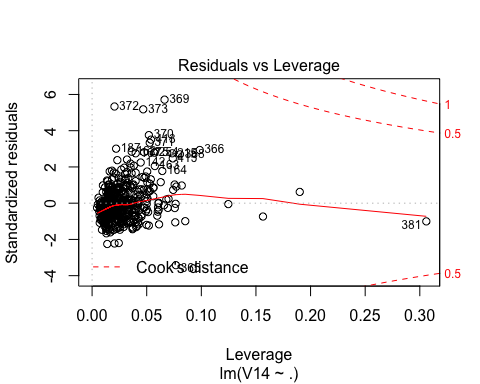
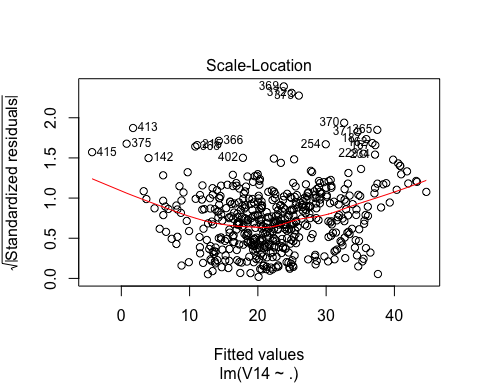
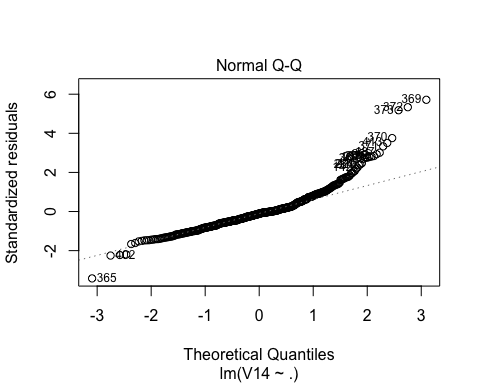
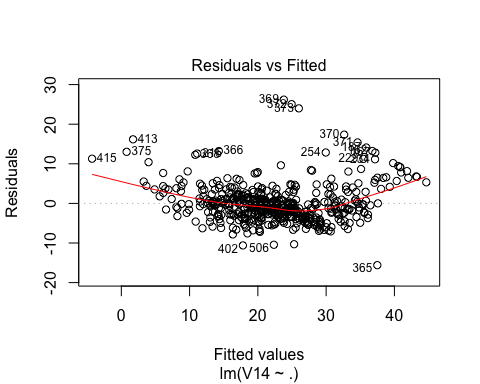
Assignment6

library(MASS)  
data = read.table("/Users/xiaoxin/Desktop/19spring/aml/6/housing.data.txt")  
model = lm(V14~., data = data)  
summary(model)

##   
## Call:  
## lm(formula = V14 ~ ., data = data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -15.595 -2.730 -0.518 1.777 26.199   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.646e+01 5.103e+00 7.144 3.28e-12 \*\*\*  
## V1 -1.080e-01 3.286e-02 -3.287 0.001087 \*\*   
## V2 4.642e-02 1.373e-02 3.382 0.000778 \*\*\*  
## V3 2.056e-02 6.150e-02 0.334 0.738288   
## V4 2.687e+00 8.616e-01 3.118 0.001925 \*\*   
## V5 -1.777e+01 3.820e+00 -4.651 4.25e-06 \*\*\*  
## V6 3.810e+00 4.179e-01 9.116 < 2e-16 \*\*\*  
## V7 6.922e-04 1.321e-02 0.052 0.958229   
## V8 -1.476e+00 1.995e-01 -7.398 6.01e-13 \*\*\*  
## V9 3.060e-01 6.635e-02 4.613 5.07e-06 \*\*\*  
## V10 -1.233e-02 3.760e-03 -3.280 0.001112 \*\*   
## V11 -9.527e-01 1.308e-01 -7.283 1.31e-12 \*\*\*  
## V12 9.312e-03 2.686e-03 3.467 0.000573 \*\*\*  
## V13 -5.248e-01 5.072e-02 -10.347 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 4.745 on 492 degrees of freedom  
## Multiple R-squared: 0.7406, Adjusted R-squared: 0.7338   
## F-statistic: 108.1 on 13 and 492 DF, p-value: < 2.2e-16

plot(model, id.n = 20)



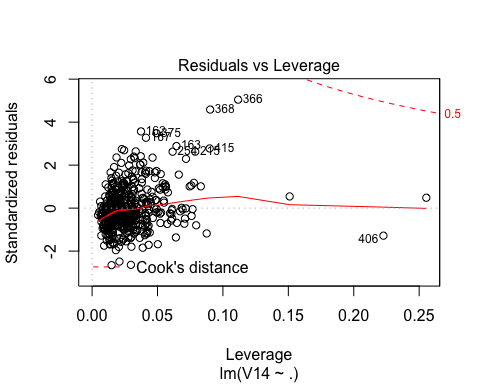
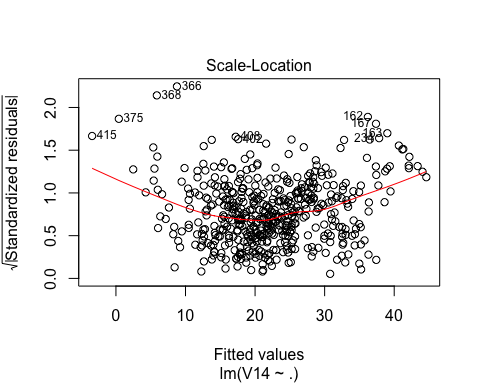
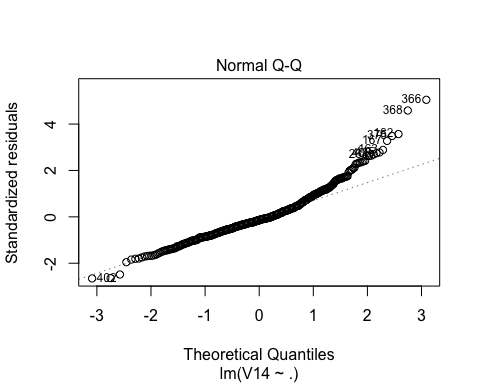
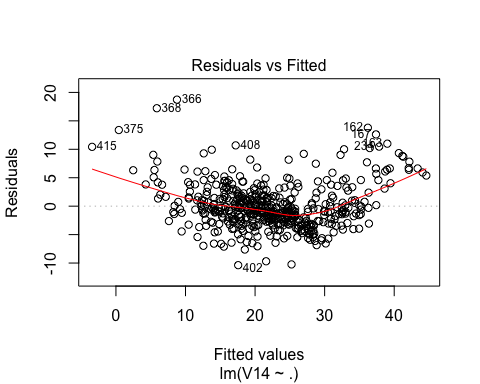
index = which(abs(stdres(model))>3)  
index

## 187 365 369 370 371 372 373 413   
## 187 365 369 370 371 372 373 413

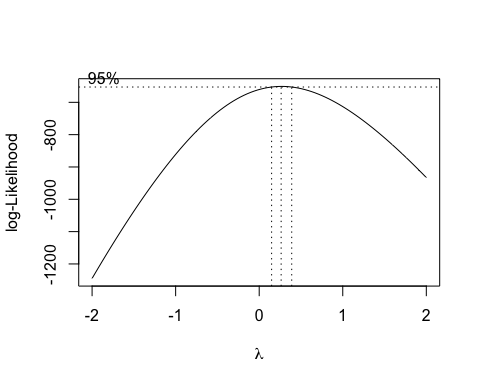
data\_fit = data[-c(187,365,369,370,371,372,373,413,381),]  
model\_fit = lm(V14~., data = data\_fit)  
summary(model\_fit)

##   
## Call:  
## lm(formula = V14 ~ ., data = data\_fit)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -10.371 -2.380 -0.564 1.659 18.729   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 24.224082 4.413242 5.489 6.54e-08 \*\*\*  
## V1 -0.072168 0.032483 -2.222 0.026767 \*   
## V2 0.034932 0.011446 3.052 0.002398 \*\*   
## V3 0.028941 0.051106 0.566 0.571457   
## V4 1.407081 0.755376 1.863 0.063104 .   
## V5 -11.980300 3.210987 -3.731 0.000213 \*\*\*  
## V6 5.004208 0.365298 13.699 < 2e-16 \*\*\*  
## V7 -0.020454 0.011099 -1.843 0.065952 .   
## V8 -1.196761 0.166796 -7.175 2.74e-12 \*\*\*  
## V9 0.212983 0.055962 3.806 0.000159 \*\*\*  
## V10 -0.012689 0.003128 -4.056 5.81e-05 \*\*\*  
## V11 -0.932464 0.109251 -8.535 < 2e-16 \*\*\*  
## V12 0.010473 0.002280 4.593 5.59e-06 \*\*\*  
## V13 -0.376034 0.044150 -8.517 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3.937 on 483 degrees of freedom  
## Multiple R-squared: 0.8028, Adjusted R-squared: 0.7975   
## F-statistic: 151.3 on 13 and 483 DF, p-value: < 2.2e-16

plot(model\_fit, id.n = 10)



trans = boxcox(model\_fit)



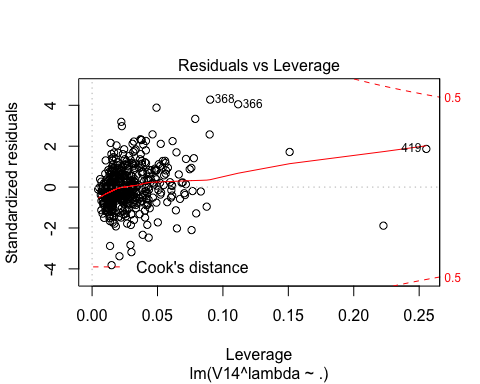
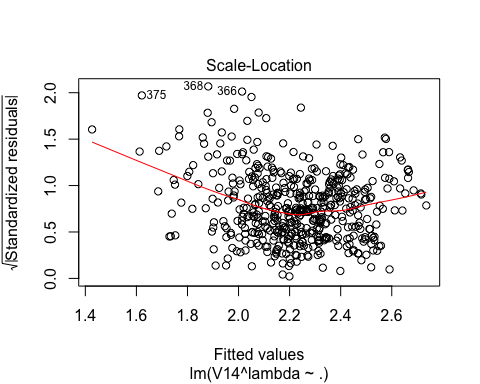
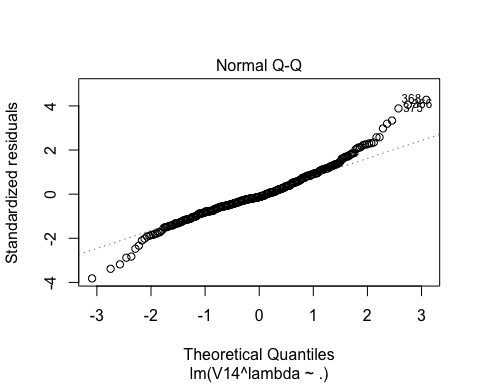
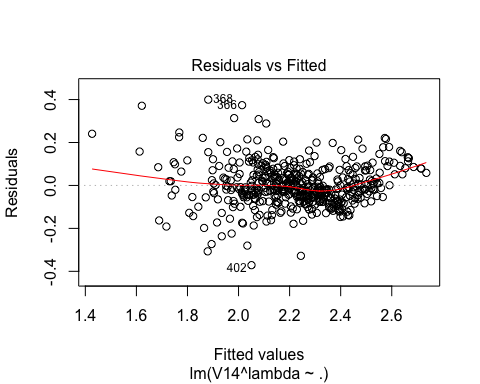
lambda = trans$x[which(trans$y == max(trans$y))]  
print(lambda)

## [1] 0.2626263

model\_trans = lm(V14^lambda~.,data = data\_fit)  
summary(model\_trans)

##   
## Call:  
## lm(formula = V14^lambda ~ ., data = data\_fit)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.37129 -0.05449 -0.01200 0.05200 0.39960   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.554e+00 1.098e-01 23.259 < 2e-16 \*\*\*  
## V1 -5.013e-03 8.082e-04 -6.202 1.20e-09 \*\*\*  
## V2 6.220e-04 2.848e-04 2.184 0.0294 \*   
## V3 1.254e-03 1.272e-03 0.986 0.3245   
## V4 3.714e-02 1.879e-02 1.976 0.0487 \*   
## V5 -3.399e-01 7.989e-02 -4.255 2.51e-05 \*\*\*  
## V6 8.770e-02 9.089e-03 9.650 < 2e-16 \*\*\*  
## V7 -3.294e-04 2.761e-04 -1.193 0.2335   
## V8 -2.626e-02 4.150e-03 -6.327 5.70e-10 \*\*\*  
## V9 6.718e-03 1.392e-03 4.825 1.88e-06 \*\*\*  
## V10 -3.616e-04 7.783e-05 -4.647 4.36e-06 \*\*\*  
## V11 -2.265e-02 2.718e-03 -8.332 8.28e-16 \*\*\*  
## V12 2.677e-04 5.674e-05 4.719 3.11e-06 \*\*\*  
## V13 -1.329e-02 1.098e-03 -12.103 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.09797 on 483 degrees of freedom  
## Multiple R-squared: 0.823, Adjusted R-squared: 0.8182   
## F-statistic: 172.7 on 13 and 483 DF, p-value: < 2.2e-16

plot(model\_trans)



y\_predict = predict(model\_trans, newdata = data\_fit[1:13])  
plot(data\_fit$V14,y\_predict, xlab = "true house price", ylab = "fitted house price",main = "fitted house price against true house price")



std\_residual = stdres(model\_trans)  
plot((data\_fit$V14)^lambda, std\_residual, ylab = "standardized residuals", xlab = "fitted value", main = "Standardized residuals against fitted values")

