

# Homework3

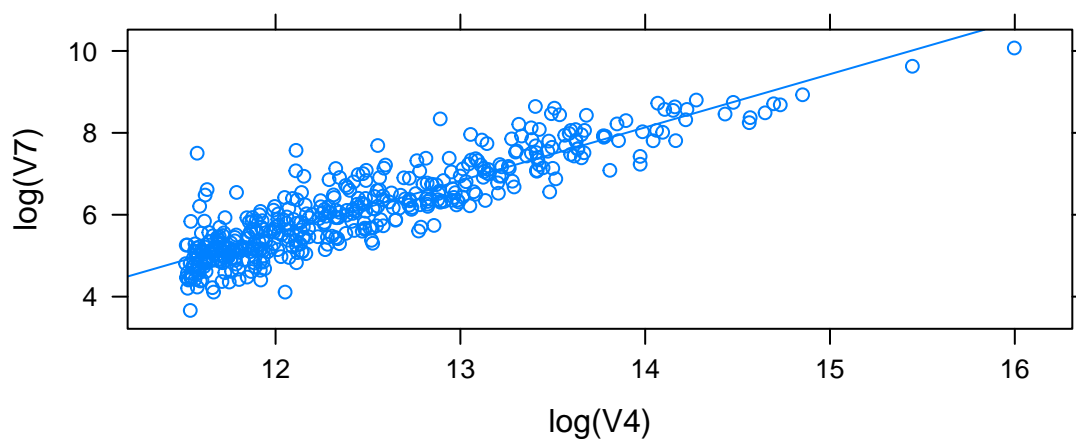
Ethan Ashby

2/19/2020

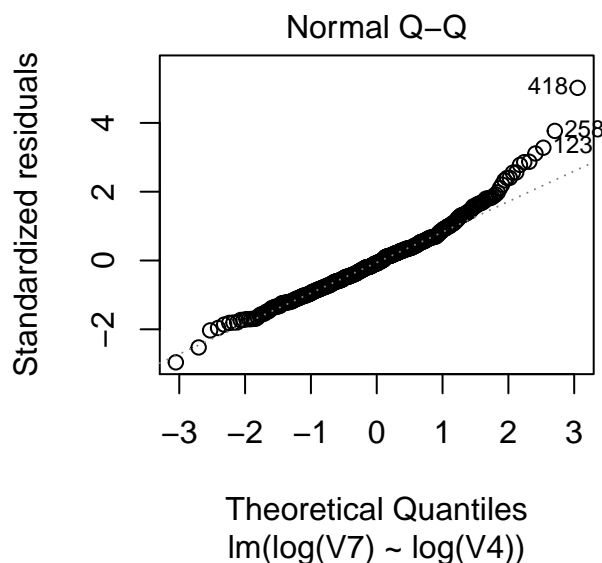
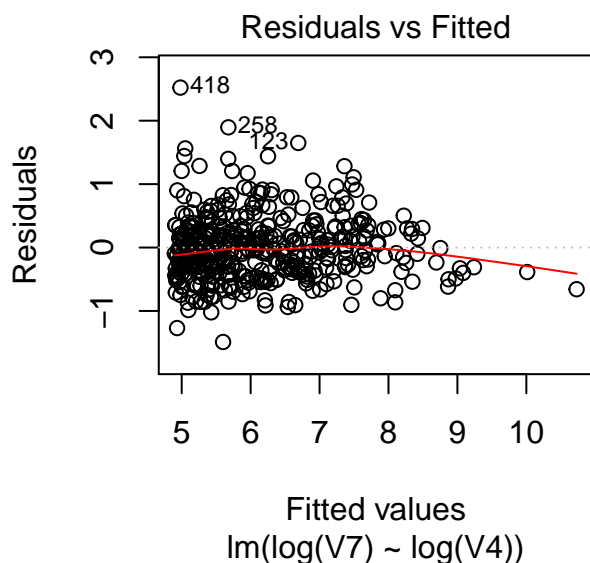
We begin by fitting a model to predict number of physicians using the estimated population.

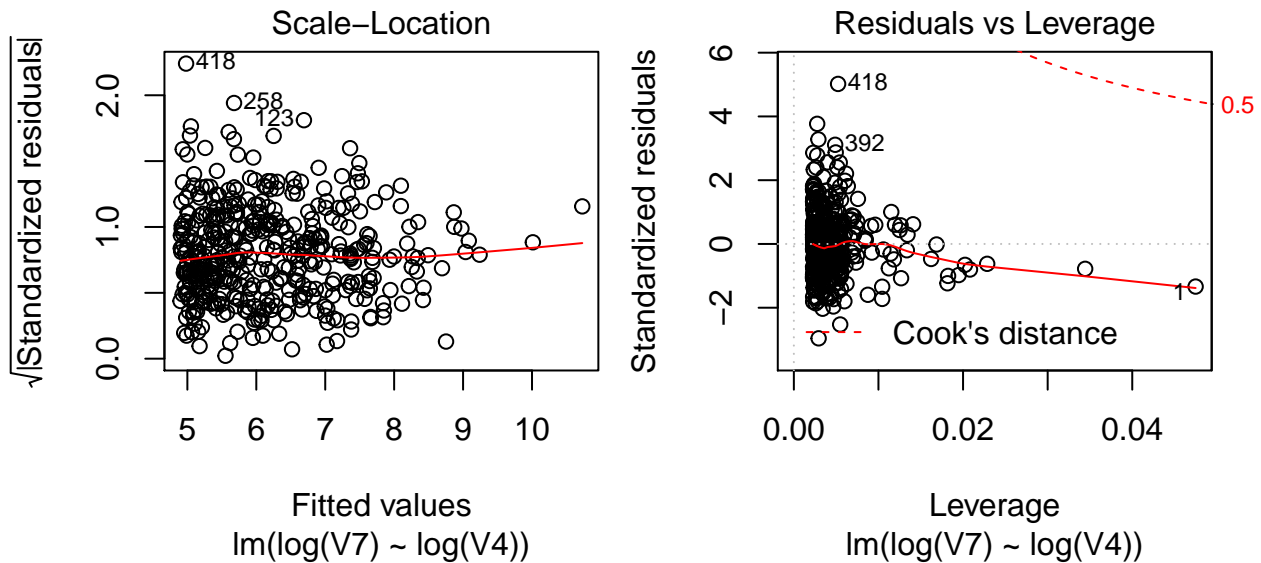
We use the box-cox procedure to inform possible data transformations. The log-likelihood is maximized at around  $\lambda=0$ , which suggests a log transformation of the y variable (number of physicians). When we refit and rerun box cox, we also see a log transform of the x variable is suggested

```
#build model predicting number of physicians using log(pop)
new_fit<-lm(log(V7)~log(V4), data=data)
xyplot(log(V7)~log(V4), data=data, type=c("r", "p"))
```



#this looks a lot better





Our diagnostic plots spell pretty good news for our updated model. The residuals vs fitted plot shows that the residuals show no trend with the fitted values, indicating constant variance and linearity are upheld in our model. Our Q-Q plot shows that despite a few outliers, that the majority of the residuals are normally distributed, which supports another of our important regression hypotheses. The scale location plot shows the standardized residuals are spread evenly above and below the line, supporting homoskedasticity. Our residuals vs leverage plot shows that there are no influential points that could be messing up our regression.

```
#rounding to 100ths digit leads to solid representation of my data
#xyplot(log(V7)~round(log(V4), 2), data=data, type=c("r", "p"))
formal.test.linearity(log(data$V4), log(data$V7), digits=-2)
```

```
## Analysis of Variance Table
##
## Model 1: y ~ x.round
## Model 2: y ~ as.factor(x.round)
##   Res.Df    RSS Df Sum of Sq   F Pr(>F)
## 1     438 111.072
## 2     230  69.627 208   41.445 0.6582 0.9989
```

And our p-value is very large, so we fail to reject the null hypothesis of linearity.

```
## [1] 2678.735 2939.674
## [1] 2678.735 7240.686
```

Confidence interval for a county of 1 million people is [2678.735, 2939.674]. Our prediction interval for a county of 1 million people is [2678.735, 7240.686].