

HW2

Frank Edwards

2/1/2019

Construction of measures

The focal outcome of my analysis is police staffing per 1,000 persons. I suggest police staffing can be partially explained by two predictors: property tax revenue per capita and racial segregation.

To construct a measure of police staffing, I divide a count of all officers by the county population, then multiply by 1,000 to provide a sensible scale for the measure. I measure property tax revenue as a per capita rate to ensure that measures are comparable across places with differing populations. I leave indices of dissimilarity untransformed.

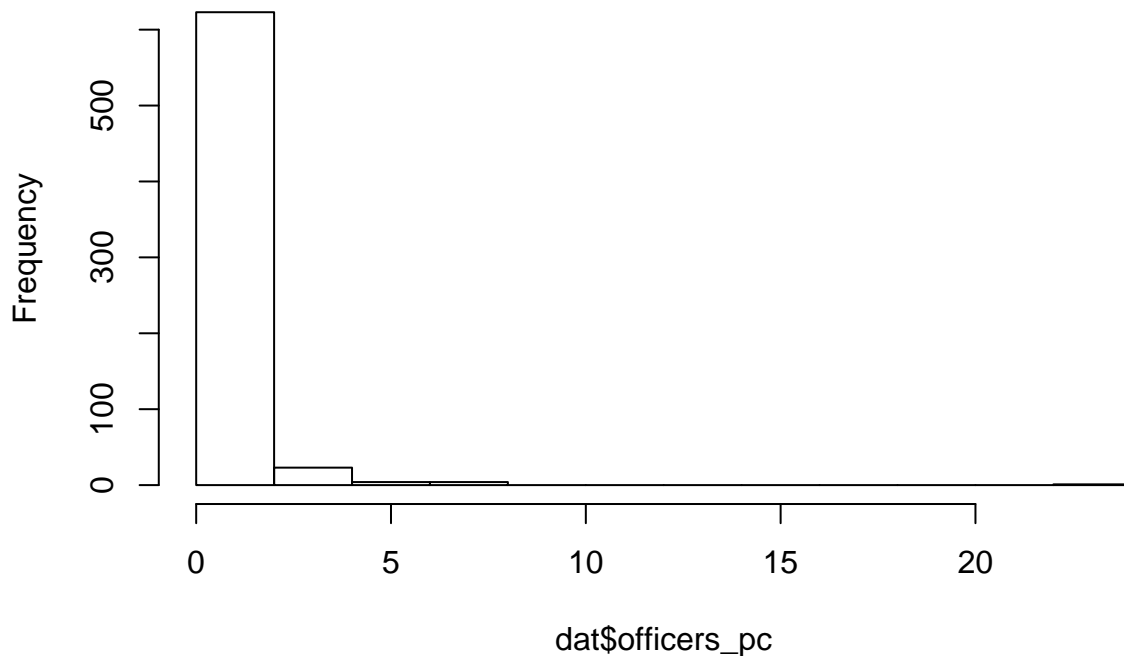
```
dat<-dat%>%  
  mutate(officers_pc = ft_sworn / pop_tot * 1000,  
         prop_tax_pc = rev_prop_tax / pop_tot)
```

Exploratory analysis of officer staffing across counties

I provide plots on both the original and log scales. Because the variable has long left-tails, driven by a small number of counties with very high staffing levels, a log transformation more clearly reveals variation across counties.

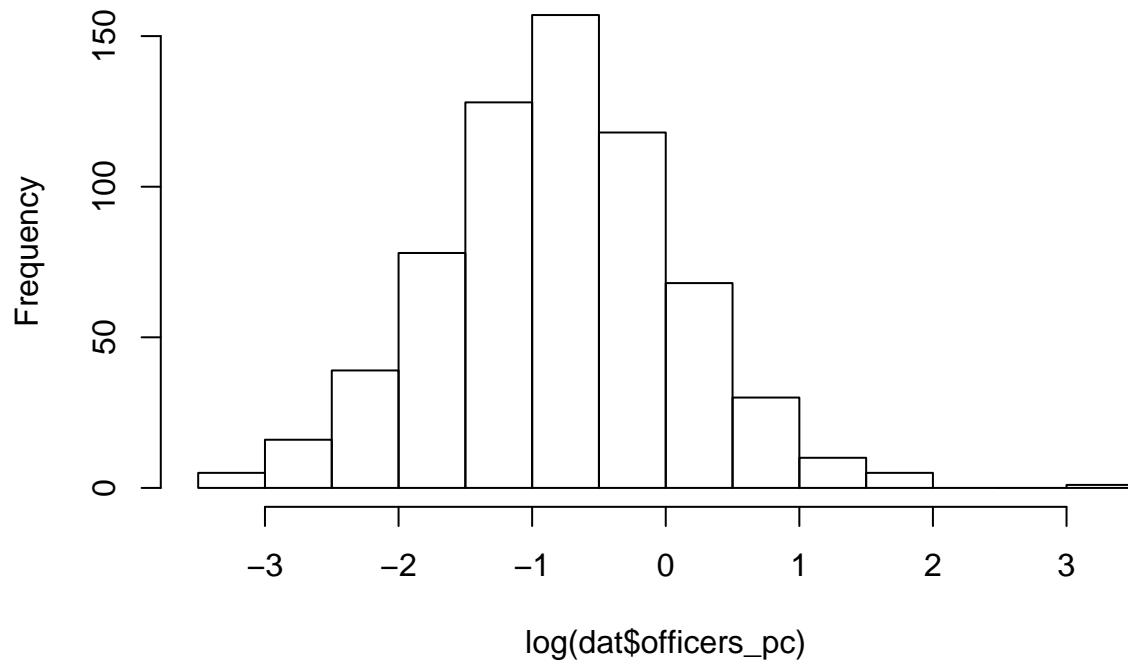
```
hist(dat$officers_pc)
```

Histogram of dat\$officers_pc



```
hist(log(dat$officers_pc))
```

Histogram of $\log(\text{dat\$officers_pc})$



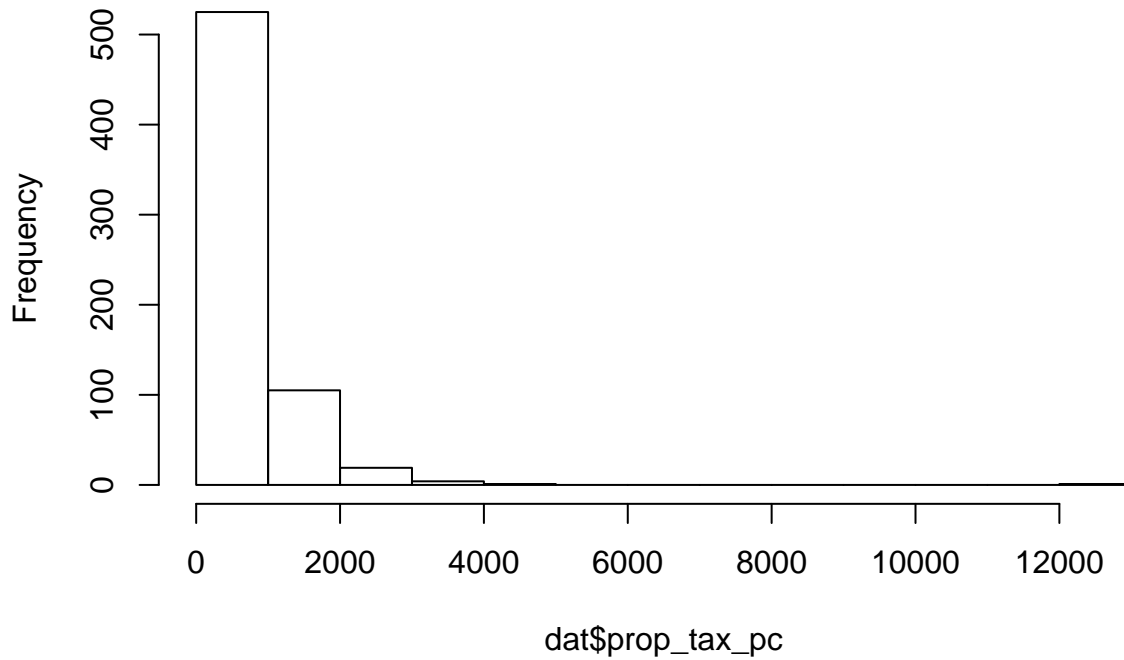
The mean county has about 0.7034727 officers per 1,000 people. The histogram shows that the modal county has about $e^{-0.5}$, or 0.6065307 officers per 1,000. The bulk of the distribution falls below this modal value. A smaller set of counties have very high staffing levels.

Exploratory analysis of focal predictors

Again, a standard histogram is not especially useful in this context, as the variable has exceptionally long tails. A small number of counties have very high per capita tax revenues, while nearly all fall below \$2000 per person.

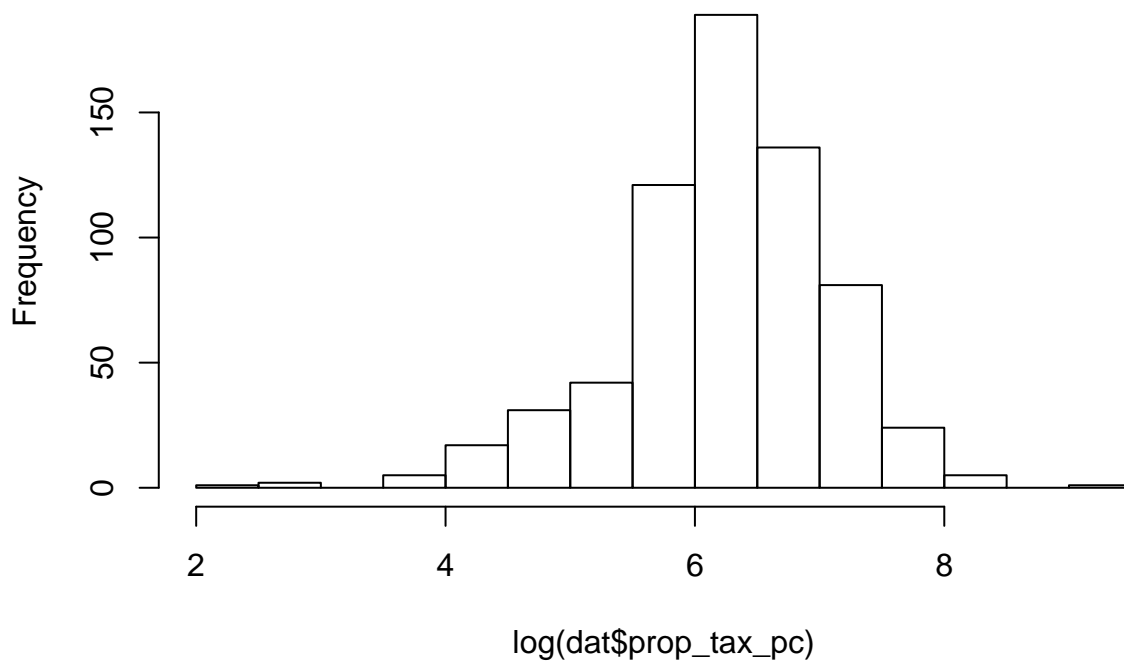
```
hist(dat$prop_tax_pc)
```

Histogram of dat\$prop_tax_pc



```
hist(log(dat$prop_tax_pc))
```

Histogram of log(dat\$prop_tax_pc)

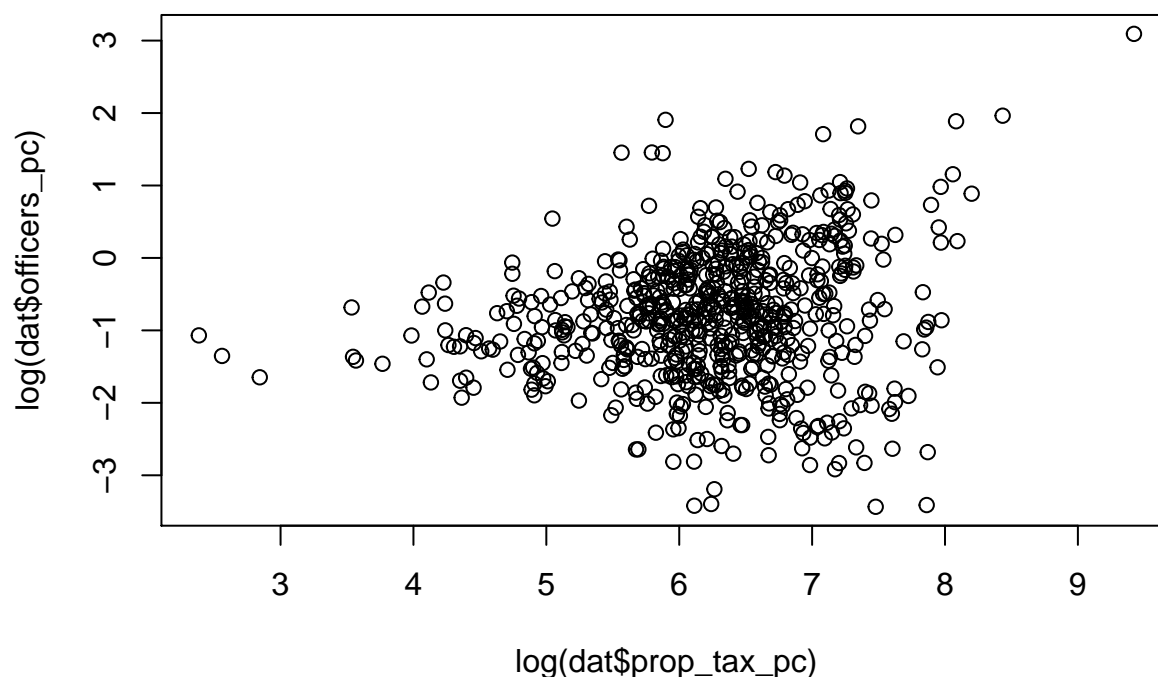


The log(property tax) histogram shows the distribution of counties more clearly, as it compresses extremely high and low values, emphasizing the center of the distribution. The modal county has property tax revenues per capita of about e^6 , or 403.4287935. The average county has revenues of 709.4305379.

Bivariate relationships between officers staffing levels and predictors

Below, I plot the log of property taxes against police officer staffing levels. We can see that there appears to be a weak positive association between tax revenues and officer staffing levels. These (untransformed) measures are correlated at $r = 0.6215803$

```
plot(log(dat$prop_tax_pc), log(dat$officers_pc))
```



##

Regression model

I model officer staffing levels as a function of property taxes and racial segregation. I suggest that places with higher tax revenues and more segregation will have more police per capita. To improve model fit, I log transform the per capita measures.

```
m1<-lm(log(officers_pc) ~ log(prop_tax_pc) + dissim_bw,
      data = dat)
library(broom)
tidy(m1)
```

```
## # A tibble: 3 x 5
##   term          estimate std.error statistic  p.value
##   <chr>          <dbl>    <dbl>    <dbl>    <dbl>
## 1 (Intercept)   -1.86      0.286    -6.48  1.79e-10
## 2 log(prop_tax_pc)  0.187    0.0417     4.49  8.49e- 6
## 3 dissim_bw     -0.00237  0.00296    -0.799 4.24e- 1
```

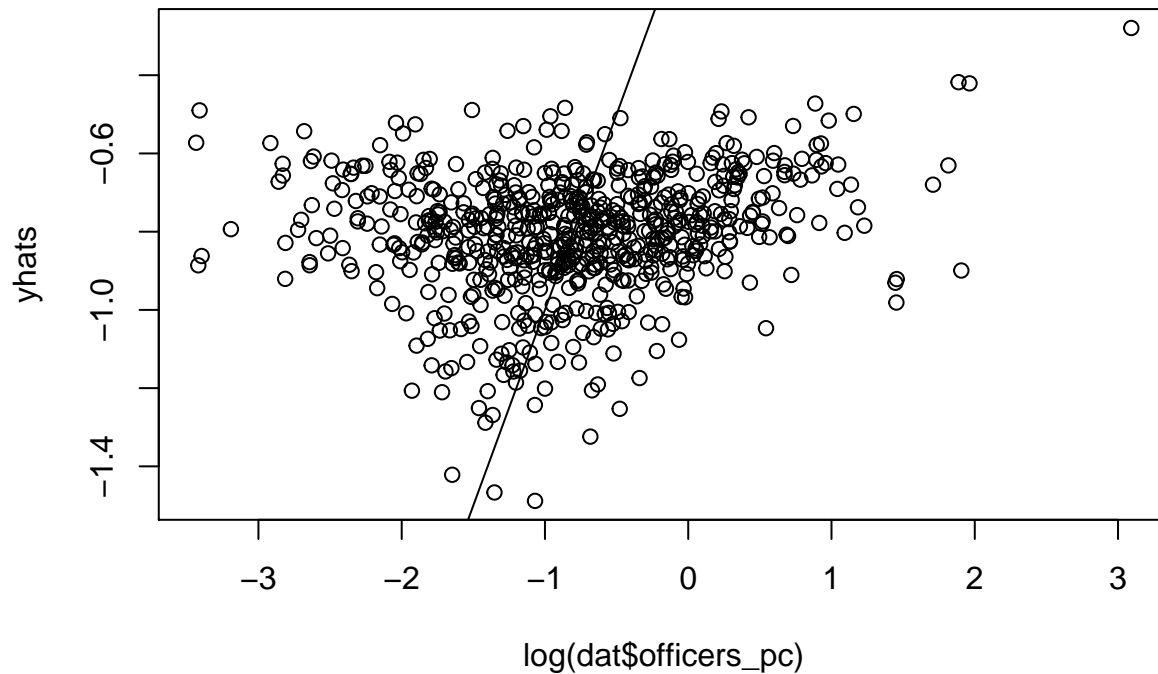
```
fake_data<-data.frame("prop_tax_pc" = quantile(dat$prop_tax_pc, 0.95),
                      "dissim_bw" = quantile(dat$dissim_bw, 0.95))
```

```
yhat<-predict(m1, newdata = fake_data)
```

This model suggests that a county with high property tax revenue (at the 95th percentile) and a county with high segregation (at the 95th percentile) will have 0.5348809 police officers per capita.

```
yhats<-predict(m1, newdata = dat)
plot(log(dat$officers_pc), yhats)
```

```
abline(0,1)
```

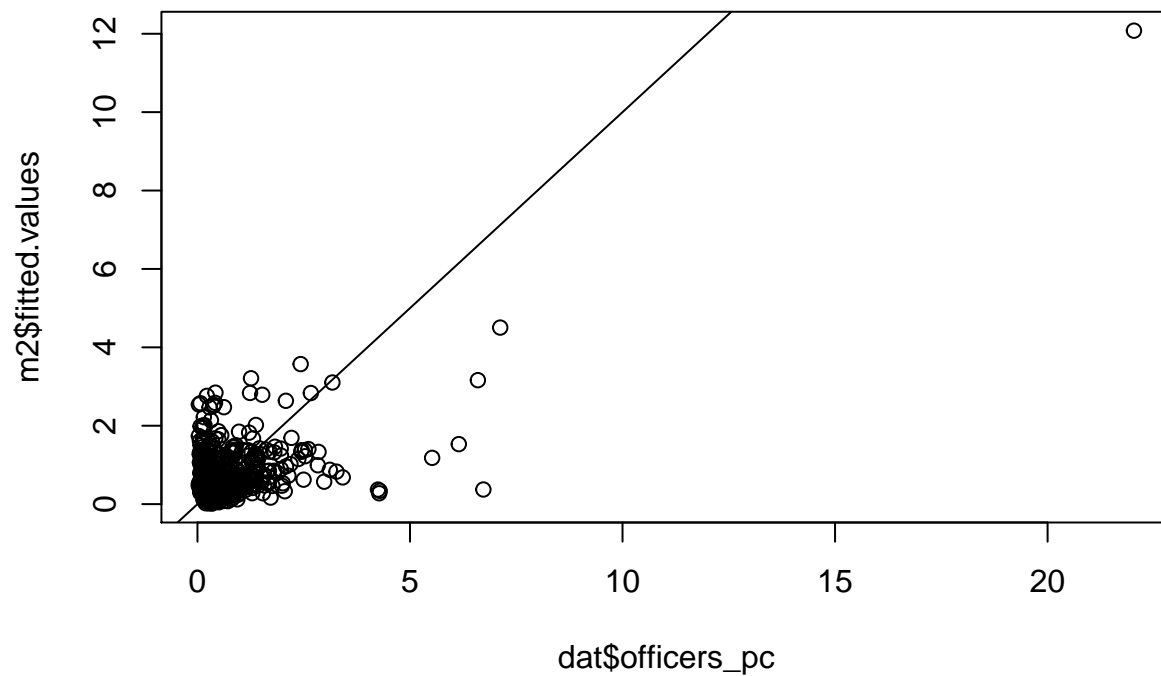


The observed vs predicted plot suggests some major problems with model fit. The model underpredicts many counties in favor of fitting well for a subset of counties with relatively small officer_pc values. This is likely a function of the log transformation, omitted variables, or a poor fit for the normal likelihood model.

```
m2<-lm((officers_pc)~ prop_tax_pc + dissim_bw,
      data = dat)
library(broom)
tidy(m2)
```

```
## # A tibble: 3 x 5
##   term      estimate std.error statistic  p.value
##   <chr>      <dbl>    <dbl>    <dbl>    <dbl>
## 1 (Intercept) -0.0131  0.151     -0.0868 9.31e- 1
## 2 prop_tax_pc  0.000975 0.0000491  19.9    5.59e-69
## 3 dissim_bw    0.000496 0.00300     0.165  8.69e- 1
```

```
plot(dat$officers_pc, m2$fitted.values)
abline(0,1)
```



Here too, we see a very poor model fit. We should consider adding additional predictors and/or seeking another kind of model (hint: Poisson is a good candidate!).

Theoretical interpretation of results

Something brilliant about criminology here.