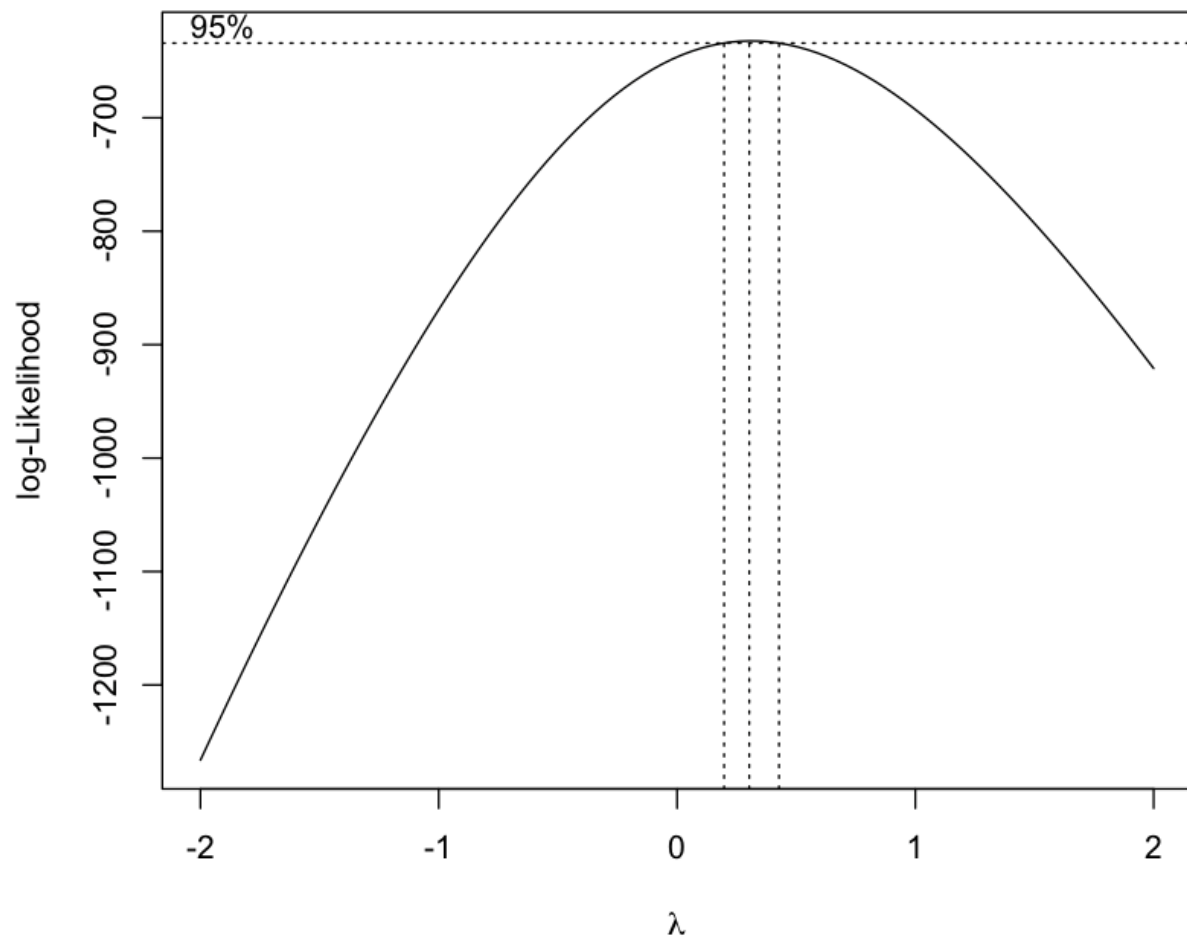


# CS 498 HW6 Report

1. Indexes of removed points:

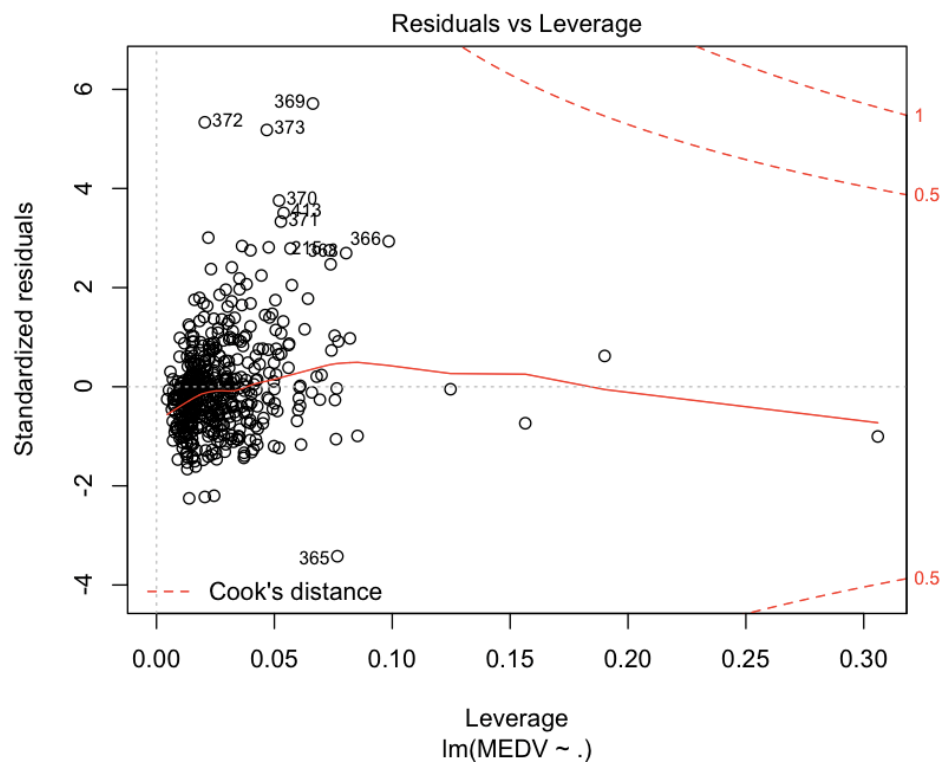
413, 366, 369, 372, 373, 370, 371, 365, 368

2. Log-likelihood vs.  $\lambda$ :

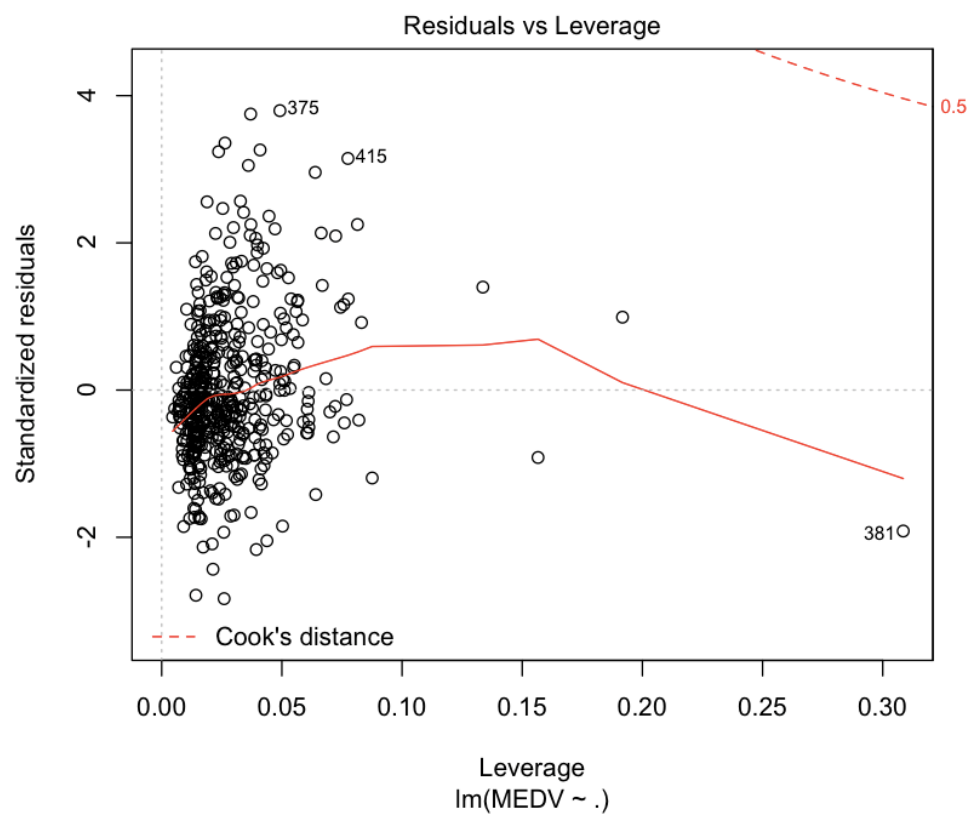


The best lambda we found is 0.3030303.

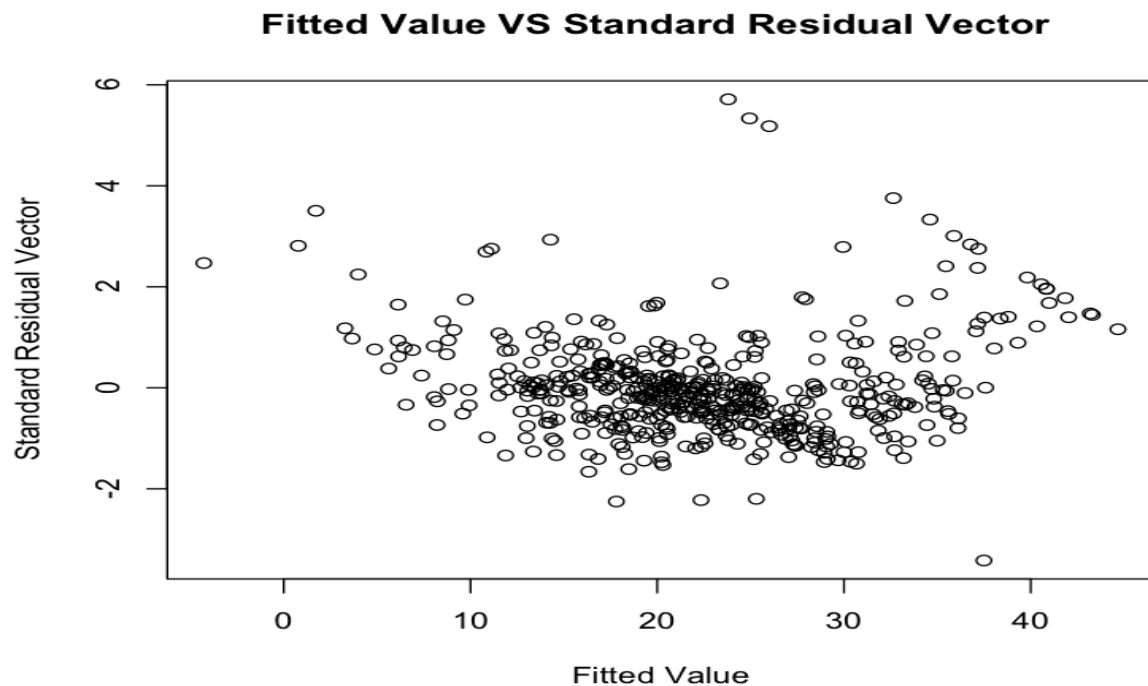
### 3. Diagnostic Plot



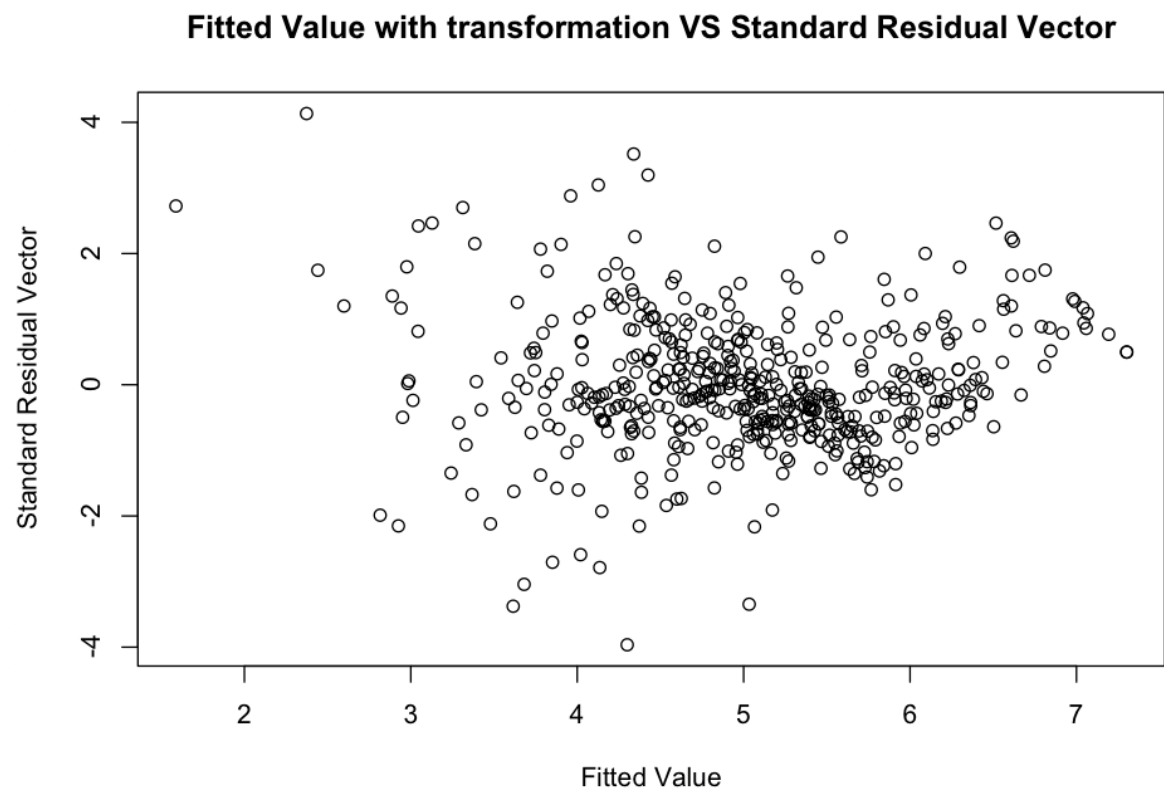
### Final Diagnostic Plot



4. Plot of Standardized Residuals vs. Fitted Value for linear regression

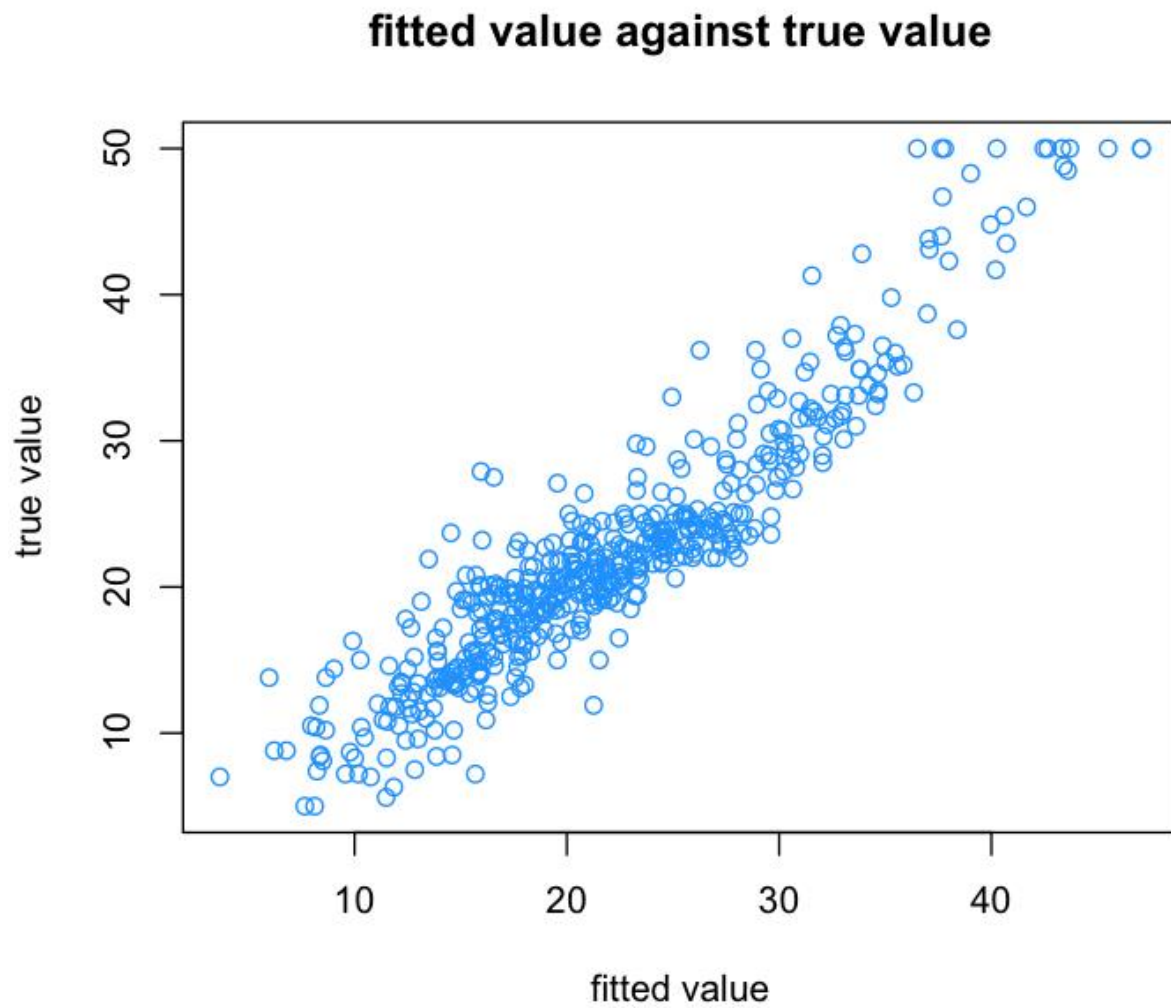


5. Plot of Standardized Residuals vs. Fitted Value for final regression



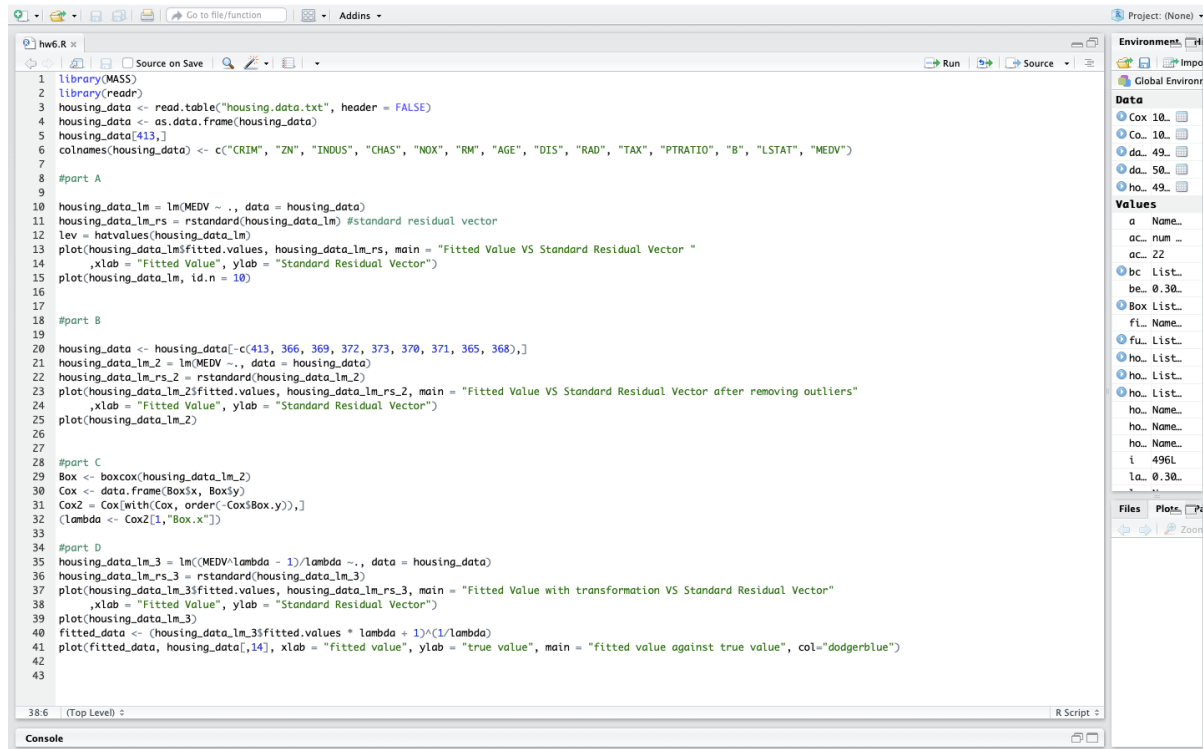
6. Observation: Before transformation, the plot does not look like random normalized. There are some points with extreme standard residuals accumulated around the origin. After the transformation, the plot looks like a random normal distribution, and the residuals are more scattered around the origin.

## 7. Final Plot of Fitted House Price vs. True House Price



Observation: The sample points are distributed around a linear line intercepting y-axis at 0. This outcome implies that fitted values are mostly close to the true values. Therefore, the model is effective and accurate.

## 8. Code Screenshot



The screenshot displays the RStudio interface with a script editor on the left containing R code, an environment pane on the right, and a console at the bottom.

```
1 library(MASS)
2 library(readr)
3 housing_data <- read.table("housing.data.txt", header = FALSE)
4 housing_data <- as.data.frame(housing_data)
5 housing_data[413,]
6 colnames(housing_data) <- c("CRIM", "ZN", "INDUS", "CHAS", "NOX", "RM", "AGE", "DIS", "RAD", "TAX", "PTRATIO", "B", "LSTAT", "MEDV")
7
8 #part A
9
10 housing_data_lm = lm(MEDV ~ ., data = housing_data)
11 housing_data_lm_rs = rstandard(housing_data_lm) #standard residual vector
12 lev = hatvalues(housing_data_lm)
13 plot(housing_data_lm$fitted.values, housing_data_lm_rs, main = "Fitted Value VS Standard Residual Vector "
14       ,xlab = "Fitted Value", ylab = "Standard Residual Vector")
15 plot(housing_data_lm, id.n = 10)
16
17
18 #part B
19
20 housing_data <- housing_data[-c(413, 366, 369, 372, 373, 370, 371, 365, 368),]
21 housing_data_lm_2 = lm(MEDV ~., data = housing_data)
22 housing_data_lm_rs_2 = rstandard(housing_data_lm_2)
23 plot(housing_data_lm_2$fitted.values, housing_data_lm_rs_2, main = "Fitted Value VS Standard Residual Vector after removing outliers"
24       ,xlab = "Fitted Value", ylab = "Standard Residual Vector")
25 plot(housing_data_lm_2)
26
27
28 #part C
29 Box <- boxcox(housing_data_lm_2)
30 Cox <- data.frame(Box$x, Box$y)
31 Cox2 = Cox[with(Cox, order(-Cox$Box.y)),]
32 (Lambda <- Cox2[1,"Box.x"])
33
34 #part D
35 housing_data_lm_3 = lm((MEDV^lambda - 1)/lambda ~., data = housing_data)
36 housing_data_lm_rs_3 = rstandard(housing_data_lm_3)
37 plot(housing_data_lm_3$fitted.values, housing_data_lm_rs_3, main = "Fitted Value with transformation VS Standard Residual Vector"
38       ,xlab = "Fitted Value", ylab = "Standard Residual Vector")
39 plot(housing_data_lm_3)
40 fitted_data <- (housing_data_lm_3$fitted.values * lambda + 1)^(1/lambda)
41 plot(fitted_data, housing_data[,14], xlab = "fitted value", ylab = "true value", main = "fitted value against true value", col="dodgerblue")
42
43
```

The environment pane on the right shows the following objects:

- Data**
  - Cox 10.
  - Co. 10.
  - da. 49.
  - da. 50.
  - ho. 49.
- Values**
  - a Name.
  - ac. num.
  - ac. 22
  - bc List.
  - be. 0.30.
  - Box List.
  - fi. Name.
  - fu. List.
  - ho. List.
  - ho. List.
  - ho. List.
  - ho. Name.
  - ho. Name.
  - ho. Name.
  - i 496L
  - la. 0.30.

The console at the bottom shows the current line number 38:6 and the message "(Top Level)".