Individual Assignment 2

Charlie Ling

9/11/2021

10. This question should be answered using the Carseats data set.

```
library(ISLR)
```

```
## Warning: package 'ISLR' was built under R version 4.0.5
```

```
attach(Carseats)
View(Carseats)
```

(a) Fit a multiple regression model to predict Sales using Price, Urban, and US.

```
lm.fit=lm(Sales~Price+Urban+US)
summary(lm.fit)
```

```
##
## Call:
## lm(formula = Sales ~ Price + Urban + US)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.9206 -1.6220 -0.0564  1.5786  7.0581
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 13.043469   0.651012  20.036  < 2e-16 ***
## Price       -0.054459   0.005242 -10.389  < 2e-16 ***
## UrbanYes    -0.021916   0.271650  -0.081    0.936
## USYes        1.200573   0.259042   4.635 4.86e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.472 on 396 degrees of freedom
## Multiple R-squared:  0.2393, Adjusted R-squared:  0.2335
## F-statistic: 41.52 on 3 and 396 DF,  p-value: < 2.2e-16
```

(b) Provide an interpretation of each coefficient in the model. Be careful—some of the variables in the model are qualitative!

```
# Sales is negetively related to price.
# Us stores have higher sales
```

(c) Write out the model in equation form, being careful to handle the qualitative variables properly.

```
# Sales=a+b1*Price+b2*Urban+b3*US
```

(d) For which of the predictors can you reject the null hypothesis H0:Bj=0?

```
# For Price and US I can reject the null hypothesis H0:Bj=0
```

(e) On the basis of your response to the previous question, fit a smaller model that only uses the predictors for which there is evidence of association with the outcome.

```
lm.fit1=lm(Sales~Price+US)
summary(lm.fit1)
```

```
##
## Call:
## lm(formula = Sales ~ Price + US)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.9269 -1.6286 -0.0574  1.5766  7.0515
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 13.03079    0.63098  20.652  < 2e-16 ***
## Price       -0.05448    0.00523 -10.416  < 2e-16 ***
## USYes        1.19964    0.25846   4.641 4.71e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.469 on 397 degrees of freedom
## Multiple R-squared:  0.2393, Adjusted R-squared:  0.2354
## F-statistic: 62.43 on 2 and 397 DF,  p-value: < 2.2e-16
```

(f) How well do the models in (a) and (e) fit the data?

```
# The models in (a) and (e) have similar Adjusted R-squared: about 23%, which
# means these two model fit the data not well.
```

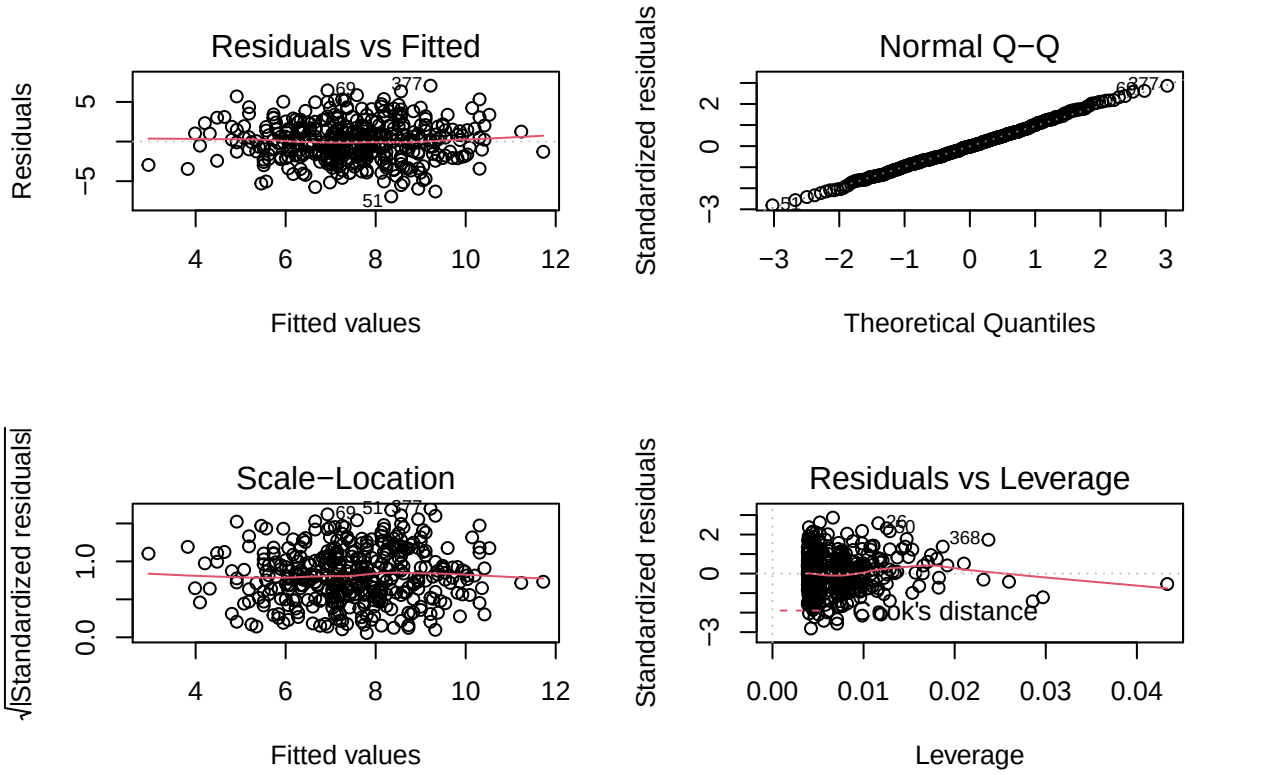
(g) Using the model from (e), obtain 95% confidence intervals for the coefficient(s).

```
confint(lm.fit1)
```

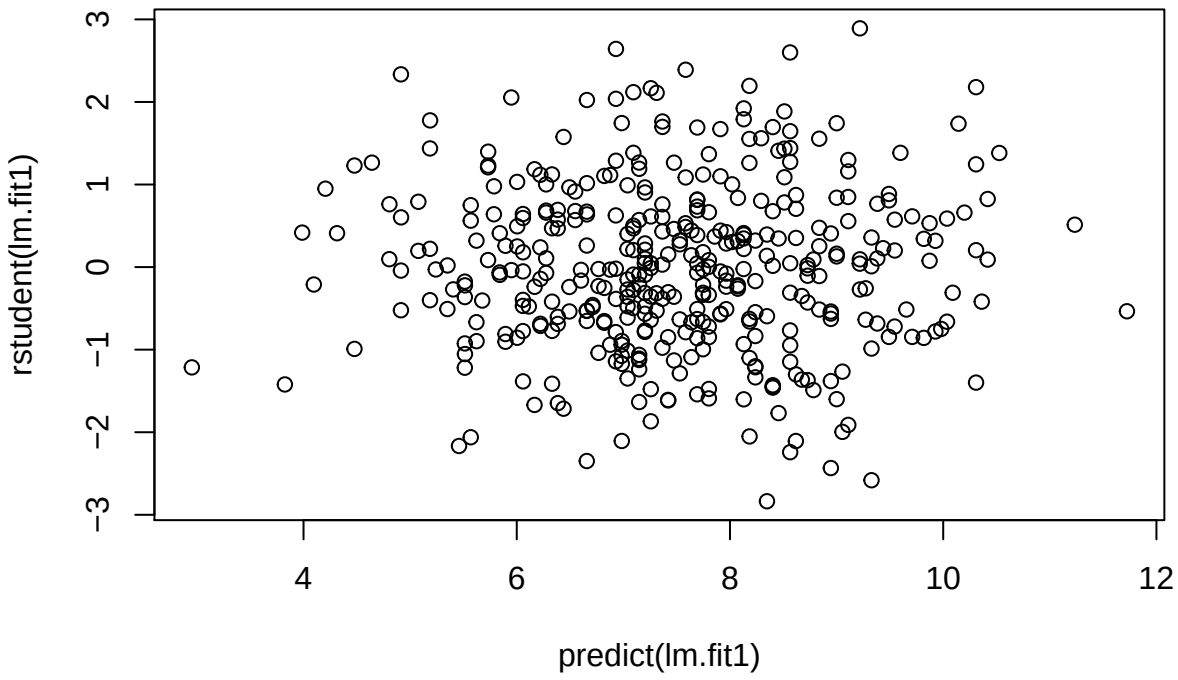
```
##                   2.5 %      97.5 %
## (Intercept) 11.79032020 14.27126531
## Price       -0.06475984 -0.04419543
## USYes        0.69151957  1.70776632
```

(h) Is there evidence of outliers or high leverage observations in the model from (e)?

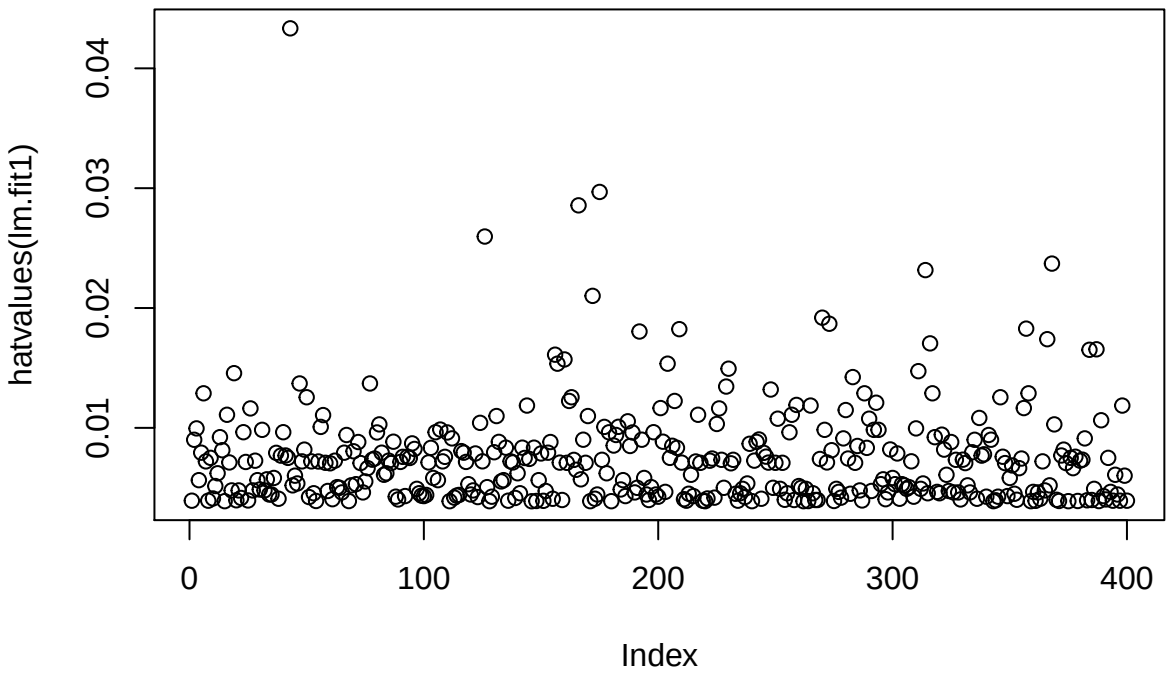
```
par(mfrow=c(2,2)) ##divide plotting region into 2*2

plot(lm.fit1)
```




```
par(mfrow=c(1,1)) ##reset plotting region
plot(predict(lm.fit1),rstudent(lm.fit1))
```




```
plot(hatvalues(lm.fit1)) ##Leverage
```




```
# There are dots are on the far right side of the Residuals vs Leverage
# plot, which means potential high-leverage points
```