annand_module02_lab01

Joseph Annand

2023-11-05

Load Libraries

```
library(ISLR2)
library(MASS)
```

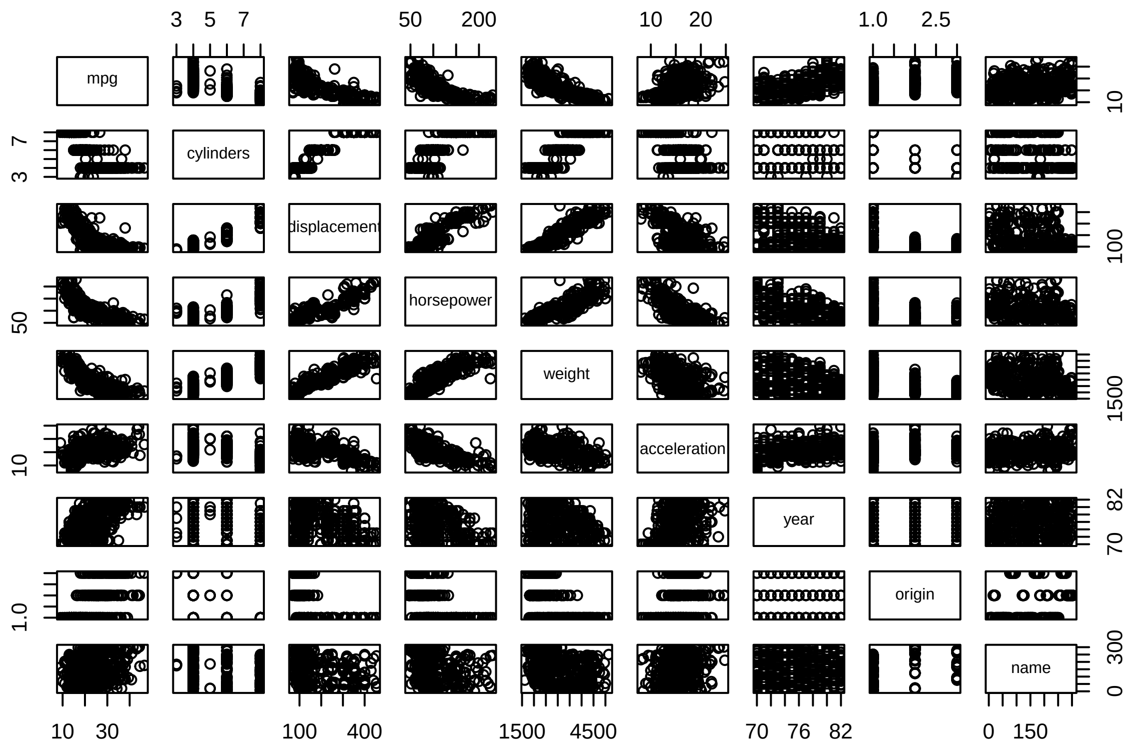
```
##
## Attaching package: 'MASS'

## The following object is masked from 'package:ISLR2':
##
##     Boston
```

Question 9

```
auto <- read.csv("Auto.csv", na.strings = "?", stringsAsFactors = T)
```

Part A

```
pairs(auto)
```



Part B

```r
auto_cor <- cor(auto[,-9], use="complete.obs")
print(auto_cor)
```

```
##                     mpg  cylinders displacement horsepower     weight
## mpg           1.0000000 -0.7776175   -0.8051269 -0.7784268 -0.8322442
## cylinders    -0.7776175  1.0000000    0.9508233  0.8429834  0.8975273
## displacement -0.8051269  0.9508233    1.0000000  0.8972570  0.9329944
## horsepower   -0.7784268  0.8429834    0.8972570  1.0000000  0.8645377
## weight       -0.8322442  0.8975273    0.9329944  0.8645377  1.0000000
## acceleration  0.4233285 -0.5046834   -0.5438005 -0.6891955 -0.4168392
## year          0.5805410 -0.3456474   -0.3698552 -0.4163615 -0.3091199
## origin        0.5652088 -0.5689316   -0.6145351 -0.4551715 -0.5850054
##              acceleration       year     origin
## mpg             0.4233285  0.5805410  0.5652088
## cylinders      -0.5046834 -0.3456474 -0.5689316
## displacement   -0.5438005 -0.3698552 -0.6145351
## horsepower     -0.6891955 -0.4163615 -0.4551715
## weight         -0.4168392 -0.3091199 -0.5850054
## acceleration    1.0000000  0.2903161  0.2127458
## year            0.2903161  1.0000000  0.1815277
## origin          0.2127458  0.1815277  1.0000000
```

Part C

```r
lm.auto <- lm(mpg ~ cylinders + displacement + horsepower + weight + acceleration + year,
              data = auto)

summary(lm.auto)
```

```
##
## Call:
## lm(formula = mpg ~ cylinders + displacement + horsepower + weight +
##     acceleration + year, data = auto)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.6927 -2.3864 -0.0801  2.0291 14.3607
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -1.454e+01  4.764e+00  -3.051  0.00244 **
## cylinders    -3.299e-01  3.321e-01  -0.993  0.32122
## displacement  7.678e-03  7.358e-03   1.044  0.29733
## horsepower   -3.914e-04  1.384e-02  -0.028  0.97745
## weight       -6.795e-03  6.700e-04 -10.141  < 2e-16 ***
## acceleration  8.527e-02  1.020e-01   0.836  0.40383
## year          7.534e-01  5.262e-02  14.318  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.435 on 385 degrees of freedom
##   (5 observations deleted due to missingness)
## Multiple R-squared:  0.8093, Adjusted R-squared:  0.8063
## F-statistic: 272.2 on 6 and 385 DF,  p-value: < 2.2e-16
```

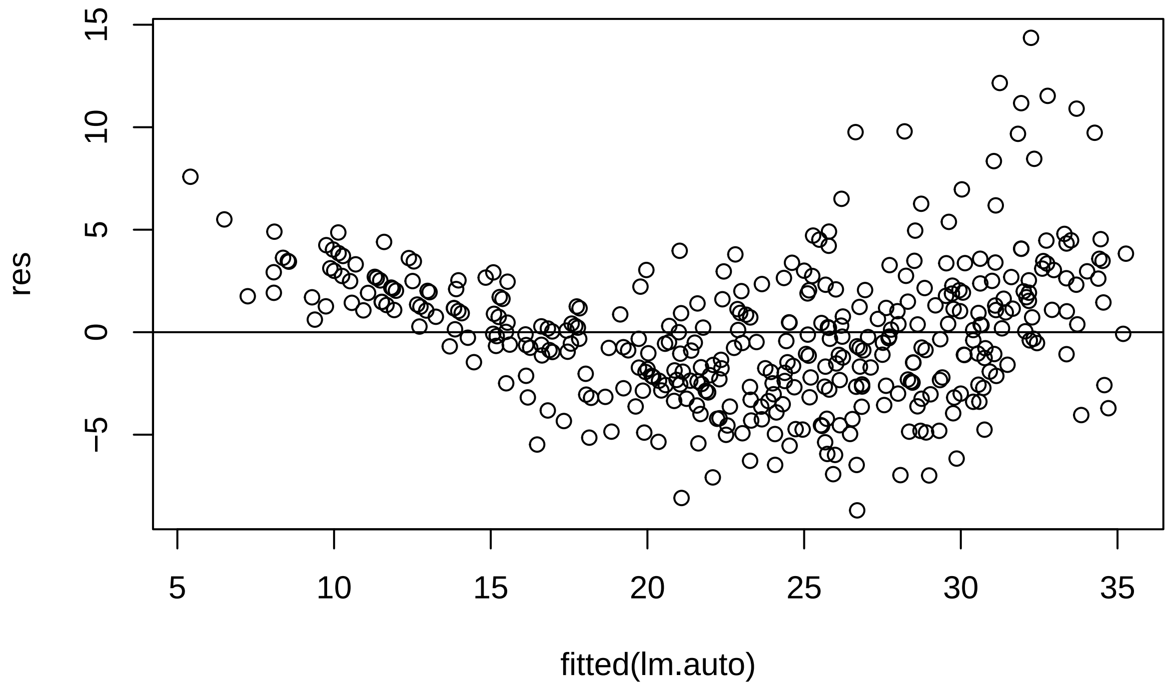
Subpart i According to the F-statistic being sufficiently larger than 1, there is a relationship between the response and the predictors.

Subpart ii Weight and Year are statistically significant predictors of mpg because the p-values for their coefficient estimates are sufficiently small.

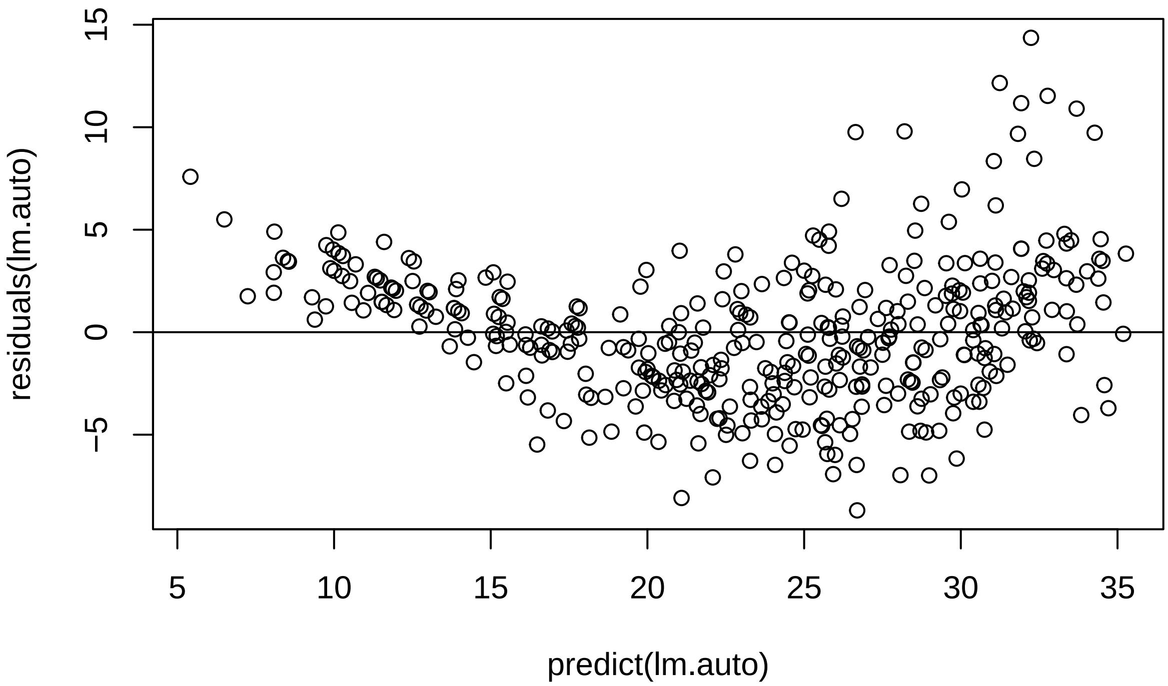
Subpart iii Coefficient for the year variable indicates that with each year younger the car was released the car gets 0.7534 miles more per gallon.

Part D

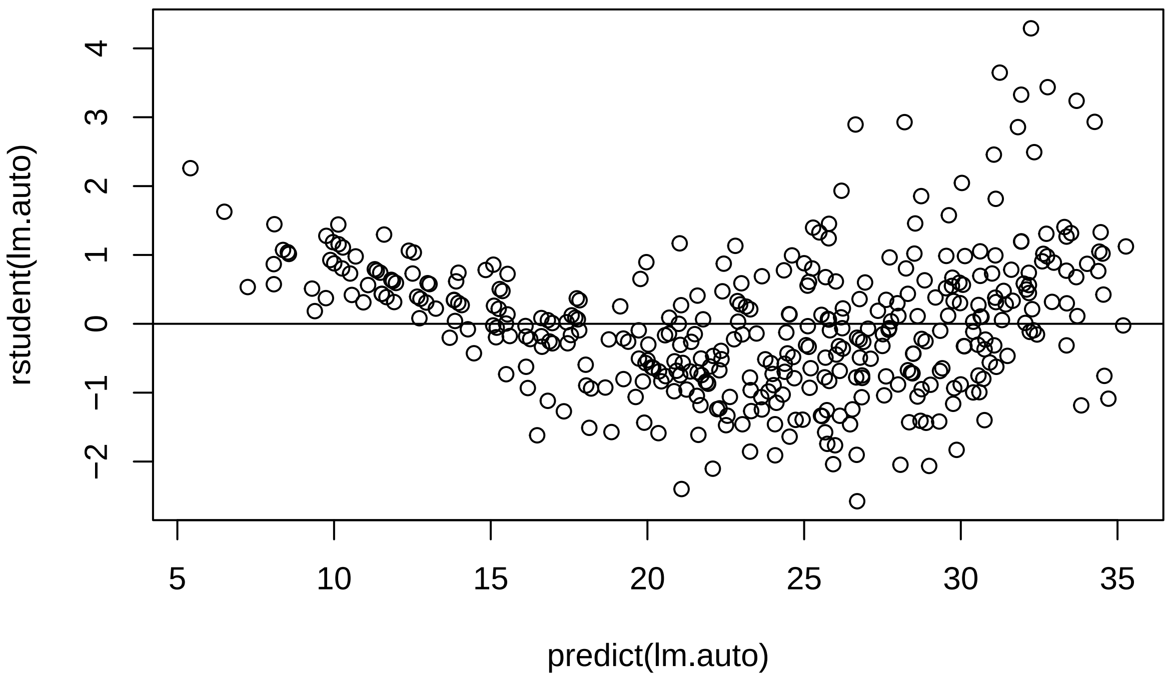
```r
#Residual plots for fitted multiple linear regression
res <- resid(lm.auto)
plot(fitted(lm.auto), res)
abline(0,0)
```



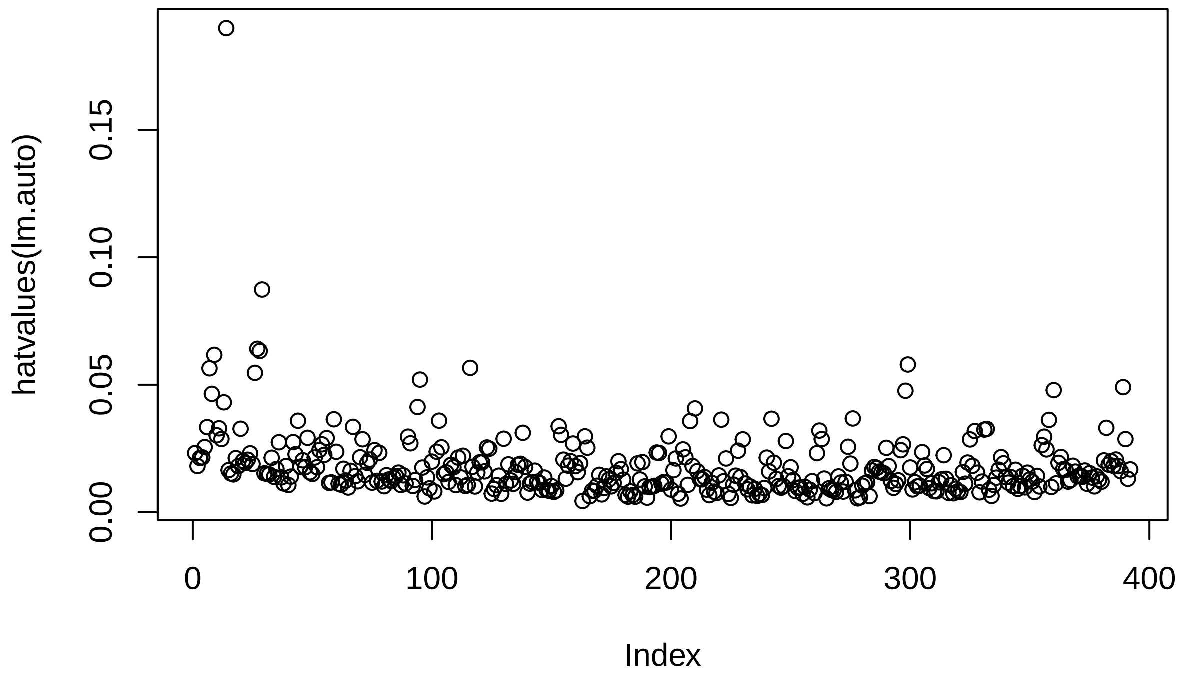

```r
plot(predict(lm.auto), residuals(lm.auto))
abline(0,0)
```




```
plot(predict(lm.auto), rstudent(lm.auto))
abline(0,0)
```




```
#Leverage plot for linear regression
plot(hatvalues(lm.auto))
```




```r
which.max(hatvalues(lm.auto))
```

```
## 14
## 14
```

The residuals plot does not show any unusually large outliers; however, it does show some non-linear shape in the data. The leverage lot shows at least one observation with unusually high leverage.

Part E

```r
# Interaction between weight and various variables
summary(lm(mpg ~ weight * year, data = auto))
```

```
##
## Call:
## lm(formula = mpg ~ weight * year, data = auto)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.0341 -1.9851 -0.0912  1.6987 12.9292
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept) -1.124e+02  1.280e+01  -8.781  < 2e-16 ***
## weight       2.821e-02  4.376e-03   6.447 3.34e-10 ***
## year         2.068e+00  1.699e-01  12.171  < 2e-16 ***
## weight:year -4.672e-04  5.857e-05  -7.977 1.66e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.187 on 393 degrees of freedom
## Multiple R-squared:  0.8354, Adjusted R-squared:  0.8341
## F-statistic: 664.9 on 3 and 393 DF,  p-value: < 2.2e-16
```

```
summary(lm(mpg ~ weight * acceleration, data = auto))
```

```
##
## Call:
## lm(formula = mpg ~ weight * acceleration, data = auto)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.5831  -2.7125  -0.3628   2.3091  15.6577
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)          2.855e+01  4.878e+00   5.854 1.01e-08 ***
## weight              -3.254e-03  1.464e-03  -2.222 0.026844 *
## acceleration         1.098e+00  3.098e-01   3.544 0.000442 ***
## weight:acceleration -2.753e-04  9.704e-05  -2.837 0.004789 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.271 on 393 degrees of freedom
## Multiple R-squared:  0.7044, Adjusted R-squared:  0.7021
## F-statistic: 312.1 on 3 and 393 DF,  p-value: < 2.2e-16
```

```
summary(lm(mpg ~ weight * cylinders, data = auto))
```

```
##
## Call:
## lm(formula = mpg ~ weight * cylinders, data = auto)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.5517  -2.6171  -0.4229   1.8263  16.7201
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)      65.8060333  3.7189927  17.695  < 2e-16 ***
## weight           -0.0129841  0.0013562  -9.574  < 2e-16 ***
## cylinders        -4.2652315  0.7226768  -5.902 7.76e-09 ***
## weight:cylinders  0.0011173  0.0002095   5.333 1.63e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 4.179 on 393 degrees of freedom
## Multiple R-squared:  0.7169, Adjusted R-squared:  0.7148
## F-statistic: 331.8 on 3 and 393 DF,  p-value: < 2.2e-16
```

```r
# Interaction between year and various variables
summary(lm(mpg ~ year * displacement, data = auto))
```

```
##
## Call:
## lm(formula = mpg ~ year * displacement, data = auto)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.9067  -2.4318  -0.2423   2.0392  17.0413
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)       -7.341e+01  8.298e+00  -8.846  < 2e-16 ***
## year               1.415e+00  1.092e-01  12.958  < 2e-16 ***
## displacement       2.559e-01  4.048e-02   6.321 7.06e-10 ***
## year:displacement -4.130e-03  5.438e-04  -7.594 2.28e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.737 on 393 degrees of freedom
## Multiple R-squared:  0.7738, Adjusted R-squared:  0.772
## F-statistic:   448 on 3 and 393 DF,  p-value: < 2.2e-16
```

```r
summary(lm(mpg ~ weight * acceleration, data = auto))
```

```
##
## Call:
## lm(formula = mpg ~ weight * acceleration, data = auto)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.5831  -2.7125  -0.3628   2.3091  15.6577
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)          2.855e+01  4.878e+00   5.854 1.01e-08 ***
## weight              -3.254e-03  1.464e-03  -2.222 0.026844 *
## acceleration         1.098e+00  3.098e-01   3.544 0.000442 ***
## weight:acceleration -2.753e-04  9.704e-05  -2.837 0.004789 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.271 on 393 degrees of freedom
## Multiple R-squared:  0.7044, Adjusted R-squared:  0.7021
## F-statistic: 312.1 on 3 and 393 DF,  p-value: < 2.2e-16
```

```r
summary(lm(mpg ~ weight * cylinders, data = auto))
```

```
##
## Call:
## lm(formula = mpg ~ weight * cylinders, data = auto)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.5517  -2.6171  -0.4229   1.8263  16.7201
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)      65.8060333  3.7189927  17.695  < 2e-16 ***
## weight           -0.0129841  0.0013562  -9.574  < 2e-16 ***
## cylinders        -4.2652315  0.7226768  -5.902 7.76e-09 ***
## weight:cylinders  0.0011173  0.0002095   5.333 1.63e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.179 on 393 degrees of freedom
## Multiple R-squared:  0.7169, Adjusted R-squared:  0.7148
## F-statistic: 331.8 on 3 and 393 DF,  p-value: < 2.2e-16
```

Strong interaction between weight and year as well as those two predictors with other variables

Part F

```r
# Log transformation of predictor variables
lm.log <- lm(mpg ~ log10(cylinders) + log10(displacement) + log10(horsepower) +
               log10(weight) + log10(acceleration) + log10(year),
             data = auto)

summary(lm.log)
```

```
##
## Call:
## lm(formula = mpg ~ log10(cylinders) + log10(displacement) + log10(horsepower) +
##     log10(weight) + log10(acceleration) + log10(year), data = auto)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.5641 -1.7873 -0.0611  1.5810 13.2714
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)          -62.413     17.650  -3.536 0.000456 ***
## log10(cylinders)       6.333      3.744   1.691 0.091585 .
## log10(displacement)   -7.843      3.121  -2.513 0.012371 *
## log10(horsepower)    -14.703      3.599  -4.085 5.36e-05 ***
## log10(weight)        -27.412      5.157  -5.316 1.80e-07 ***
## log10(acceleration)  -12.263      3.735  -3.283 0.001119 **
```

```
## log10(year)          126.240      8.278  15.250  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.103 on 385 degrees of freedom
##   (5 observations deleted due to missingness)
## Multiple R-squared:  0.8444, Adjusted R-squared:  0.8419
## F-statistic: 348.1 on 6 and 385 DF,  p-value: < 2.2e-16
```

In addition to weight and year, for horsepower and acceleration, we reject the null hypothesis that the parameter estimate is zero. F-statistic is much greater than 1.

```r
# Square root transformation of predictors
lm.sqrt <- lm(mpg ~ sqrt(cylinders) + sqrt(displacement) + sqrt(horsepower) +
               sqrt(weight) + sqrt(acceleration) + sqrt(year), data = auto)

summary(lm.sqrt)
```

```
##
## Call:
## lm(formula = mpg ~ sqrt(cylinders) + sqrt(displacement) + sqrt(horsepower) +
##     sqrt(weight) + sqrt(acceleration) + sqrt(year), data = auto)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.0770 -1.9915 -0.2719  1.7993 13.9583
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)        -45.0956     9.3107  -4.843 1.85e-06 ***
## sqrt(cylinders)      1.0224     1.5417   0.663   0.5076
## sqrt(displacement)  -0.1794     0.2132  -0.841   0.4007
## sqrt(horsepower)    -0.5345     0.3090  -1.730   0.0845 .
## sqrt(weight)        -0.6222     0.0807  -7.709 1.09e-13 ***
## sqrt(acceleration)  -0.9155     0.8524  -1.074   0.2835
## sqrt(year)          12.7588     0.8777  14.537  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.281 on 385 degrees of freedom
##   (5 observations deleted due to missingness)
## Multiple R-squared:  0.826,  Adjusted R-squared:  0.8233
## F-statistic: 304.7 on 6 and 385 DF,  p-value: < 2.2e-16
```

In addition to weight and year, for horsepower, we reject the null hypothesis that the parameter estimate is zero. F-statistic is much greater than 1.

```r
# X^2 transformation of predictor variables
lm.square <- lm(mpg ~ I(cylinders^2) + I(displacement^2) + I(horsepower^2)
                + I(weight^2) + I(acceleration^2) + I(year^2), data = auto)

summary(lm.square)
```

```
##
## Call:
## lm(formula = mpg ~ I(cylinders^2) + I(displacement^2) + I(horsepower^2) +
##     I(weight^2) + I(acceleration^2) + I(year^2), data = auto)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.9076 -2.6160 -0.0569  2.1774 14.7696
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)        3.084e+00  2.437e+00   1.265  0.20654
## I(cylinders^2)    -9.796e-02  2.626e-02  -3.730  0.00022 ***
## I(displacement^2)  4.477e-05  1.428e-05   3.135  0.00185 **
## I(horsepower^2)    1.975e-05  5.101e-05   0.387  0.69886
## I(weight^2)       -1.014e-06  9.272e-08 -10.934  < 2e-16 ***
## I(acceleration^2)  5.966e-03  2.808e-03   2.124  0.03429 *
## I(year^2)          5.078e-03  3.683e-04  13.788  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.695 on 385 degrees of freedom
##   (5 observations deleted due to missingness)
## Multiple R-squared:  0.7793, Adjusted R-squared:  0.7759
## F-statistic: 226.6 on 6 and 385 DF,  p-value: < 2.2e-16
```

In addition to weight and year, for all variables except horsepower, we reject the null hypothesis that the parameter estimate is zero. F-statistic is much greater than 1.

Question 10

```r
# Import dataset
carseats <- Carseats
```

Part A

```r
lm.carseat <- lm(Sales ~ Price + Urban + US, data = carseats)
summary(lm.carseat)
```

```
##
## 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

When Price increases by one unit and Urban = US = No, sales decrease by 0.54459. When Urban = Yes, Sales will be 0.021916 units less than if Urban = No. When US = Yes, Sales will be 1.200573 units more than if US = No.

Part C

Sales = B0 + B1 * Price + {dummy variable}

{dummy variable} = {B2 if Urban is Yes and US is No, B3 if Urban is No and US is Yes, B2 + B3 if Urban and US are Yes, or 0 if Urban and US are No}

Part D

We can reject the null hypothesis for Price and US.

Part E

```r
# Update model to remove Urban predictor variable
lm.carseat1 <- update(lm.carseat, ~ . - Urban)
summary(lm.carseat1)
```

```
##
## 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

Both models from parts a and e fit the data well based on the F-statistics and p-values.

```r
anova(lm.carseat, lm.carseat1)
```

```
## Analysis of Variance Table
##
## Model 1: Sales ~ Price + Urban + US
## Model 2: Sales ~ Price + US
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1    396 2420.8
## 2    397 2420.9 -1  -0.03979 0.0065 0.9357
```

Anova shows that there is no statistical difference between the two models from parts a and e.

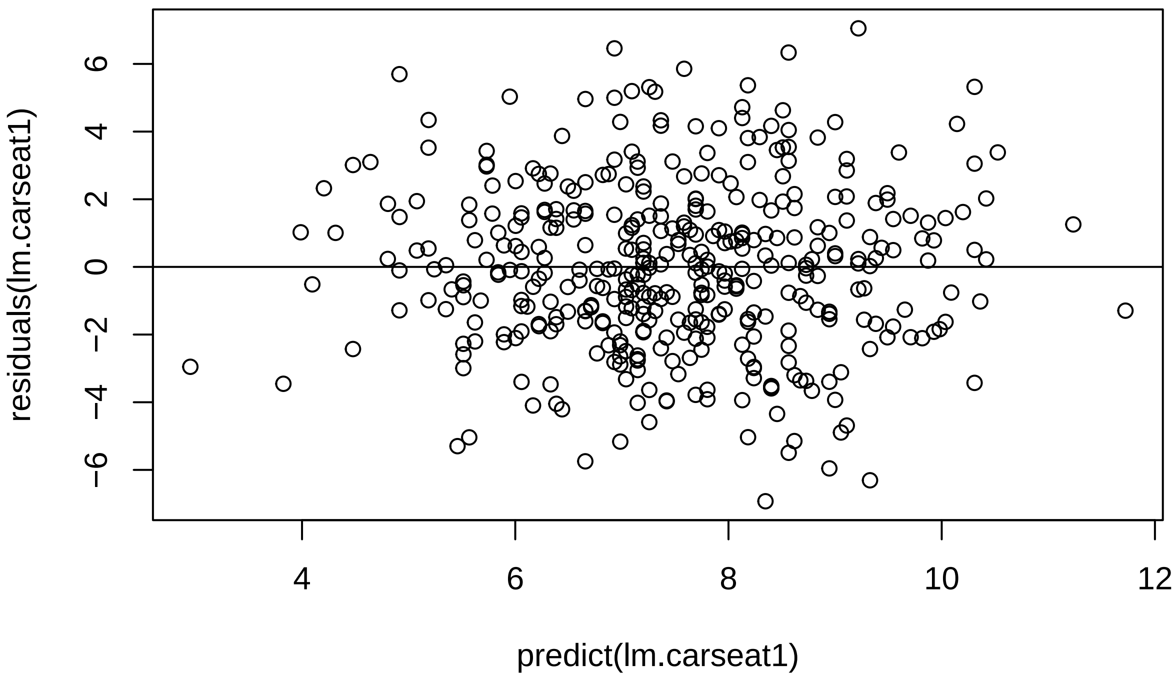
Part G

```r
confint(lm.carseat1)
```

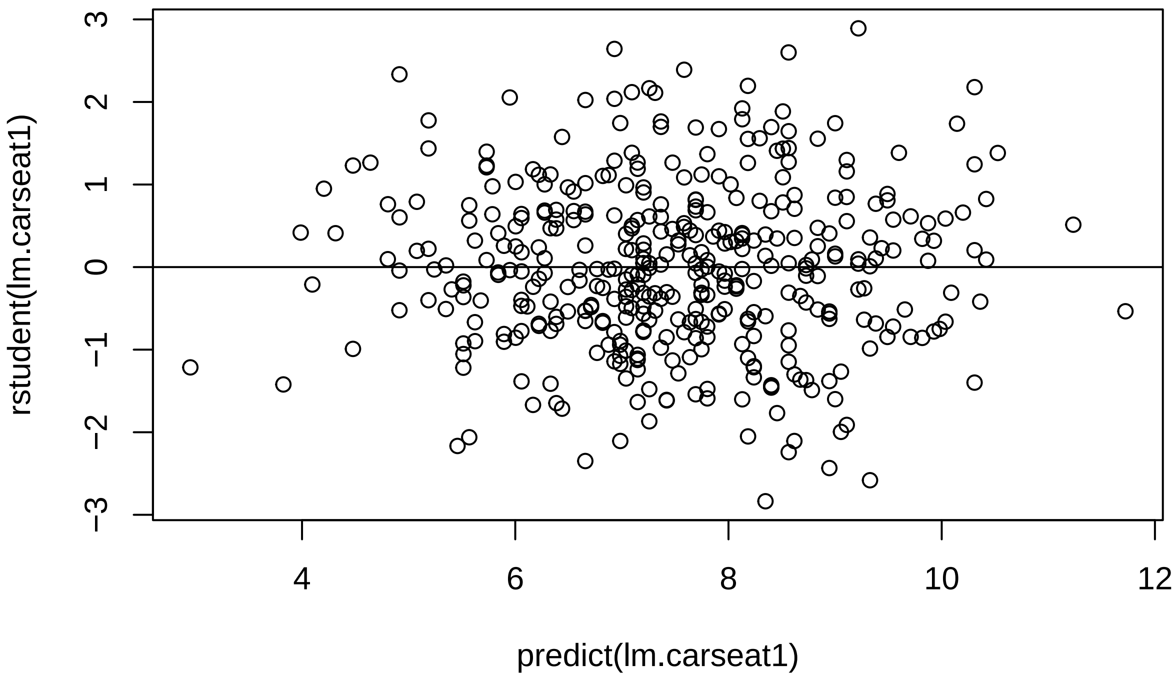
```
##                   2.5 %      97.5 %
## (Intercept) 11.79032020 14.27126531
## Price       -0.06475984 -0.04419543
## USYes        0.69151957  1.70776632
```

Part H

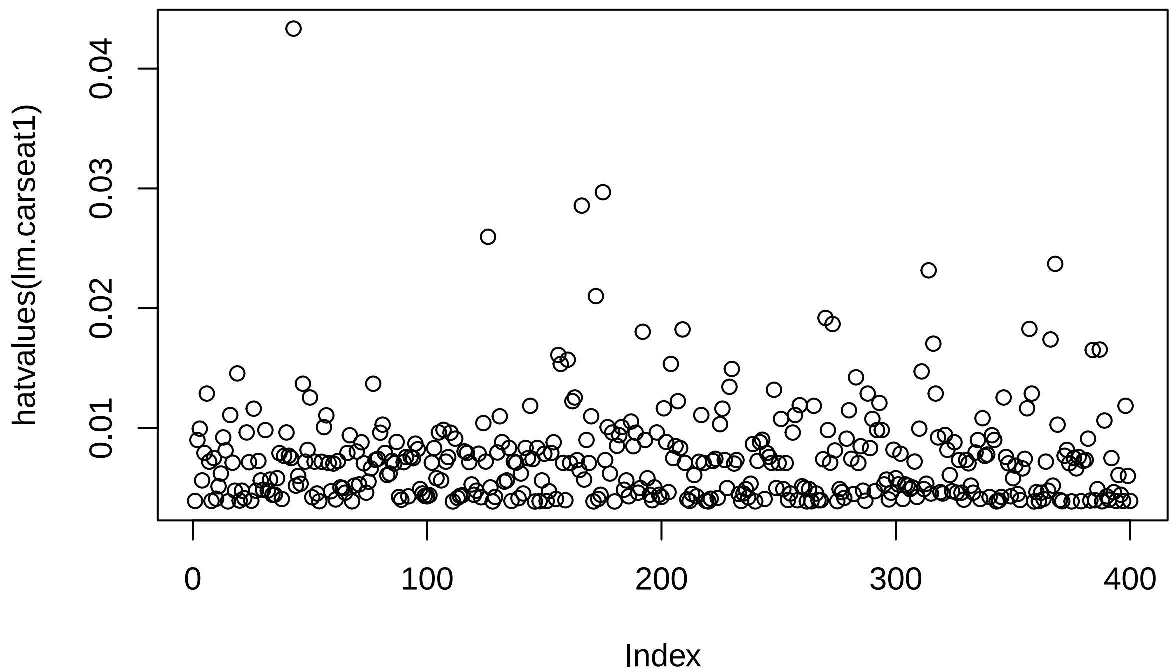
```r
plot(predict(lm.carseat1), residuals(lm.carseat1))
abline(0,0)
```




```
plot(predict(lm.carseat1), rstudent(lm.carseat1))
abline(0,0)
```




```
plot(hatvalues(lm.carseat1))
```




```r
which.max(hatvalues(lm.carseat1))
```

```
## 43
## 43
```

No evidence of outliers, but at least one observation with unusually high leverage.

Question 14

Part A

```r
set.seed(1)
x1 <- runif(100)
x2 <- 0.5 * x1 + rnorm(100) / 10
y <- 2 + 2 * x1 + 0.3 * x2 + rnorm(100)
```

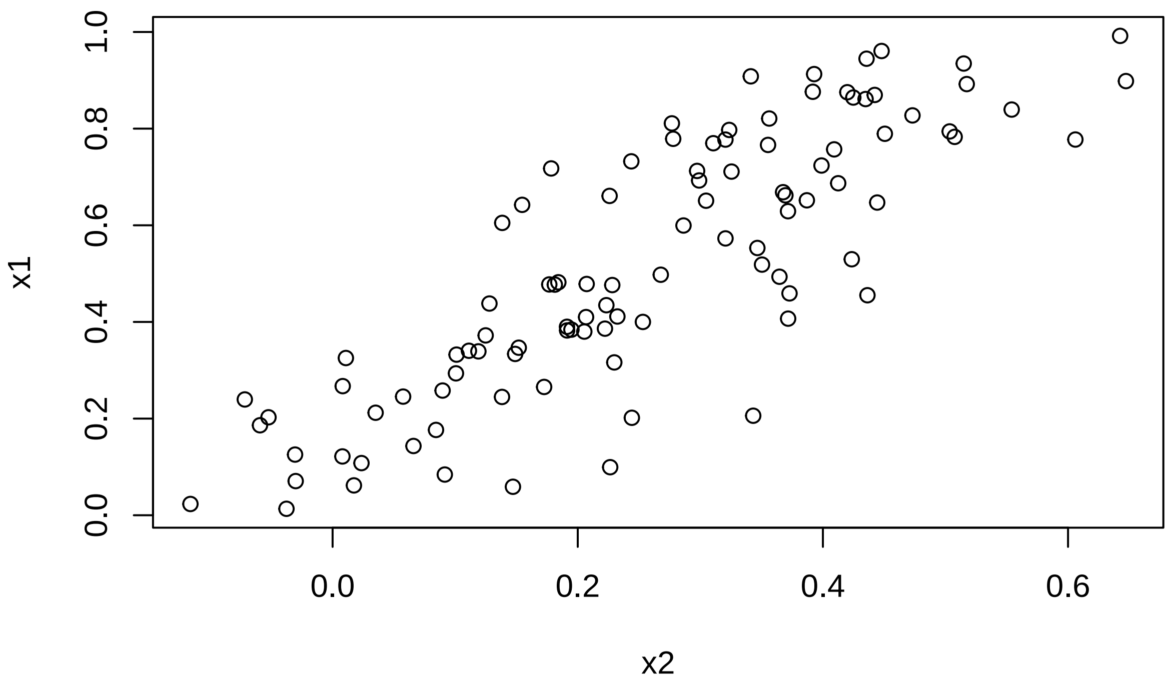
The regression coefficients are B1 = 2 and B2 = 0.3

Part B

```r
cor(x1, x2)
```

```
## [1] 0.8351212
```

```r
plot(x2, x1)
```



Part C

```r
lm.collin <- lm(y ~ x1 + x2)

summary(lm.collin)
```

```
##
## Call:
## lm(formula = y ~ x1 + x2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.8311 -0.7273 -0.0537  0.6338  2.3359
##
## Coefficients:
```

```
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept)   2.1305     0.2319   9.188 7.61e-15 ***
## x1            1.4396     0.7212   1.996   0.0487 *  
## x2            1.0097     1.1337   0.891   0.3754    
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 1.056 on 97 degrees of freedom
## Multiple R-squared:  0.2088, Adjusted R-squared:  0.1925 
## F-statistic:  12.8 on 2 and 97 DF,  p-value: 1.164e-05
```

B0 is 2.1305, B1 is 1.4396, and B2 is 1.0097. We can reject the null hypothesis that B1 = 0 because the p-value for the estimate is less than 0.05. We cannot reject the null hypothesis that B2 = 0 because the p-value is greater than 0.05. The F-statistic for the multiple linear regression is >1 and its corresponding p-value is much <0.05.

Part D

```r
lm.collin1 <- lm(y ~ x1)
summary(lm.collin1)
```

```
## 
## Call:
## lm(formula = y ~ x1)
## 
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -2.89495 -0.66874 -0.07785  0.59221  2.45560 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept)   2.1124     0.2307   9.155 8.27e-15 ***
## x1            1.9759     0.3963   4.986 2.66e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 1.055 on 98 degrees of freedom
## Multiple R-squared:  0.2024, Adjusted R-squared:  0.1942 
## F-statistic: 24.86 on 1 and 98 DF,  p-value: 2.661e-06
```

We can reject the null hypothesis that B1 = 0 because the p-value for the estimate is less than 0.05.

Part E

```r
lm.collin2 <- lm(y ~ x2)
summary(lm.collin2)
```

```
## 
## Call:
```

```
## lm(formula = y ~ x2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.62687 -0.75156 -0.03598  0.72383  2.44890
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.3899     0.1949   12.26  < 2e-16 ***
## x2            2.8996     0.6330    4.58 1.37e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.072 on 98 degrees of freedom
## Multiple R-squared:  0.1763, Adjusted R-squared:  0.1679
## F-statistic: 20.98 on 1 and 98 DF,  p-value: 1.366e-05
```

We can reject the null hypothesis that B1 = 0 because the p-value is <0.05.

Part F

These results do contradict each other as we fail to reject the null hypothesis for the B2 parameter estimate in part c but reject the null hypothesis for the same parameter estimate in part e. We can tell from part b, though, that x1 and x2 are collinear, meaning the power fo the hypothesis tests, and thus the probability of correctly detecting a non-zero coefficient, is reduced.

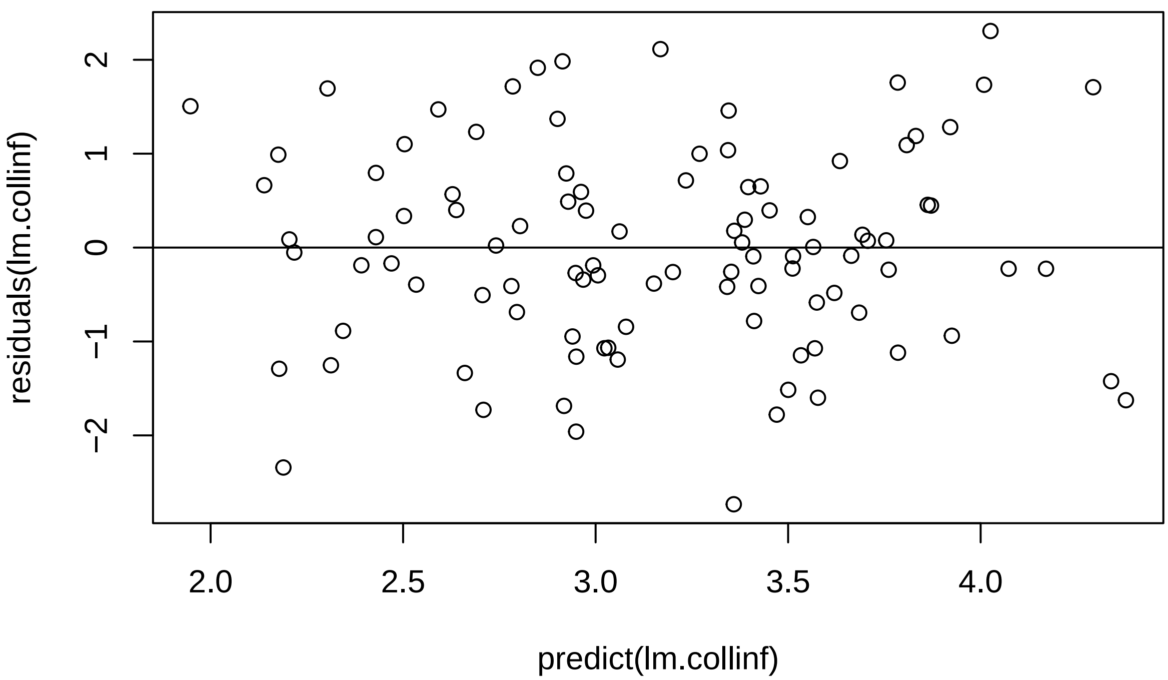
Part F

```r
x1 <- c(x1, 0.1)
x2 <- c(x2, 0.8)
y <- c(y, 6)

lm.collinf <- lm(y ~ x1 + x2)
summary(lm.collinf)
```

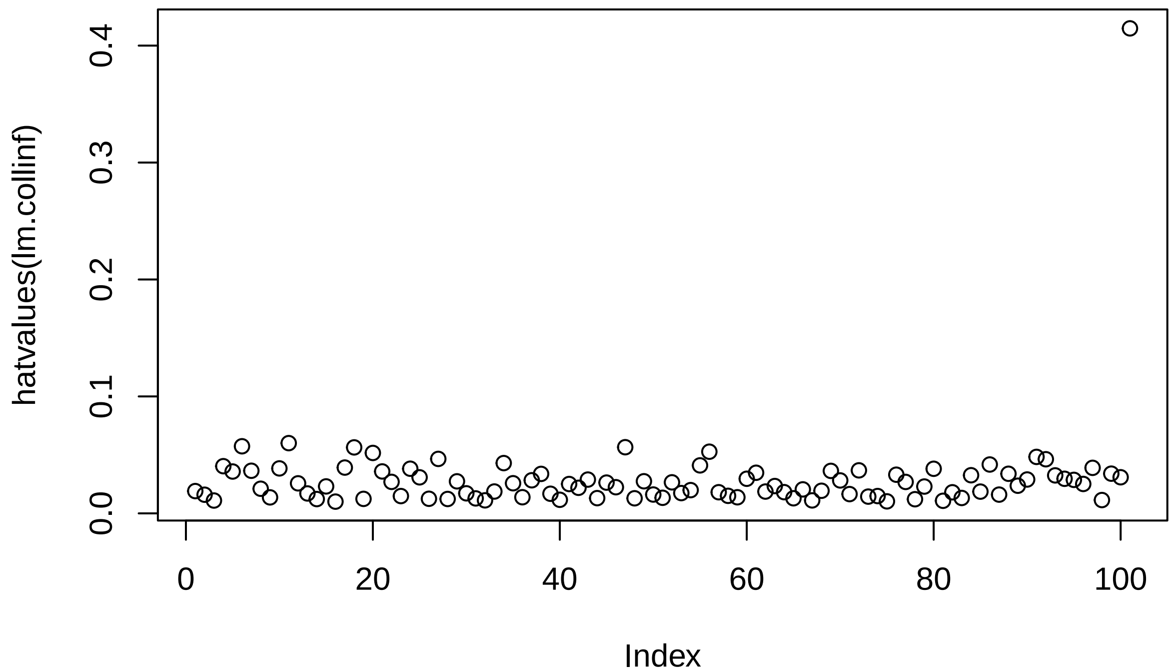
```
##
## Call:
## lm(formula = y ~ x1 + x2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.73348 -0.69318 -0.05263  0.66385  2.30619
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.2267     0.2314   9.624 7.91e-16 ***
## x1            0.5394     0.5922   0.911  0.36458
## x2            2.5146     0.8977   2.801  0.00614 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 1.075 on 98 degrees of freedom
## Multiple R-squared:  0.2188, Adjusted R-squared:  0.2029
## F-statistic: 13.72 on 2 and 98 DF,  p-value: 5.564e-06
```

```r
plot(predict(lm.collinf), residuals(lm.collinf))
abline(0,0)
```




```r
plot(hatvalues(lm.collinf))
```

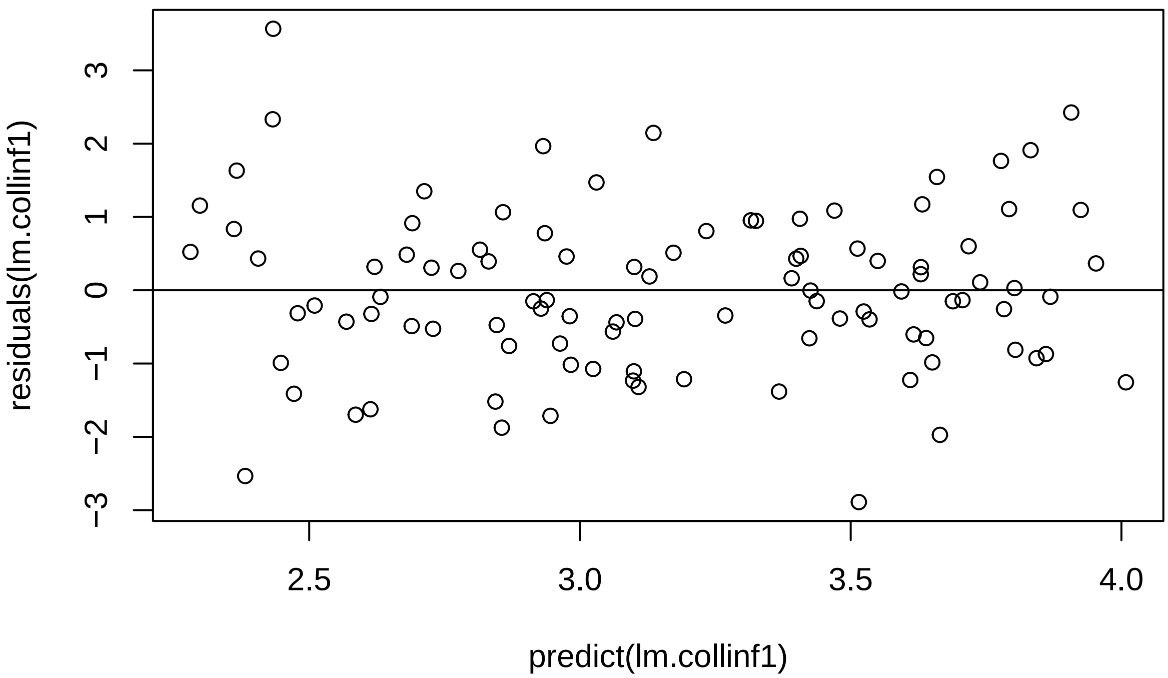


We now fail to reject the null hypothesis that B1 = 0 and we may reject the null hypothesis that B2 = 0. The observation does not appear to be an outlier but does have high leverage.

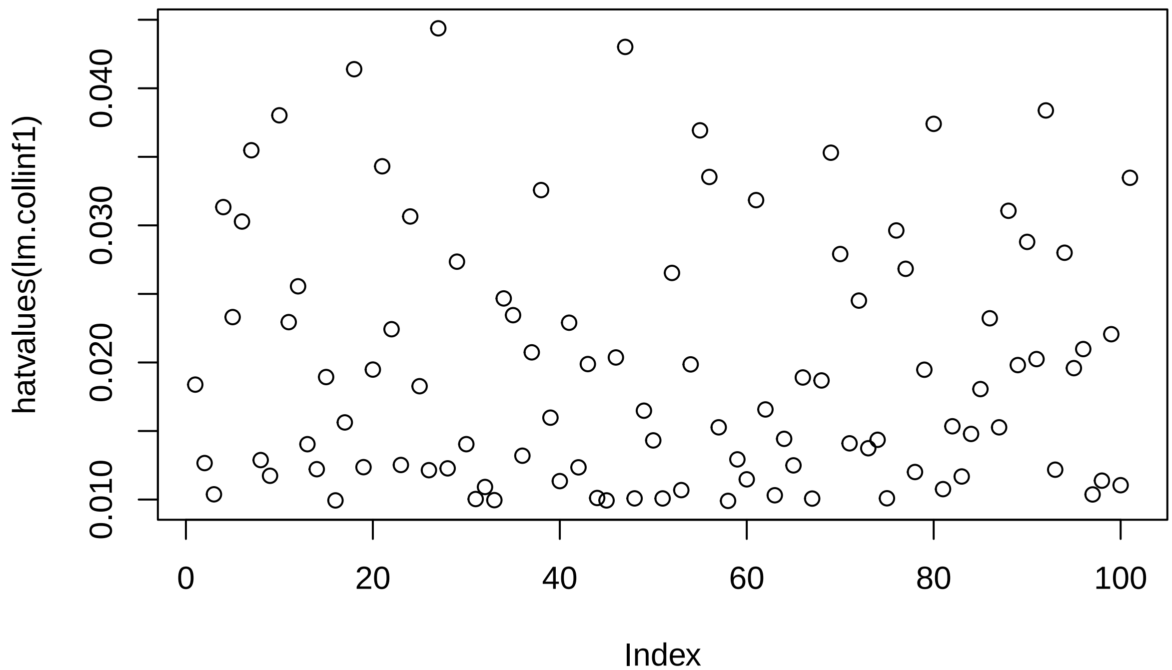
```r
lm.collinf1 <- lm(y ~ x1)
summary(lm.collinf1)
```

```
##
## Call:
## lm(formula = y ~ x1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.8897 -0.6556 -0.0909  0.5682  3.5665
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.2569     0.2390   9.445 1.78e-15 ***
## x1            1.7657     0.4124   4.282 4.29e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.111 on 99 degrees of freedom
## Multiple R-squared:  0.1562, Adjusted R-squared:  0.1477
## F-statistic: 18.33 on 1 and 99 DF,  p-value: 4.295e-05
```

```r
plot(predict(lm.collinf1), residuals(lm.collinf1))
abline(0,0)
```




```r
plot(hatvalues(lm.collinf1))
```

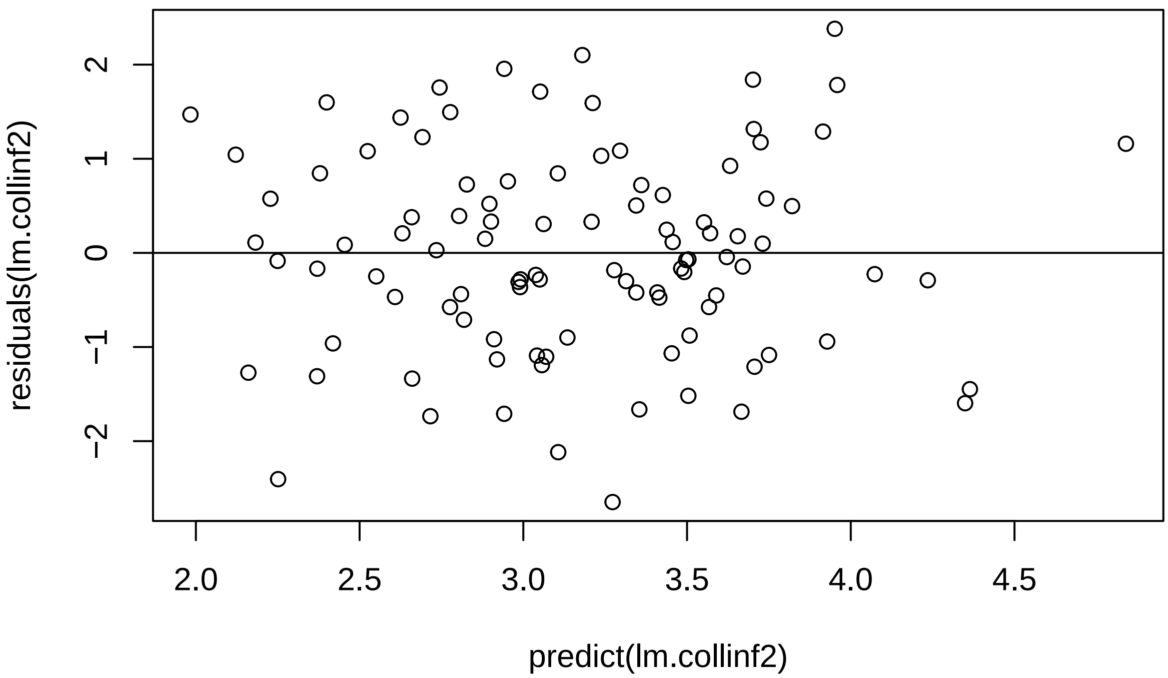


We reject the null hypothesis for the estimate of B1, which is the same as in part d. The residuals have constant variability and there are no obvious outliers. No observation stands out with an unusually high leverage.

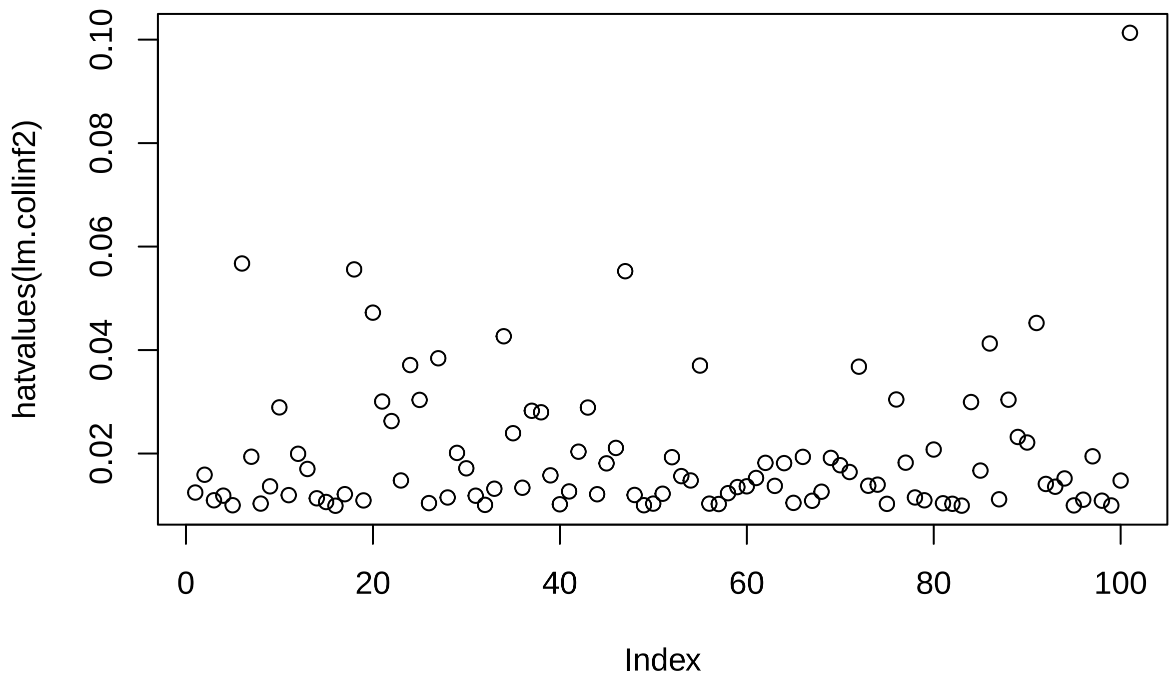
```
lm.collinf2 <- lm(y ~ x2)
summary(lm.collinf2)
```

```
##
## Call:
## lm(formula = y ~ x2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.64729 -0.71021 -0.06899  0.72699  2.38074
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.3451     0.1912  12.264  < 2e-16 ***
## x2            3.1190     0.6040   5.164 1.25e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.074 on 99 degrees of freedom
## Multiple R-squared:  0.2122, Adjusted R-squared:  0.2042
## F-statistic: 26.66 on 1 and 99 DF,  p-value: 1.253e-06
```

```r
plot(predict(lm.collinf2), residuals(lm.collinf2))
abline(0,0)
```




```r
plot(hatvalues(lm.collinf2))
```



Similar to part e, we reject the null hypothesis that B1 = 0. Residual plot does show any particularly large outliers. There is a high-leverage point.