CS498 AML,AMO HW6

WORAWICH CHAIYAKUNAPRUK, Worapol Setwipatanachai

TOTAL POINTS

100 / 100

QUESTION 1

1 code for regression and resulting model. **0 / 0**

- ✓ **- 0 pts** **Correct**

QUESTION 2

2 a screenshot of your diagnostic plot and a few sentences of your explanation. **50 / 50**

- **+ 10 pts** Correct
- **+ 50 Point adjustment**

QUESTION 3

3 a screenshot of your new diagnostic plot. **20 / 20**

- **+ 0 pts** Correct
- **+ 20 Point adjustment**

QUESTION 4

4 a screenshot of your code for subproblem 2. **10 / 10**

- **+ 0 pts** Correct
- **+ 0 pts** Incorrect
- **+ 10 Point adjustment**

QUESTION 5

5 a screenshot of Box-Cox transformation plot and the best value you chose. **10 / 10**

- ✓ **- 0 pts** **Correct**
- **- 10 pts** Incorrect

QUESTION 6

6 result of the standardized residuals of the regression after Box-Cox transformation and a plot of fitted house price against true house price. **10 / 10**

- ✓ **- 0 pts** **Correct**
- **- 10 pts** Incorrect

QUESTION 7

7 code for subproblems 3 and 4. **0 / 0**

- ✓ **- 0 pts** **Correct**
- **- 5 pts** Click here to replace this description.

QUESTION 8

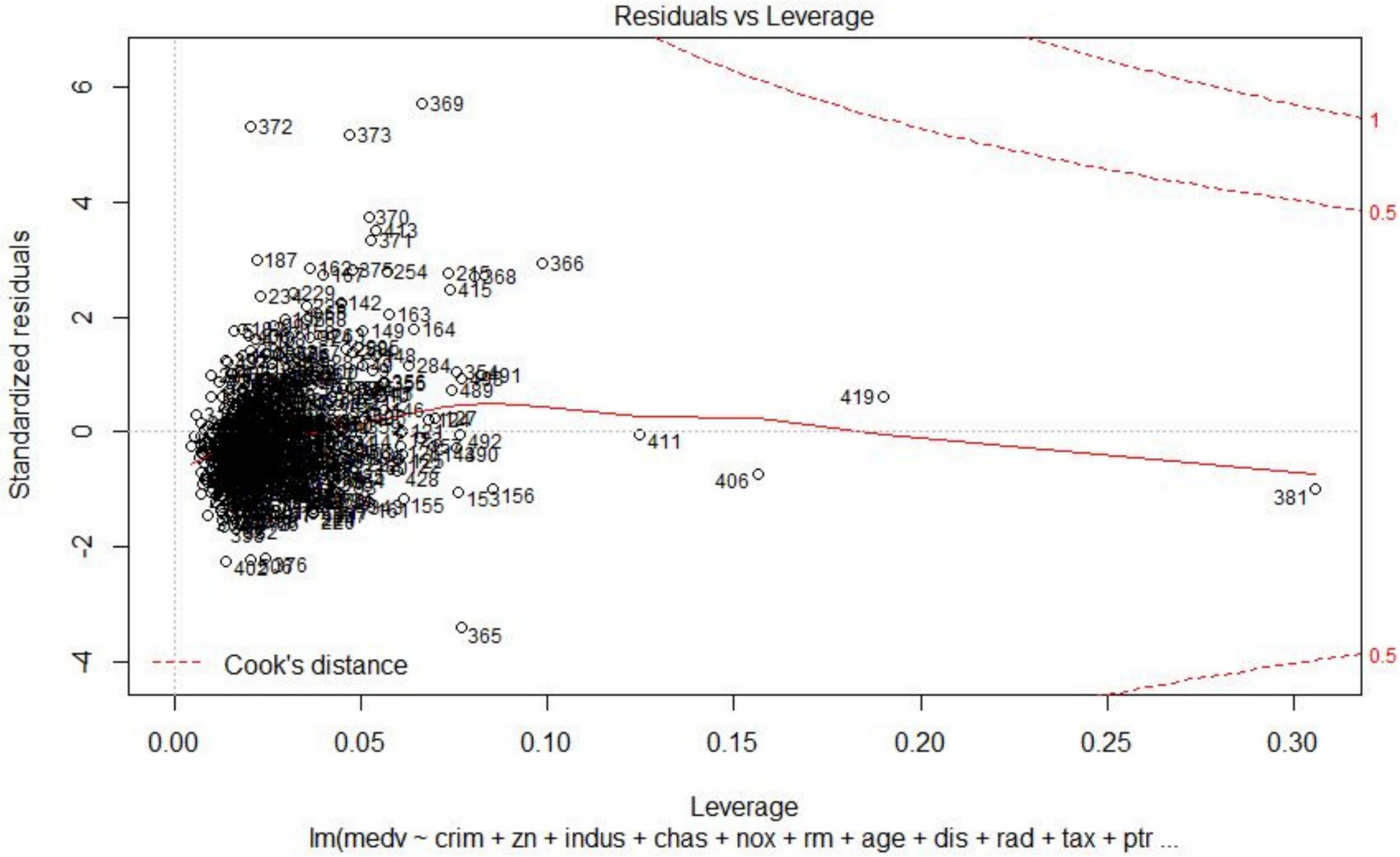
8 late penalty **0 / 0**

- ✓ **- 0 pts** **Correct**
- **- 5 pts** 1 day
- **- 10 pts** 2 days
- **- 15 pts** 3 days
- **- 20 pts** 4 days
- **- 30 pts** max

(**0 points**) **Page 1**: code for regression and resulting model.

```
library(MASS)
data(Boston)
fit1 <- lm(medv ~ crim + zn + indus + chas + nox + rm + age + dis + rad + tax +
ptratio + black + lstat, data=Boston)
par(mfrow=c(1,1)) # Change the panel layout to 2 x 2
plot(fit1, id.n = 3)
```

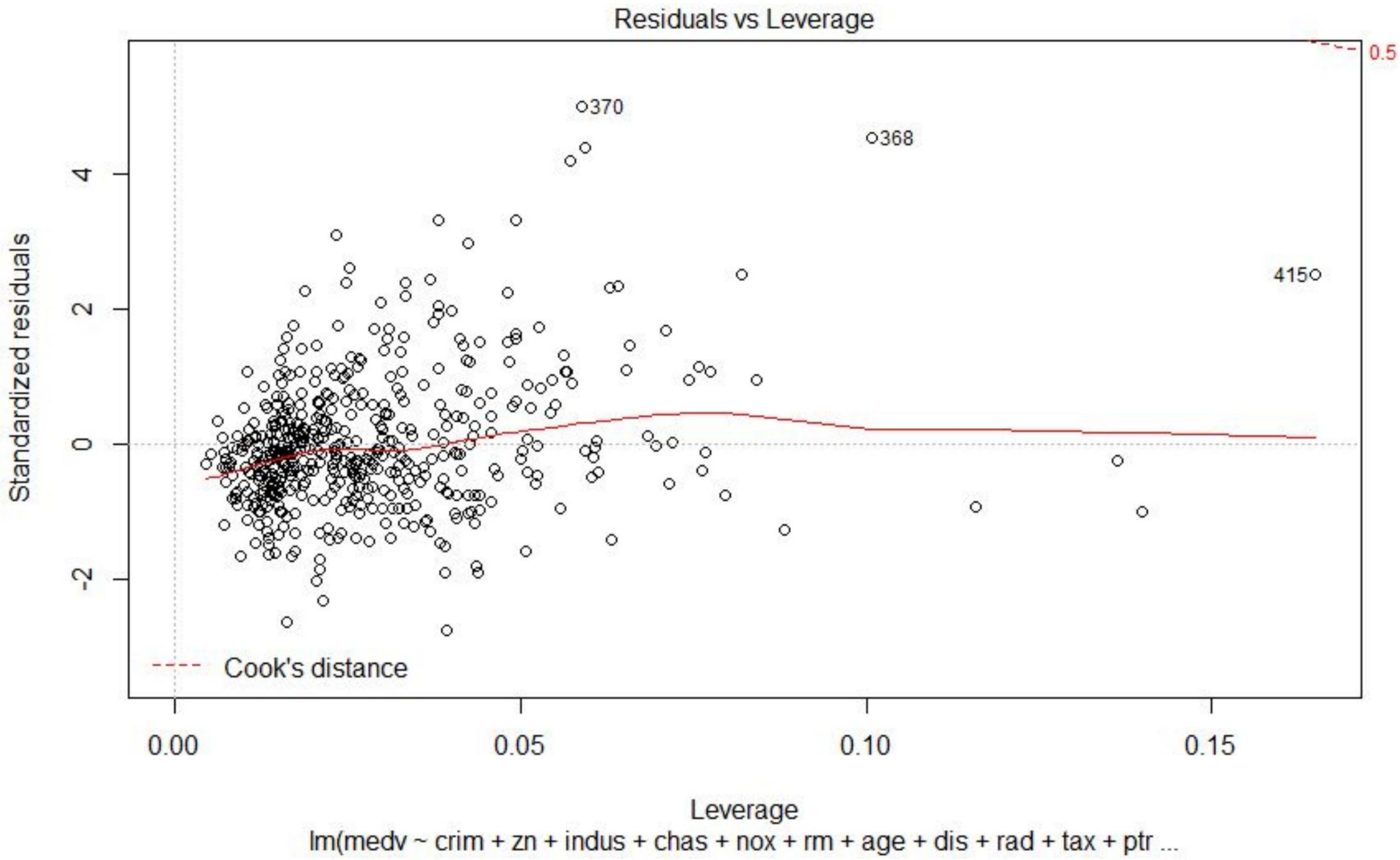
(**50 points**) **Page 2**: a screenshot of your diagnostic plot and a few sentences of your explanation.



The index 381, 419, 406, 411, 365, 369, 373, 372, and 366 have data that can be considered outliers.

- Index 381, 419, 406, 366, and 411 have high leverage (>0.1), which would make the prediction model depends heavily on these incorrect values.
- Index 365, 369, 373, and 372 have high standardized residual (>3), which mean they are very unlikely to be measured (less than 1% chance).

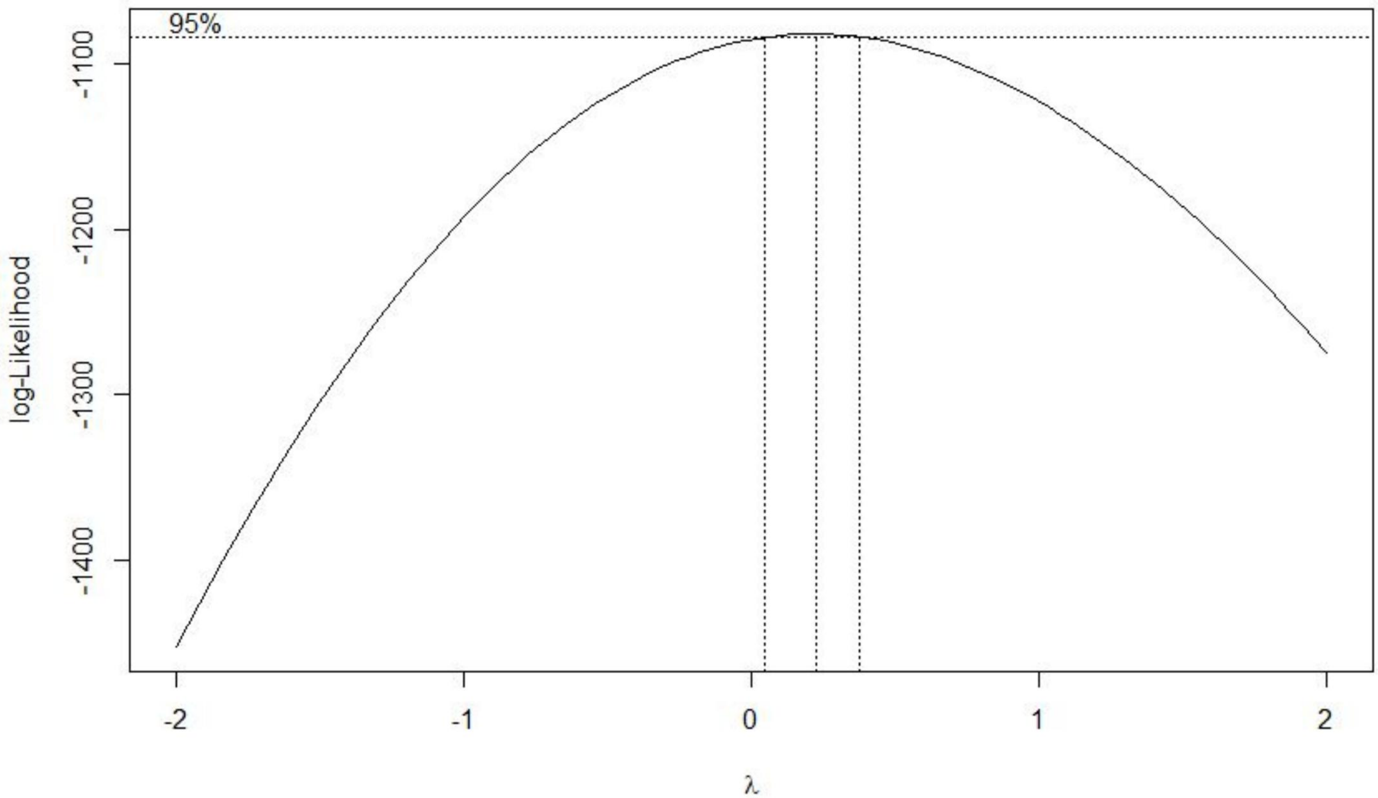
(**20 points**) **Page 3**: a screenshot of your new diagnostic plot.



(**10 points**) **Page 4**: a screenshot of your code for subproblem 2.

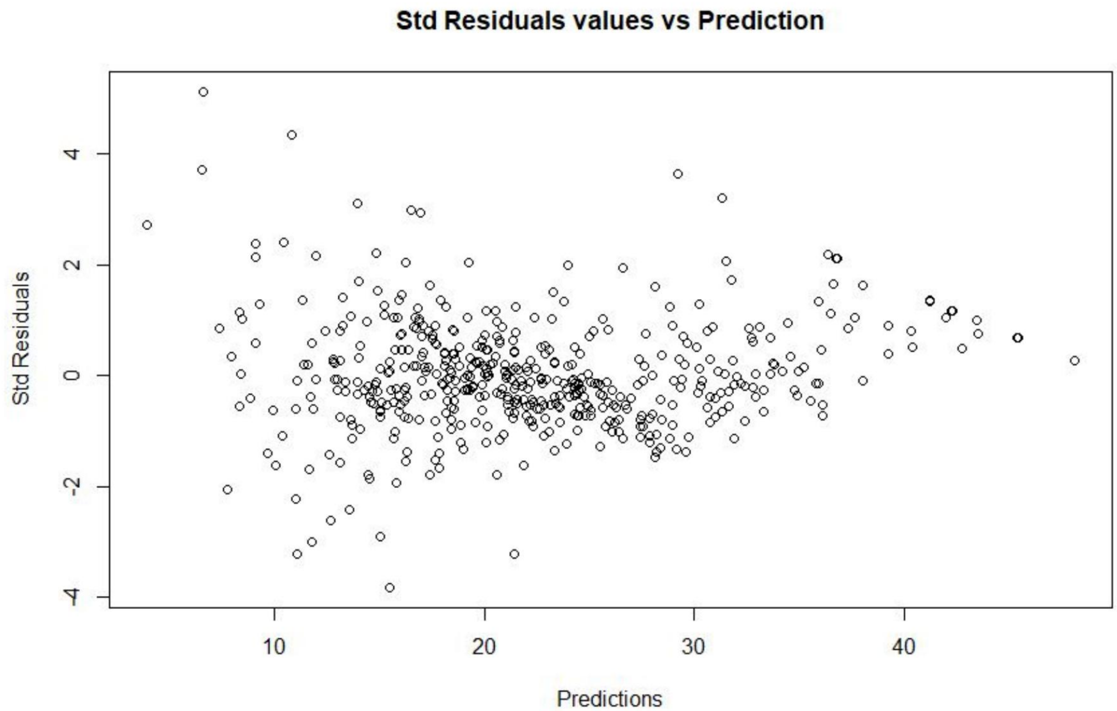
```
Boston2 <- Boston[-c(381,419,406,411,365,369,373,372,366),]
fit2 <- lm(medv ~ crim + zn + indus + chas + nox + rm + age + dis + rad + tax +
ptratio + black + lstat, data=Boston2)
par(mfrow=c(1,1))
plot(fit2, id.n = 3)
```

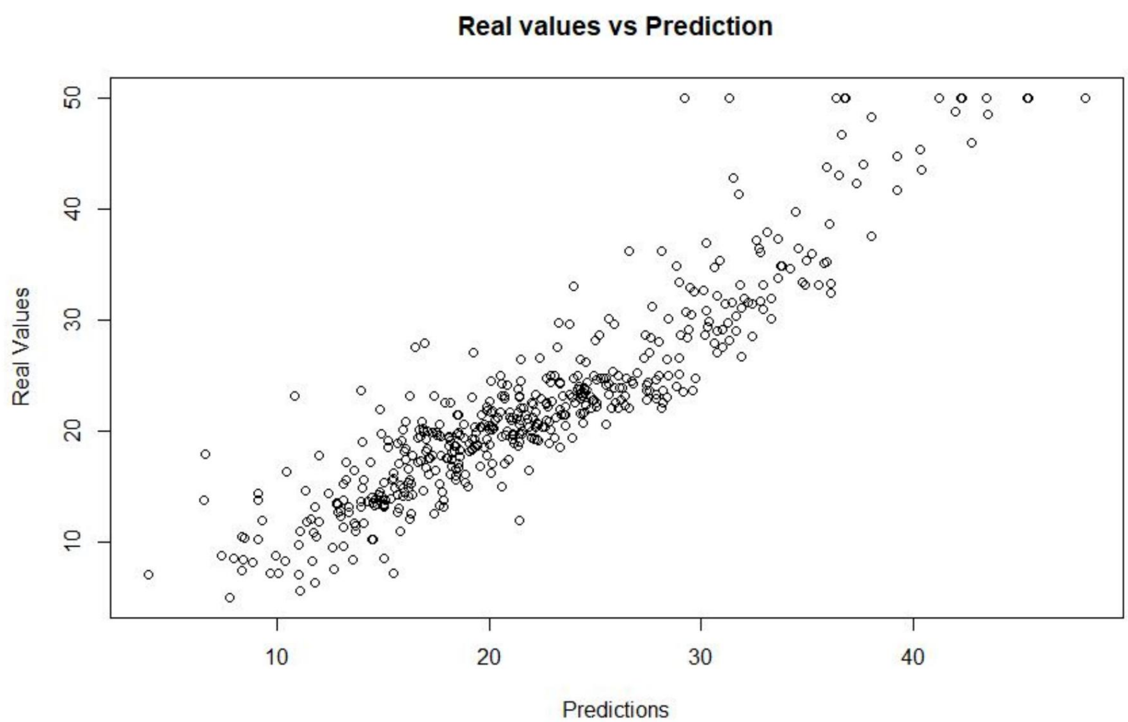
(**10 points**) **Page 5**: a screenshot of Box-Cox transformation plot and the best value you chose.



We chose 0.222222 for lambda.

(**10 points**) **Page 6**: result of the standardized residuals of the regression after Box-Cox transformation and a plot of fitted house price against true house price.





(**0 points**) **Page 7**: code for subproblems 3 and 4.

```
# Apply Box-Cox transformation
par(mfrow=c(1,1))
Box = boxcox(Boston2$medv ~ 1)
Cox = data.frame(Box$x, Box$y)            # Create a data frame with the results
Cox2 = Cox[with(Cox, order(-Cox$Box.y)),] # Order the new data frame by
decreasing y
Cox2[1,]                                  # Display the lambda with the greatest
log likelihood
lambda = Cox2[1, "Box.x"]                 # Extract that lambda
T_box = (Boston2$medv ^ lambda - 1)/lambda   # Transform the original data
Boston2$boxcox <- T_box

# Fit the Box-Cox transformed data
fit3 <- lm(boxcox ~ crim + zn + indus + chas + nox + rm + age + dis + rad + tax
+ ptratio + black + lstat, data=Boston2)
Boston2.stdres <- rstandard(fit3)
par(mfrow=c(1,1))

# Plot "Real values vs Prediction"
plot((((fit3$fitted.values)*0.2222222222)+1)^(1/0.2222222222), Boston2$medv,
ylab = "Real Values", xlab = "Predictions", main = "Real values vs Prediction")

# Plot "Std Residuals values vs Prediction"
plot((((fit3$fitted.values)*0.2222222222)+1)^(1/0.2222222222), Boston2.stdres,
ylab = "Std Residuals", xlab = "Predictions", main = "Std Residuals values vs
Prediction")
```