[STAT 4610] HW-1 / Michael Ghattas

Michael Ghattas

9/6/2022

Problem 10

Start:

```
library("ISLR")
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
```

Part (a)

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

```
##
## 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
```

Part (b)

- If the price increases by USD 1000 while other predictors held constant, sales would decrease by ~54.46 units.
- A store location in relation to Urban areas has no affect on sales.
- US based stores will on average sell 1200 more carseats than international stores.

Part (c)

- $Sales = \beta_0 + \beta_1 \cdot Price + \beta_2 \cdot UrbanYes + \beta_3 \cdot USYes$

Part (d)

- Urban: Its p-value is not statistically significant (p-value = 0.936).

Part (e)

```
lm.fit2 = lm(Sales ~ Price + US, data = Carseats)
summary(lm.fit2)
```

```
##
## 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
```

Part (f)

- Given the R^2 values we can note that the two models are weak, with only ~23.5% of the data explained.

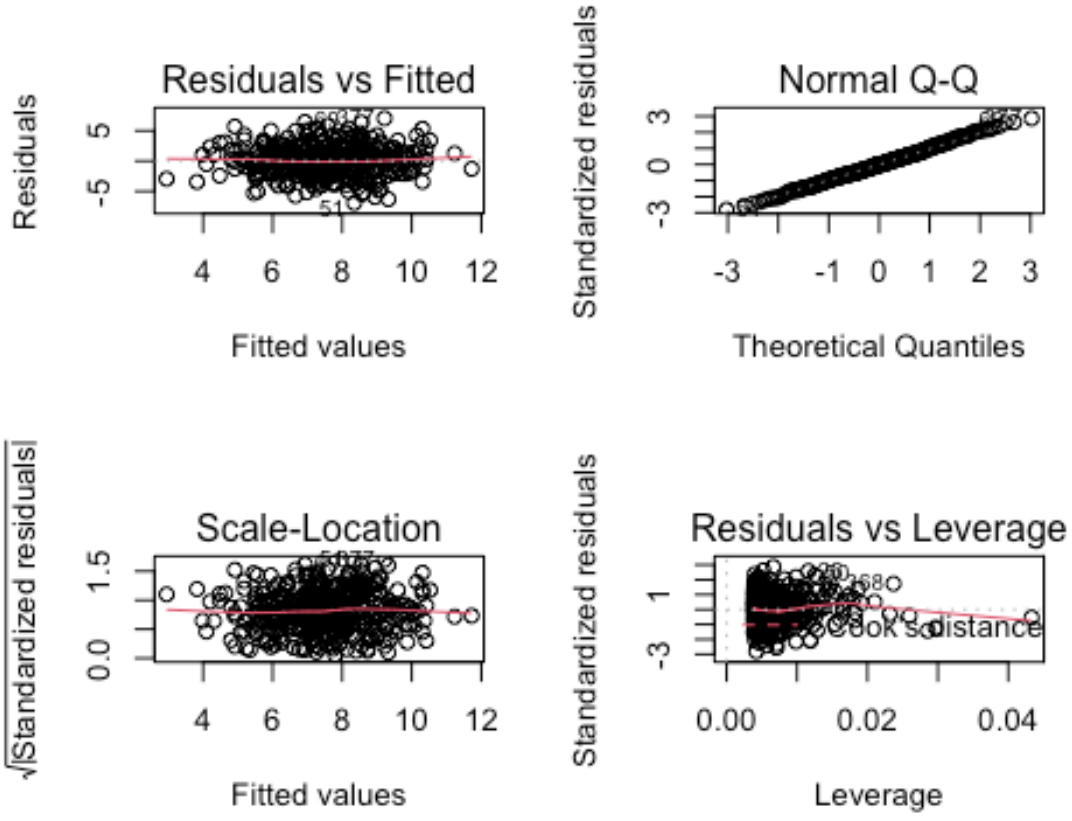
Part (g)

```
confint(lm.fit2)
```

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

Part (h)

```
par(mfrow = c(2, 2))
plot(lm.fit2)
```



- No treatment is needed, given the Normal Q-Q plot and the Residuals vs Leverage plot, we can note that outliers/high-leverage points are not significantly visible.

End.