do It Anyway

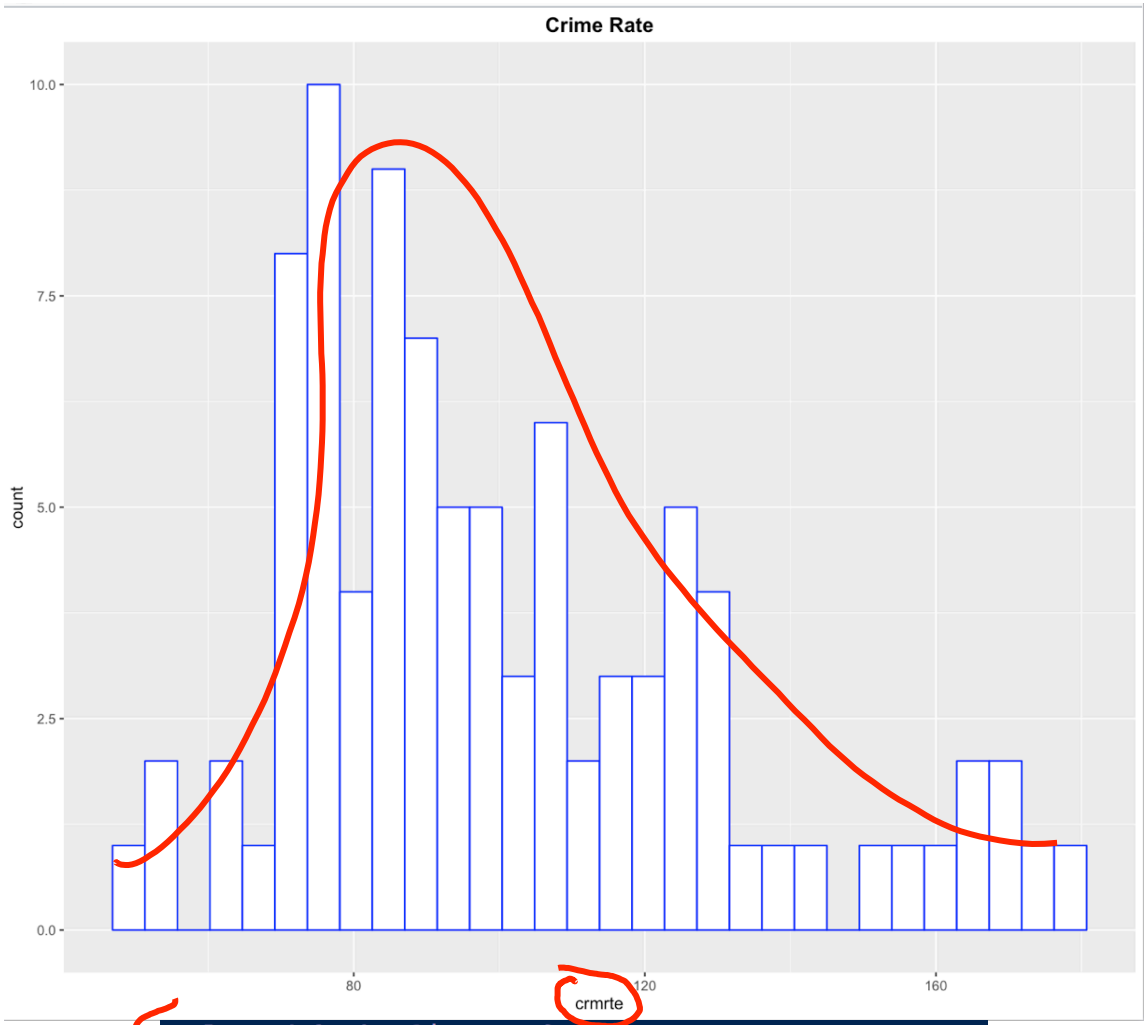
- Let's use a dataset that comes with Wooldridge's text: crime2.

```
'data.frame':   92 obs. of  34 variables:
 $ pop      : num  229528 246815 814054 933177 374974 ...
 $ crimes   : num  17136 17306 75654 83960 31352 ...
 $ unem     : num  8.2 3.7 8.1 5.4 9 ...
 $ officers: int  326 321 1621 1803 633 685 245 259 504 563 ...
 $ pcinc    : int  8532 12155 7551 11363 8343 11729 7592 10802
 $ west     : int  1 1 1 1 1 1 1 1 1 1 ...
 $ nrtheast: int  0 0 0 0 0 0 0 0 0 0 ...
 $ south    : int  0 0 0 0 0 0 0 0 0 0 ...
 $ year     : int  82 87 82 87 82 87 82 87 82 87 ...
 $ area     : num  44.6 44.6 375 375 49.8 ...
 $ d87      : int  0 1 0 1 0 1 0 1 0 1 ...
 $ popden   : num  5146 5534 2171 2488 7530 ...
 $ crmrte   : num  74.7 70.1 92.9 90 83.6 ...
 $ offarea  : num  7.31 7.2 4.32 4.81 12.71 ...
 $ lawexpc  : num  851 2262 875 1070 1122 ...
 $ polpc    : num  1.42 1.3 1.99 1.93 1.69 ...
 $ lpop     : num  12.3 12.4 13.6 13.7 12.8 ...
 $ loffic   : num  5.79 5.77 7.39 7.5 6.45 ...
 $ lpcinc   : num  9.05 9.41 8.93 9.34 9.03 ...
 $ llawexpc: num  6.75 7.72 6.77 6.98 7.02 ...
 $ lpopden : num  8.55 8.62 7.68 7.82 8.93 ...
 $ lcrimes : num  9.75 9.76 11.23 11.34 10.35 ...
 $ larea    : num  3.8 3.8 5.93 5.93 3.91 ...
 $ lcrmrte : num  4.31 4.25 4.53 4.5 4.43 ...
 $ clcrimes: num  NA 0.00987 NA 0.10417 NA ...
 $ clpop    : num  NA 0.0726 NA 0.1366 NA ...
 $ clcrmrte: num  NA -0.0627 NA -0.0324 NA ...
 $ lpolpc   : num  0.351 0.263 0.689 0.659 0.524 ...
 $ clpolpc : num  NA -0.0881 NA -0.0302 NA ...
 $ cllawexp: num  NA 0.978 NA 0.201 NA ...
 $ cunem    : num  NA -4.5 NA -2.7 NA ...
 $ clpopden: num  NA 0.0726 NA 0.1366 NA ...
 $ lcrmrt_1: num  NA 4.31 NA 4.53 NA ...
 $ ccrmrte : num  NA -4.54 NA -2.96 NA ...
```

First Few Observations in the Dataset

```
> head(crime2)
     pop crimes unem officers pcinc west nrtheast south year  area d87   popden
1 229528  17136  8.2      326  8532    1        0     0   82  44.6   0 5146.368
2 246815  17306  3.7      321 12155    1        0     0   87  44.6   1 5533.969
3 814054  75654  8.1     1621  7551    1        0     0   82 375.0   0 2170.811
4 933177  83960  5.4     1803 11363    1        0     0   87 375.0   1 2488.472
5 374974  31352  9.0      633  8343    1        0     0   82  49.8   0 7529.599
6 406297  31364  5.9      685 11729    1        0     0   87  49.8   1 8158.574
    crmrte   offarea   lawexpc    polpc     lpop   loffic   lpcinc llawexpc  lpopden
1 74.65756  7.309417  850.8599 1.420306 12.34378 5.786897 9.051579 6.746247 8.546046
2 70.11729  7.197309 2262.4399 1.300569 12.41639 5.771441 9.405496 7.724199 8.618661
3 92.93487  4.322667  875.0800 1.991268 13.60978 7.390799 8.929436 6.774315 7.682856
4 89.97221  4.808000 1069.6400 1.932109 13.74635 7.497207 9.338118 6.975078 7.819424
5 83.61113 12.710844 1121.8999 1.688117 12.83461 6.450470 9.029179 7.022779 8.926597
6 77.19476 13.755020 1545.6000 1.685959 12.91484 6.529419 9.369820 7.343167 9.006824
    lcrimes    larea  lcrmrte     clcrimes      clpop    clcrmrte    lpolpc
1  9.748937 3.797734 4.312912           NA         NA          NA 0.3508723
2  9.758808 3.797734 4.250169 0.0098714828 0.07261372 -0.06274271 0.2628021
3 11.233926 5.926926 4.531899           NA         NA          NA 0.6887718
4 11.338096 5.926926 4.499501 0.1041698456 0.13656807 -0.03239822 0.6586123
5 10.353033 3.908015 4.426177           NA         NA          NA 0.5236138
6 10.353416 3.908015 4.346332 0.0003833771 0.08022785 -0.07984495 0.5223344
       clpolpc  cllawexp cunem   clpopden lcrmrt_1   ccrmrte
1           NA        NA    NA         NA       NA        NA
2 -0.088070214 0.9779520  -4.5 0.07261467 4.312912 -4.540268
3           NA        NA    NA         NA       NA        NA
4 -0.030159533 0.2007623  -2.7 0.13656807 4.531899 -2.962654
5           NA        NA    NA         NA       NA        NA
6 -0.001279354 0.3203883  -3.1 0.08022785 4.426177 -6.416374
```

Histogram of the Variable of Interest: Crimes




```
> length(crime2$crmrte)
[1] 92
> summary(crime2$crmrte)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  50.02   77.22   92.54  100.80  118.90  179.40
```

Berkeley SCHOOL OF INFORMATION