Analysts of the Boston Dataset

Dominick Terranova dterranova1@une.edu

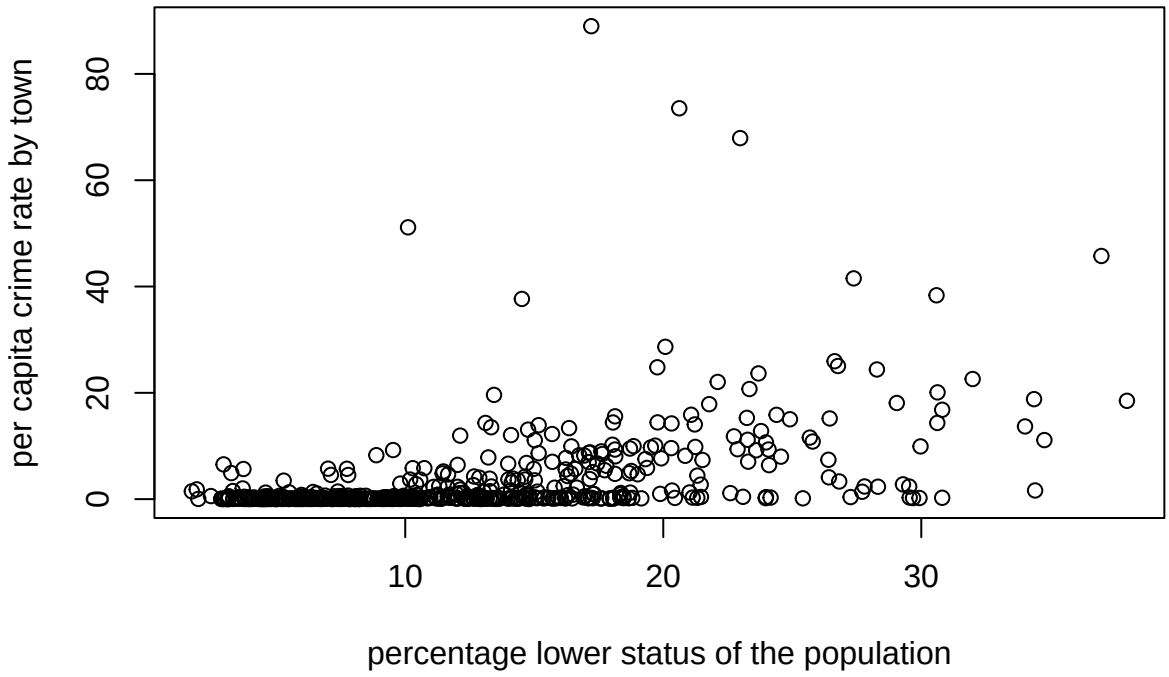
This report is due on 10/26/2023

1 Boston Dataset

```
library(MASS)

dataBoston <- Boston

plot(dataBoston$lstat, dataBoston$crim, xlab = "percentage lower status of the population",
     ylab ="per capita crime rate by town" )
```

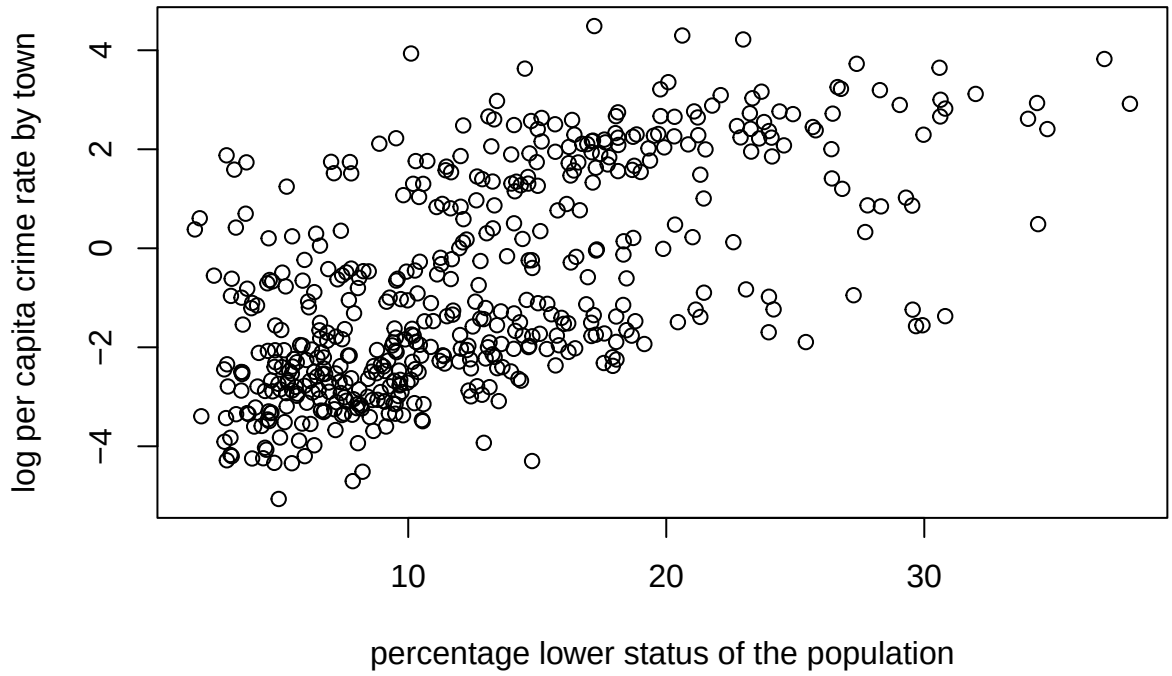



```
lm(dataBoston$crim~dataBoston$lstat, data = dataBoston)
```

```
##
## Call:
## lm(formula = dataBoston$crim ~ dataBoston$lstat, data = dataBoston)
##
## Coefficients:
##      (Intercept)  dataBoston$lstat
##          -3.3305            0.5488
```

#2 Transformed Response Variable

```
plot(dataBoston$lstat, log(dataBoston$crim), xlab = "percentage lower status of the population",
     ylab ="log per capita crime rate by town" )
```



There exist a positive relationship between per capita crime rate by town and percentage lower status of the population and there is a low variation in the variables as compared to the other plot. The strength of the relationship is stronger than what exists without the transposed response variable.

#3 Model Comparison

```
summary( lm(dataBoston$crim~dataBoston$lstat, data = dataBoston) )
```

```
##
## Call:
## lm(formula = dataBoston$crim ~ dataBoston$lstat, data = dataBoston)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -13.925  -2.822  -0.664   1.079  82.862
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -3.33054    0.69376  -4.801 2.09e-06 ***
## dataBoston$lstat  0.54880    0.04776  11.491  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.664 on 504 degrees of freedom
## Multiple R-squared:  0.2076, Adjusted R-squared:  0.206
## F-statistic:   132 on 1 and 504 DF,  p-value: < 2.2e-16
```

```
summary( lm(log(dataBoston$crim)~dataBoston$lstat, data = dataBoston) )
```

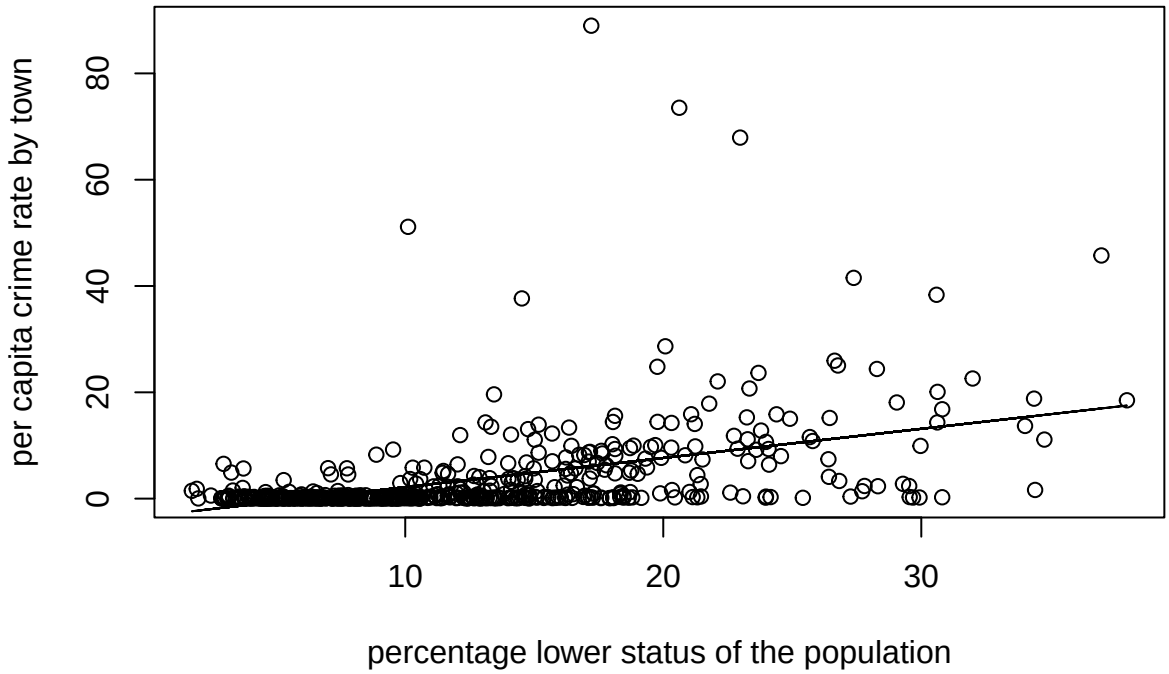
```
##
## Call:
```

```
## lm(formula = log(dataBoston$crim) ~ dataBoston$lstat, data = dataBoston)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.0520 -1.3202 -0.3165  1.3216  5.1974
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -3.18092    0.15267  -20.84   <2e-16 ***
## dataBoston$lstat  0.18972    0.01051   18.05   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.687 on 504 degrees of freedom
## Multiple R-squared:  0.3926, Adjusted R-squared:  0.3914
## F-statistic: 325.8 on 1 and 504 DF,  p-value: < 2.2e-16
```

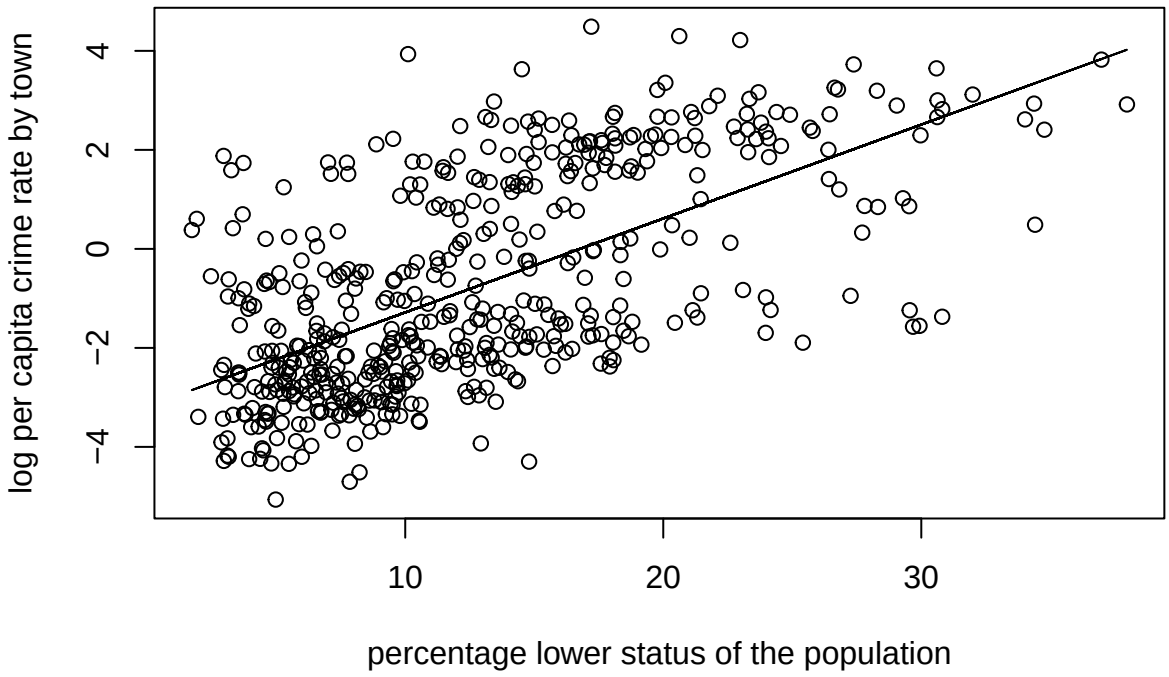
The model with the log transformed response variable has a better fit.

#4 Fitting Regression Lines

```
plot(dataBoston$lstat, dataBoston$crim, xlab = "percentage lower status of the population",
     ylab ="per capita crime rate by town" ,
     points(dataBoston$lstat, lm(dataBoston$crim~dataBoston$lstat)$fitted, type = "l"   ) )
```




```
plot(dataBoston$lstat, log(dataBoston$crim), xlab = "percentage lower status of the population",
     ylab ="log per capita crime rate by town"  )
points(dataBoston$lstat, lm(log(dataBoston$crim)~dataBoston$lstat)$fitted, type = "l"   )
```


log per capita crime rate by town
4
2
0
−2
−4
10
20
30
percentage lower status of the population