Zixiao Wu 515491-Individual Assignment 2

Zixiao Wu

2022-09-16

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com (http://rmarkdown.rstudio.com).

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
library(MASS)
library(ISLR)
Carseats = na.omit(Carseats)
head(Carseats)
```

```
##   Sales CompPrice Income Advertising Population Price ShelveLoc Age Education
## 1  9.50       138     73           11        276   120       Bad  42        17
## 2 11.22       111     48           16        260    83      Good  65        10
## 3 10.06       113     35           10        269    80    Medium  59        12
## 4  7.40       117    100            4        466    97    Medium  55        14
## 5  4.15       141     64            3        340   128       Bad  38        13
## 6 10.81       124    113           13        501    72       Bad  78        16
##   Urban  US
## 1   Yes Yes
## 2   Yes Yes
## 3   Yes Yes
## 4   Yes Yes
## 5   Yes  No
## 6    No Yes
```

```
attach(Carseats)
```

#(a)

```
sales = lm(Sales ~ Price + Urban + US, data=Carseats)
summary(sales)
```

```
##
## Call:
## lm(formula = Sales ~ Price + Urban + US, data = Carseats)
##
## 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)

1. the intercept represents the number of car seats sold when all other predictors = 0;
2. the coefficient of price = -0.05, means when price increases by 1 and other predictors are constant, sales decrease by 54.
3. the coefficient of urban = -0.02, but it is not significant, means it does not affect the sales.
4. the coefficient ofus = 1.2, means when the sales is in the us, the sales will increase by 1200.

#(c)

```
Sales = 13.04 - 0.02 * Urban + 1.20 * US - 0.05 * Price
      = 13.04 - 0.02 * Urban + 1.20 * US - 0.05 * Price,                    (Urban=1, US=1)
      = 13.04 - 0.02 * Urban + 1.20 * Price,                                (Urban=1, US=0)
      = 13.04 + 1.20 * US - 0.05 * Price,                                   (Urban=0, US=1)
      = 13.04 - 0.05 * Price,                                               (Urban=0, US=0)
```

#(d)

we can reject the H0 hypothesis for the intercept, US and Price according to the p-value in the summary of the model above.

#(e)

```
sales2 = lm(Sales ~ Price+US, data= Carseats)
summary(sales2)
```

```
##
## Call:
## lm(formula = Sales ~ Price + US, data = Carseats)
##
## 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)

```
The Residual standard error goes from 2.47 to 2.46, remian the same, and the Adjusted R-squared remian in 0.23.
So neither model is very well, given the low value of and R-squared.
```

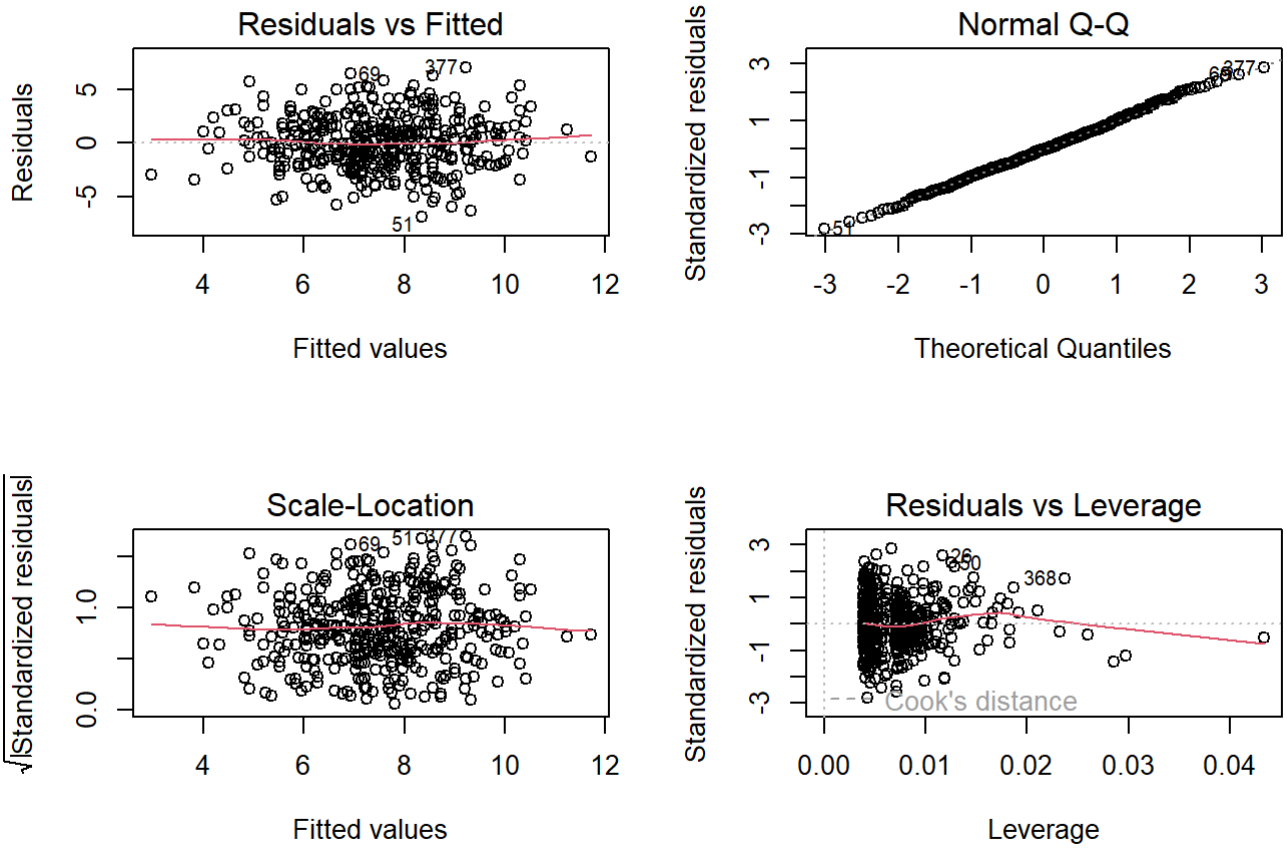
#(g)

```
confint(sales2)
```

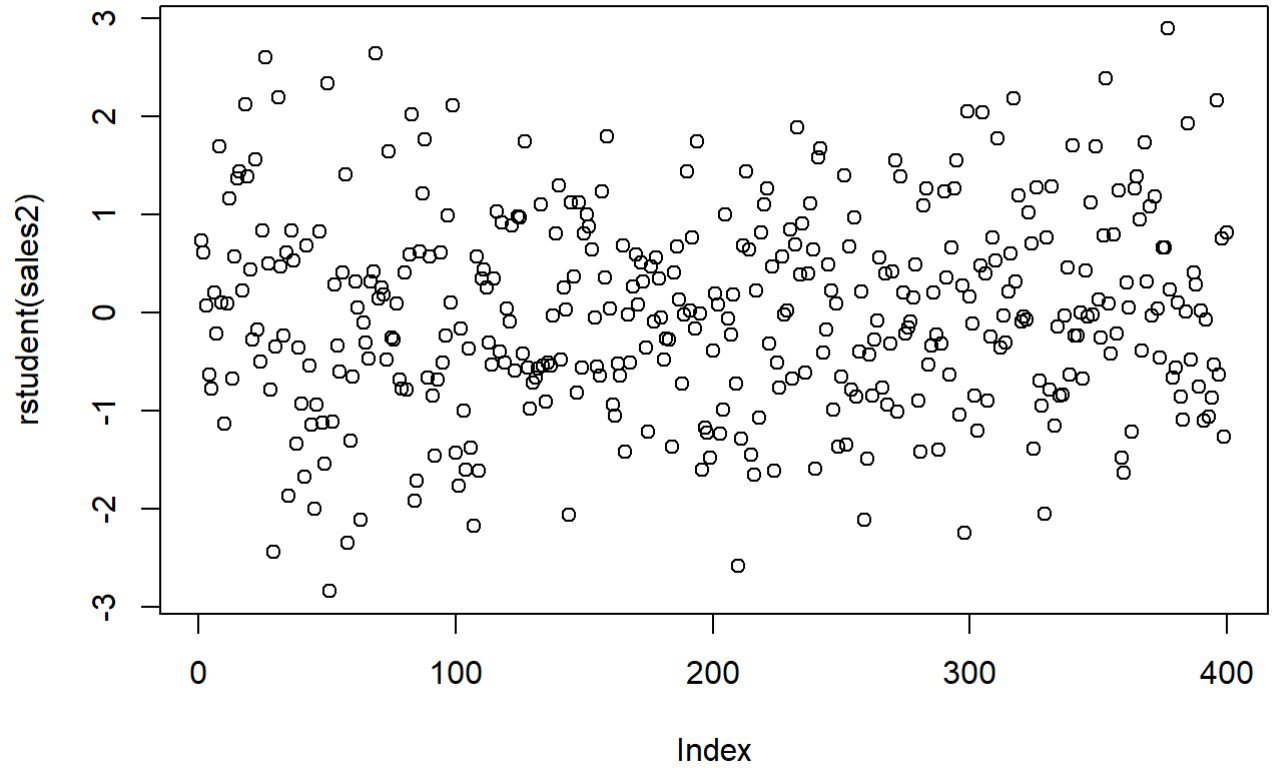
```
##                   2.5 %      97.5 %
## (Intercept) 11.79032020 14.27126531
## Price       -0.06475984 -0.04419543
## USYes        0.69151957  1.70776632
```

#(h)

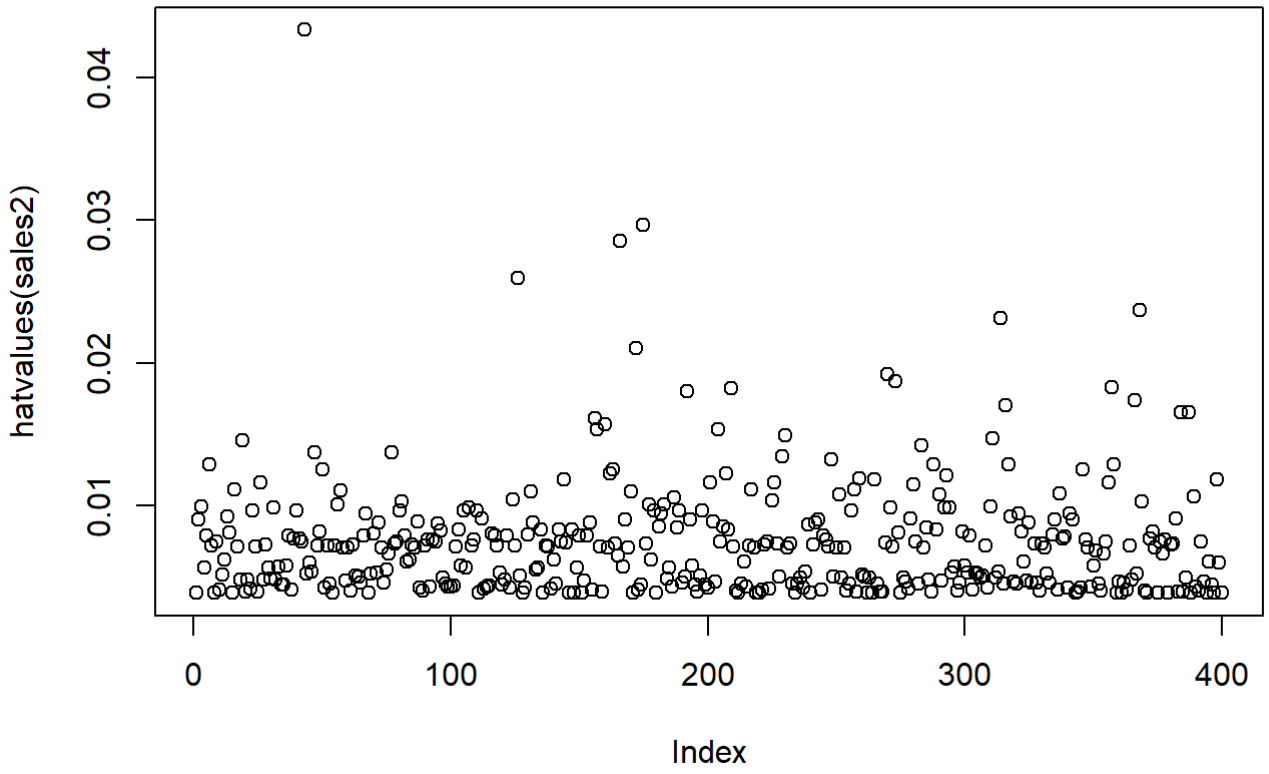
```
par(mfrow=c(2,2))
plot(sales2)
```




```
par(mfrow=c(1,1))
plot(rstudent(sales2))
```



```
plot(hatvalues(sales2))
```




```
rstudent(sales2)[which(rstudent(sales2)>3)]
```

```
## named numeric(0)
```

Based on the Residuals vs Leverage plot and Hatvalues plot, we can see there are no evidence of high leverage observation.