Anusha muppalla
16286311

ISL LAB ASSIGNMENT-2

2. This question involves the use of simple linear regression on the Auto data set.
(a) Use the lm() function to perform a simple linear regression with mpg as the response and horsepower as the predictor. Use the summary() function to print the results. Comment on the output. For example:

i. Is there a relationship between the predictor and the response?

```
> library(ISLR)
> data("Auto")
>
> lm.fit <- lm(mpg ~ horsepower, data=Auto)
> summary(lm.fit)

Call:
lm(formula = mpg ~ horsepower, data = Auto)

Residuals:
     Min       1Q   Median       3Q      Max
-13.5710  -3.2592  -0.3435   2.7630  16.9240

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 39.935861   0.717499   55.66   <2e-16 ***
horsepower  -0.157845   0.006446  -24.49   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.906 on 390 degrees of freedom
Multiple R-squared:  0.6059,    Adjusted R-squared:  0.6049
F-statistic: 599.7 on 1 and 390 DF,  p-value: < 2.2e-16

>
```

ii. How strong is the relationship between the predictor and the response ?

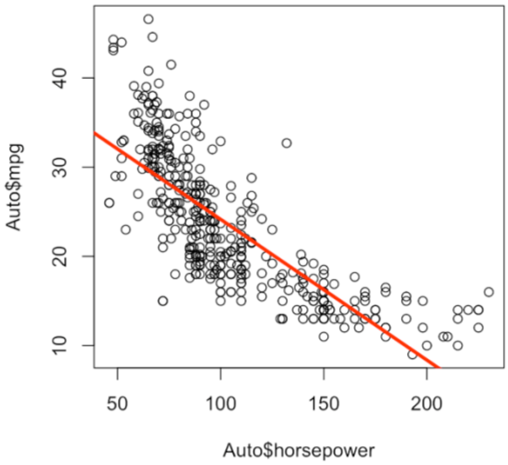
To calculate the residual error relative to the response we use the mean of the response and the RSE. The mean of mpg is 23.4459184. The RSE of the lm.fit was 4.9057569 which indicates a percentage error of 20.9237141%. We may also note that as the $R2$ is equal to 0.6059483, almost 60.5948258% of the variability in “mpg” can be explained using “horsepower”.

iii. Is the relationship between the predictor and the response positive or negative?

As the coeficient of “horsepower” is negative, the relationship is also negative. The more horsepower an automobile has the linear regression indicates the less mpg fuel efficiency the automobile will have.

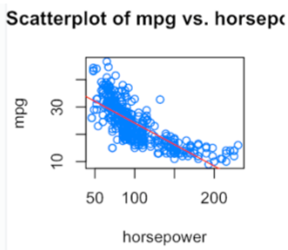
iv. What is the predicted mpg associated with a horsepower of 98? What are the associated 95% confidence and prediction intervals?

```
Console

Residuals:
     Min       1Q   Median       3Q      Max
-13.5710  -3.2592  -0.3435   2.7630  16.9240

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 39.935861   0.717499   55.66   <2e-16 ***
horsepower  -0.157845   0.006446  -24.49   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.906 on 390 degrees of freedom
Multiple R-squared:  0.6059,	Adjusted R-squared:  0.6049
F-statistic: 599.7 on 1 and 390 DF,  p-value: < 2.2e-16

> predict(lm.fit, data.frame("horsepower"=98), interval="confidence")
       fit      lwr      upr
1 24.46708 23.97308 24.96108
> predict(lm.fit, data.frame("horsepower"=98), interval="prediction")
       fit     lwr      upr
1 24.46708 14.8094 34.12476
> plot(Auto$horsepower, Auto$mpg)
> abline(lm.fit, lwd=3, col="red")
>
```



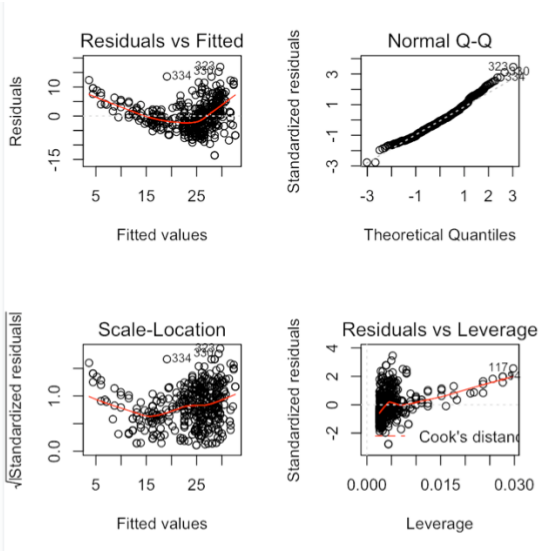
2.(b) Plot the response and the predictor. Use the abline() function to display the least squares regression line.

```
Residuals:
     Min       1Q   Median       3Q      Max
-13.5710  -3.2592  -0.3435   2.7630  16.9240

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 39.935861   0.717499   55.66   <2e-16 ***
horsepower  -0.157845   0.006446  -24.49   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.906 on 390 degrees of freedom
Multiple R-squared:  0.6059,	Adjusted R-squared:  0.6049
F-statistic: 599.7 on 1 and 390 DF,  p-value: < 2.2e-16

> predict(fit, data.frame(horsepower = 98), interval = "confidence")
       fit      lwr      upr
1 24.46708 23.97308 24.96108
> predict(fit, data.frame(horsepower = 98), interval = "prediction")
       fit     lwr      upr
1 24.46708 14.8094 34.12476
> plot(Auto$horsepower, Auto$mpg, main = "Scatterplot of mpg vs. horsep
ower", xlab = "horsepower", ylab = "mpg", col = "blue")
> abline(fit, col = "red")
>
```



2.(c) Use the plot() function to produce diagnostic plots of the least squares regression fit. Comment on any problems you see with the fit.

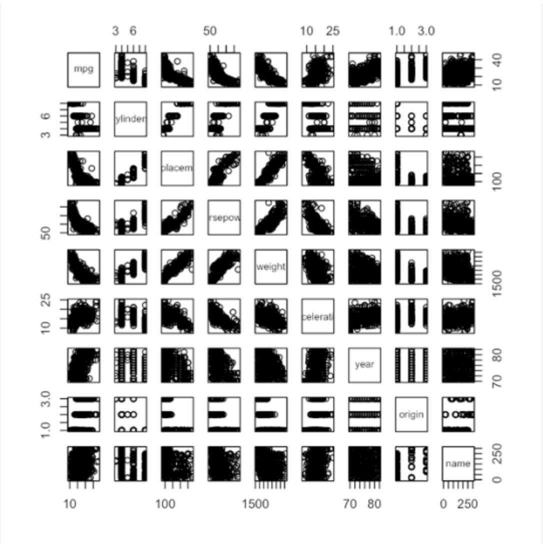
```
-13.5710  -3.2592  -0.3435   2.7630  16.9240

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 39.935861   0.717499   55.66   <2e-16 ***
horsepower  -0.157845   0.006446  -24.49   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.906 on 390 degrees of freedom
Multiple R-squared:  0.6059,	Adjusted R-squared:  0.6049
F-statistic: 599.7 on 1 and 390 DF,  p-value: < 2.2e-16

> predict(fit, data.frame(horsepower = 98), interval = "confidence")
       fit      lwr      upr
1 24.46708 23.97308 24.96108
> predict(fit, data.frame(horsepower = 98), interval = "prediction")
       fit     lwr      upr
1 24.46708 14.8094 34.12476
> plot(Auto$horsepower, Auto$mpg, main = "Scatterplot of mpg vs. horsepower", xlab = "horsepower", ylab = "mpg", col = "blue")
> abline(fit, col = "red")
> par(mfrow = c(2, 2))
> plot(fit)
>
>
```



3.This question involves the use of multiple linear regression on the Auto data set.

(a) Produce a scatterplot matrix which includes all of the variables in the data set.

```
Console ~/
> library(ISLR)
> data("Auto")
>
> pairs(Auto)
>
```



(b) Compute the matrix of correlations between the variables using the function cor(). You will need to exclude the name variable, which is qualitative.

```
>
> cor(Auto[1:8])
                    mpg  cylinders displacement horsepower
mpg           1.0000000 -0.7776175   -0.8051269 -0.7784268
cylinders    -0.7776175  1.0000000    0.9508233  0.8429834
displacement -0.8051269  0.9508233    1.0000000  0.8972570
horsepower   -0.7784268  0.8429834    0.8972570  1.0000000
weight       -0.8322442  0.8975273    0.9329944  0.8645377
acceleration  0.4233285 -0.5046834   -0.5438005 -0.6891955
year          0.5805410 -0.3456474   -0.3698552 -0.4163615
origin        0.5652088 -0.5689316   -0.6145351 -0.4551715
                 weight acceleration       year     origin
mpg          -0.8322442    0.4233285  0.5805410  0.5652088
cylinders     0.8975273   -0.5046834 -0.3456474 -0.5689316
displacement  0.9329944   -0.5438005 -0.3698552 -0.6145351
horsepower    0.8645377   -0.6891955 -0.4163615 -0.4551715
weight        1.0000000   -0.4168392 -0.3091199 -0.5850054
acceleration -0.4168392    1.0000000  0.2903161  0.2127458
year         -0.3091199    0.2903161  1.0000000  0.1815277
origin       -0.5850054    0.2127458  0.1815277  1.0000000
> |
```

(c) Use the lm() function to perform a multiple linear regression with mpg as the response and all other variables except name as the predictors. Use the summary() function to print the results. Comment on the output. For instance:

```
Console ~/
summary"
> autolm2 <- lm(mpg ~ . -name, data = Auto)
> summary(autolm2)

Call:
lm(formula = mpg ~ . - name, data = Auto)

Residuals:
    Min      1Q  Median      3Q     Max
-9.5903 -2.1565 -0.1169  1.8690 13.0604

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)  -17.218435   4.644294  -3.707  0.00024 ***
cylinders     -0.493376   0.323282  -1.526  0.12780
displacement   0.019896   0.007515   2.647  0.00844 **
horsepower    -0.016951   0.013787  -1.230  0.21963
weight        -0.006474   0.000652  -9.929  < 2e-16 ***
acceleration   0.080576   0.098845   0.815  0.41548
year           0.750773   0.050973  14.729  < 2e-16 ***
origin         1.426141   0.278136   5.127 4.67e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.328 on 384 degrees of freedom
Multiple R-squared:  0.8215,    Adjusted R-squared:  0.8182
F-statistic: 252.4 on 7 and 384 DF,  p-value: < 2.2e-16

>
```

1.Is there a relationship between the predictors and the response?

The F-statistic is very high which indicates that there is most likely a strong relationship between the predictors and the response.

2.Which predictors appear to have a statistically significant relationship to the response?

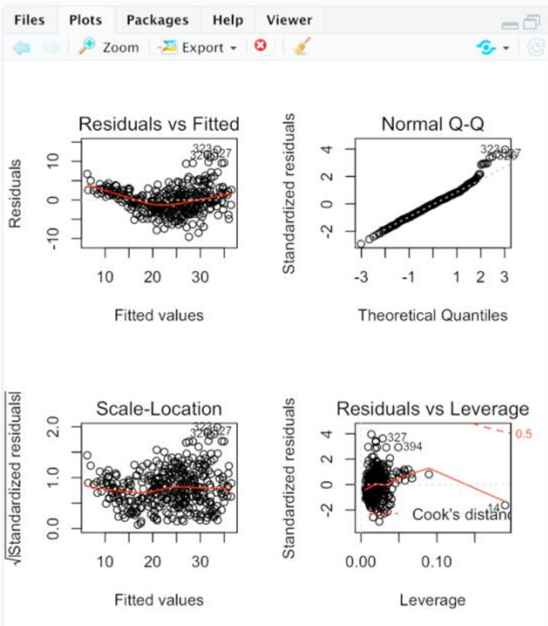
The low p-values for displacement, weight, year, and origin indicate a statistically significant relationship to mpg.

3. What does the coefficient for the year variable suggest?

Each additional year improves fuel efficiency by approximately 0.75 mpg.

(d) Use the plot() function to produce diagnostic plots of the linear regression fit. Comment on any problems you see with the fit. Do the residual plots suggest any unusually large outliers? Does the leverage plot identify any observations with unusually high leverage?

```
Console ~/
> summary(autolm2)

Call:
lm(formula = mpg ~ . - name, data = Auto)

Residuals:
    Min      1Q  Median      3Q     Max
-9.5903 -2.1565 -0.1169  1.8690 13.0604

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)  -17.218435   4.644294  -3.707  0.00024 ***
cylinders     -0.493376   0.323282  -1.526  0.12780
displacement   0.019896   0.007515   2.647  0.00844 **
horsepower    -0.016951   0.013787  -1.230  0.21963
weight        -0.006474   0.000652  -9.929  < 2e-16 ***
acceleration   0.080576   0.098845   0.815  0.41548
year           0.750773   0.050973  14.729  < 2e-16 ***
origin         1.426141   0.278136   5.127 4.67e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.328 on 384 degrees of freedom
Multiple R-squared:  0.8215,	Adjusted R-squared:  0.8182
F-statistic: 252.4 on 7 and 384 DF,  p-value: < 2.2e-16

> par(mfrow=c(2,2))
> plot(autolm2)
>
```



(e) Use the ∗ and : symbols to fit linear regression models with interaction effects. Do any interactions appear to be statistically significant?

```
> autolm3 <- lm(mpg ~ . * ., data = Auto[,1:8])
> summary(autolm3)

Call:
lm(formula = mpg ~ . * ., data = Auto[, 1:8])

Residuals:
    Min      1Q  Median      3Q     Max
-7.6303 -1.4481  0.0596  1.2739 11.1386

Coefficients:
                            Estimate Std. Error t value Pr(>|t|)
(Intercept)                3.548e+01  5.314e+01   0.668  0.50475
cylinders                  6.989e+00  8.248e+00   0.847  0.39738
displacement              -4.785e-01  1.894e-01  -2.527  0.01192
horsepower                 5.034e-01  3.470e-01   1.451  0.14769
weight                     4.133e-03  1.759e-02   0.235  0.81442
acceleration              -5.859e+00  2.174e+00  -2.696  0.00735
year                       6.974e-01  6.097e-01   1.144  0.25340
origin                    -2.090e+01  7.097e+00  -2.944  0.00345
cylinders:displacement    -3.383e-03  6.455e-03  -0.524  0.60051
cylinders:horsepower       1.161e-02  2.420e-02   0.480  0.63157
cylinders:weight           3.575e-04  8.955e-04   0.399  0.69000
cylinders:acceleration     2.779e-01  1.664e-01   1.670  0.09584
cylinders:year            -1.741e-01  9.714e-02  -1.793  0.07389
cylinders:origin           4.022e-01  4.926e-01   0.816  0.41482
displacement:horsepower   -8.491e-05  2.885e-04  -0.294  0.76867
displacement:weight        2.472e-05  1.470e-05   1.682  0.09342
displacement:acceleration -3.479e-03  3.342e-03  -1.041  0.29853
displacement:year          5.934e-03  2.391e-03   2.482  0.01352
displacement:origin        2.398e-02  1.947e-02   1.232  0.21875
```

```
displacement:origin          2.398e-02  1.947e-02   1.232  0.21875
horsepower:weight           -1.968e-05  2.924e-05  -0.673  0.50124
horsepower:acceleration     -7.213e-03  3.719e-03  -1.939  0.05325
horsepower:year             -5.838e-03  3.938e-03  -1.482  0.13916
horsepower:origin            2.233e-03  2.930e-02   0.076  0.93931
weight:acceleration          2.346e-04  2.289e-04   1.025  0.30596
weight:year                 -2.245e-04  2.127e-04  -1.056  0.29182
weight:origin               -5.789e-04  1.591e-03  -0.364  0.71623
acceleration:year            5.562e-02  2.558e-02   2.174  0.03033
acceleration:origin          4.583e-01  1.567e-01   2.926  0.00365
year:origin                  1.393e-01  7.399e-02   1.882  0.06062

(Intercept)
cylinders
displacement                *
horsepower
weight
acceleration                **
year
origin                      **
cylinders:displacement
cylinders:horsepower
cylinders:weight
cylinders:acceleration      .
cylinders:year              .
cylinders:origin
displacement:horsepower
displacement:weight         .
displacement:acceleration
displacement:year           *
displacement:origin
```

```
year
origin                      **
cylinders:displacement
cylinders:horsepower
cylinders:weight
cylinders:acceleration      .
cylinders:year              .
cylinders:origin
displacement:horsepower
displacement:weight         .
displacement:acceleration
displacement:year           *
displacement:origin
horsepower:weight
horsepower:acceleration     .
horsepower:year
horsepower:origin
weight:acceleration
weight:year
weight:origin
acceleration:year           *
acceleration:origin         **
year:origin                 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.695 on 363 degrees of freedom
Multiple R-squared:  0.8893,    Adjusted R-squared:  0.8808
F-statistic: 104.2 on 28 and 363 DF,  p-value: < 2.2e-16

>
```

(f) Try a few different transformations of the variables, such as log(X), X, X . Comment on your findings.

```
> autolmlog <- lm(mpg ~ log(horsepower) + log(weight) + log(accelerati
n), data = Auto)
> summary(autolmlog)

Call:
lm(formula = mpg ~ log(horsepower) + log(weight) + log(acceleration),
    data = Auto)

Residuals:
     Min       1Q   Median       3Q      Max
-10.8237  -2.5240  -0.2389   2.0105  15.3681

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)        190.152      8.255  23.035  < 2e-16 ***
log(horsepower)    -11.799      1.933  -6.103 2.53e-09 ***
log(weight)        -12.306      1.820  -6.762 5.03e-11 ***
log(acceleration)   -5.363      1.970  -2.723  0.00677 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.961 on 388 degrees of freedom
Multiple R-squared:  0.7445,    Adjusted R-squared:  0.7425
F-statistic: 376.8 on 3 and 388 DF,  p-value: < 2.2e-16

>
```

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

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

```
sqrt(acceleration)  -1.6296     1.0218  -1.595    0.112
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.088 on 388 degrees of freedom
Multiple R-squared:  0.7278,    Adjusted R-squared:  0.7257
F-statistic: 345.9 on 3 and 388 DF,  p-value: < 2.2e-16

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

Carseats {ISLR} R Documentation

Sales of Child Car Seats

Description

A simulated data set containing sales of child car seats at 400 different stores.

Usage

`Carseats`

Format

A data frame with 400 observations on the following 11 variables.

`Sales`

Unit sales (in thousands) at each location

`CompPrice`

Price charged by competitor at each location

`Income`

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

```
Console ~/
> str(Carseats)
'data.frame':   400 obs. of  11 variables:
 $ Sales      : num  9.5 11.22 10.06 7.4 4.15 ...
 $ CompPrice  : num  138 111 113 117 141 124 115 136 132 132 ...
 $ Income     : num  73 48 35 100 64 113 105 81 110 113 ...
 $ Advertising: num  11 16 10 4 3 13 0 15 0 0 ...
 $ Population : num  276 260 269 466 340 501 45 425 108 131 ...
 $ Price      : num  120 83 80 97 128 72 108 120 124 124 ...
 $ ShelveLoc  : Factor w/ 3 levels "Bad","Good","Medium": 1 2 3 3 1 1 3
 2 3 3 ...
 $ Age        : num  42 65 59 55 38 78 71 67 76 76 ...
 $ Education  : num  17 10 12 14 13 16 15 10 10 17 ...
 $ Urban      : Factor w/ 2 levels "No","Yes": 2 2 2 2 2 1 2 2 1 1 ...
 $ US         : Factor w/ 2 levels "No","Yes": 2 2 2 2 1 2 1 2 1 2 ...
> 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
```

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

The model may be written as

$$Sales=13.0434689+(-0.0544588)\times Price+(-0.0219162)\times Urban+(1.2005727)\times US+\varepsilon$$

with Urban=1 if the store is in an urban location and 0 if not, and US=1 if the store is in the US and 0 if not.

(d) For which of the predictors can you reject the null hypothesis $H_0 : \beta_j = 0$

We can reject the null hypothesis for the “Price” and “US” variables.

(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 out- come.

```
> 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

> 
```

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

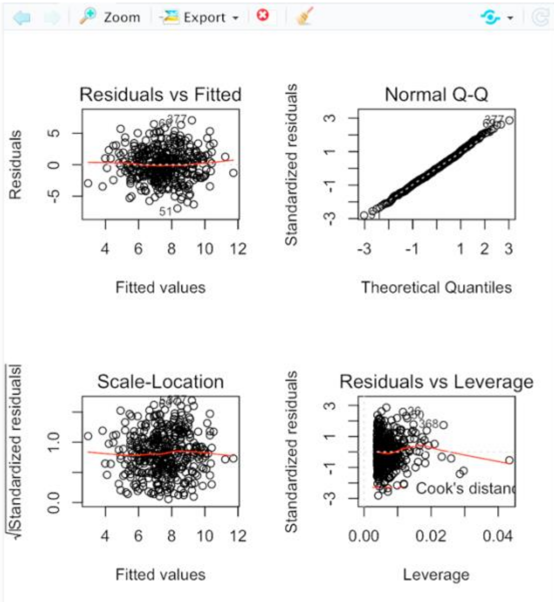
The R2 for the smaller model is marginally better than for the bigger model. Essentially about 23.9262888% of the variability is explained by the model.

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

```
> confint(lm.fit2)
                  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)?

```
> 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

> confint(lm.fit2)
                  2.5 %      97.5 %
(Intercept) 11.79032020 14.27126531
Price       -0.06475984 -0.04419543
USYes        0.69151957  1.70776632
> par(mfrow=c(2,2))
> plot(lm.fit2)
>
```



5.In this problem we will investigate the t-statistic for the null hypothesis $H_0: \beta = 0$ in simple linear regression without an intercept. To begin, we generate a predictor x and a response y as follows.

(a) Perform a simple linear regression of y onto x, without an intercept. Report the coefficient estimate $\hat{\beta}$, the standard error of this coefficient estimate, and the t- statistic and p-value associated with the null hypothesis $H_0: \beta = 0$. Comment on these results. (You can perform regression without an intercept using the command lm(y ∼ x + 0).)

```
> set.seed(1)
> x <- rnorm(100)
> y <- 2 * x + rnorm(100)
> fit5 <- lm(y ~ x + 0)
> summary(fit5)

Call:
lm(formula = y ~ x + 0)

Residuals:
    Min      1Q  Median      3Q     Max 
-1.9154 -0.6472 -0.1771  0.5056  2.3109 

Coefficients:
  Estimate Std. Error t value Pr(>|t|)    
x   1.9939     0.1065   18.73   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9586 on 99 degrees of freedom
Multiple R-squared:  0.7798,	Adjusted R-squared:  0.7776 
F-statistic: 350.7 on 1 and 99 DF,  p-value: < 2.2e-16

> 
```

(b) Now perform a simple linear regression of x onto y without an intercept, and report the coefficient estimate, its standard error, and the corresponding t-statistic and p-values associated with the null hypothesis $H_0 : \beta = 0$. Comment on these results.

```
> fit6 <- lm(x ~ y + 0)
> summary(fit6)

Call:
lm(formula = x ~ y + 0)

Residuals:
    Min      1Q  Median      3Q     Max 
-0.8699 -0.2368  0.1030  0.2858  0.8938 

Coefficients:
  Estimate Std. Error t value Pr(>|t|)    
y  0.39111    0.02089   18.73   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4246 on 99 degrees of freedom
Multiple R-squared:  0.7798,	Adjusted R-squared:  0.7776 
F-statistic: 350.7 on 1 and 99 DF,  p-value: < 2.2e-16

> 
```

(c) What is the relationship between the results obtained in (a) and (b)?

We obtain the same value for the t-statistic and consequently the same value for the corresponding p-value. Both results in (a) and (b) reflect the same line created in (a). In other words, $y=2x+\varepsilon$ could also be written $x=0.5(y-\varepsilon)$.

(d) For the regression of Y onto X without an intercept, the t-statistic for $H_0 : \beta = 0$

```
> n <- length(x)
> t <- sqrt(n - 1)*(x %*% y)/sqrt(sum(x^2) * sum(y^2) - (x %*% y)^2)
> as.numeric(t)
[1] 18.72593
> 
```

(e) Using the results from (d), argue that the t-statistic for the regression of y onto x is the same as the t-statistic for the regression of x onto y.

It is easy to see that if we replace xi by yi in the formula for the t-statistic, the result would be the same.

(f) In R, show that when regression is performed with an intercept, the t-statistic for H_0 : $\beta_1 = 0$. is the same for the regression of y onto x as it is for the regression of x onto y.

```
> fit7 <- lm(y ~ x)
> summary(fit7)

Call:
lm(formula = y ~ x)

Residuals:
    Min      1Q  Median      3Q     Max 
-1.8768 -0.6138 -0.1395  0.5394  2.3462 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -0.03769    0.09699  -0.389    0.698    
x            1.99894    0.10773  18.556   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9628 on 98 degrees of freedom
Multiple R-squared:  0.7784,	Adjusted R-squared:  0.7762 
F-statistic: 344.3 on 1 and 98 DF,  p-value: < 2.2e-16

> 
```

```
> fit8 <- lm(x ~ y)
> summary(fit8)

Call:
lm(formula = x ~ y)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.90848 -0.28101  0.06274  0.24570  0.85736 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.03880    0.04266    0.91    0.365    
y            0.38942    0.02099   18.56   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4249 on 98 degrees of freedom
Multiple R-squared:  0.7784,	Adjusted R-squared:  0.7762 
F-statistic: 344.3 on 1 and 98 DF,  p-value: < 2.2e-16

> 
```