HW4

Adhrit Srivastav, eid: Ams22362

9/30/2021

3.47 Driving fatalities and speed limits

```
speed = read.csv("C:/Users/adhri/OneDrive/Documents/R/App_Reg_and_Time_Series/chpt3/datasets/Speed.csv")
attach(speed)

# full model
modelYSYS = lm(FatalityRate ~ Year + StateControl + Year*StateControl)
summary(modelYSYS)
```

```
##
## Call:
## lm(formula = FatalityRate ~ Year + StateControl + Year * StateControl)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.103571 -0.020769  0.004048  0.022473  0.091667
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)        2.162e+02  1.303e+01   16.59 6.19e-12 ***
## Year              -1.076e-01  6.548e-03  -16.44 7.19e-12 ***
## StateControl      -1.614e+02  1.447e+01  -11.15 3.07e-09 ***
## Year:StateControl  8.097e-02  7.264e-03   11.15 3.08e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04243 on 17 degrees of freedom
## Multiple R-squared:  0.9831, Adjusted R-squared:  0.9801
## F-statistic:   329 on 3 and 17 DF,  p-value: 2.998e-15
```

```
anova(modelYSYS)
```

```
## Analysis of Variance Table
##
## Response: FatalityRate
##                   Df  Sum Sq Mean Sq  F value    Pr(>F)
## Year               1 1.55026 1.55026 860.9841 5.288e-16 ***
## StateControl       1 0.00292 0.00292   1.6211    0.2201
## Year:StateControl  1 0.22373 0.22373 124.2562 3.082e-09 ***
```

```
## Residuals         17 0.03061 0.00180
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# reduced model
modelY = lm(FatalityRate ~ Year)
summary(modelY)
```

```
##
## Call:
## lm(formula = FatalityRate ~ Year)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.18959 -0.07550 -0.02576  0.09346  0.24606
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 91.320887   8.374227    10.9 1.28e-09 ***
## Year        -0.044870   0.004193   -10.7 1.75e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1164 on 19 degrees of freedom
## Multiple R-squared:  0.8577, Adjusted R-squared:  0.8502
## F-statistic: 114.5 on 1 and 19 DF,  p-value: 1.75e-09
```

```r
anova(modelY)
```

```
## Analysis of Variance Table
##
## Response: FatalityRate
##           Df  Sum Sq Mean Sq F value   Pr(>F)
## Year       1 1.55026 1.55026  114.49 1.75e-09 ***
## Residuals 19 0.25726 0.01354
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# reduced model
# = lm(FatalityRate ~ Year + StateControl)
#summary(modelYS)
#anova(modelYS)

t.test(Year*StateControl, FatalityRate)
```

```
##
##  Welch Two Sample t-test
##
## data:  Year * StateControl and FatalityRate
## t = 5.693, df = 20, p-value = 1.429e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
```

```
##   783.7487 1690.2494
## sample estimates:
##   mean of x   mean of y
## 1238.714286    1.715238
```

```
#anova(modelY, modelYS)
anova(modelY, modelYSYS)
```

```
## Analysis of Variance Table
##
## Model 1: FatalityRate ~ Year
## Model 2: FatalityRate ~ Year + StateControl + Year * StateControl
##   Res.Df     RSS Df Sum of Sq      F    Pr(>F)
## 1     19 0.25726
## 2     17 0.03061  2   0.22665 62.939 1.386e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

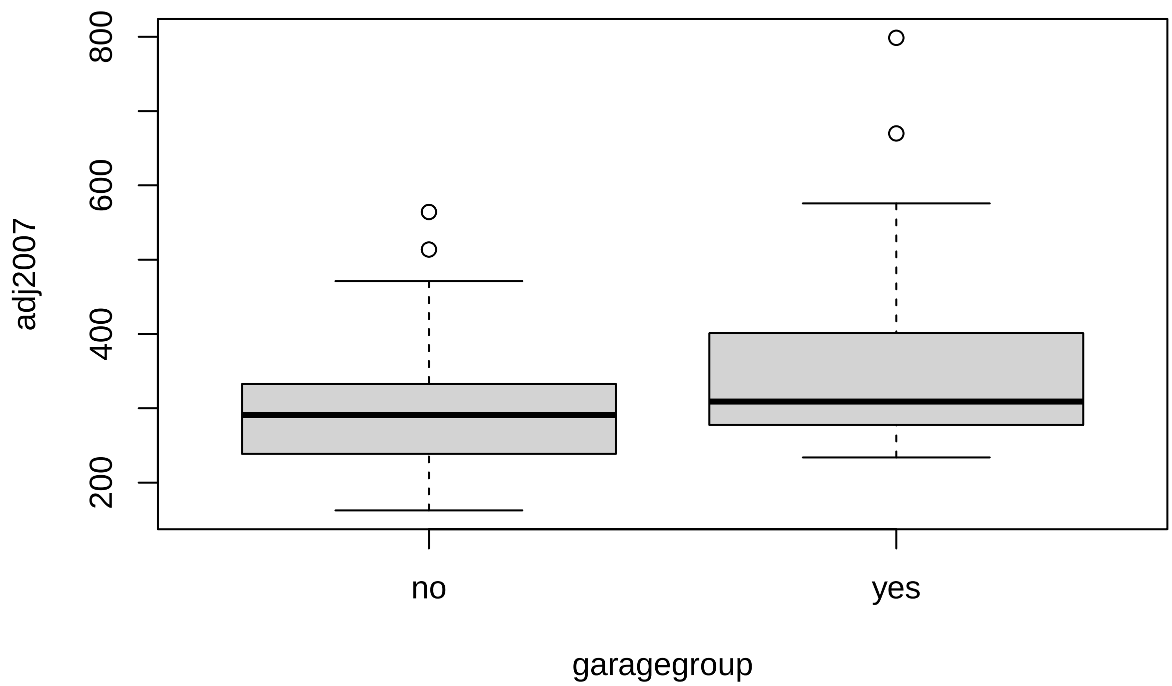
a) There is a significant difference between the slopes of the two lines because the F-statistic is 62.939 with a p-value of 1.386e-08 which gives us evidence that at least one of StateControl and Year:StateControl is useful for predicting FatalityRate. There is not a significant difference between the intercepts because the p-values are very small for both.

b) The nested F-test is similar to the one used in part A in which it was determined that there is a significant difference in slopes. The F-statistic is 62.939 with a p-value of 1.386e-08 which indicates that the B_3 slope is significant. This result is the same as a t-test for that coefficient because when run with that coefficient, a p-value of 1.429e-05 is produced, indicating statistical significance of the coefficient.

3.48 Real estate near Rails to Trails

```
rail = read.csv("C:/Users/adhri/OneDrive/Documents/R/App_Reg_and_Time_Series/chpt3/datasets/RailsTrails
attach(rail)
```

```
# a
boxplot(adj2007 ~ garagegroup)
```




```
t.test(adj2007 ~ garagegroup)
```

```
##
##  Welch Two Sample t-test
##
## data:  adj2007 by garagegroup
## t = -2.7145, df = 94.013, p-value = 0.007896
## alternative hypothesis: true difference in means between group no and group yes is not equal to 0
## 95 percent confidence interval:
##  -93.36936 -14.48237
## sample estimates:
##  mean in group no mean in group yes
##          300.0728          353.9987
```

```
# b
modelD = lm(adj2007 ~ distance)
summary(modelD)
```

```
##
## Call:
## lm(formula = adj2007 ~ distance)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -190.55  -58.19  -17.48   25.22  444.41
```

```
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  388.204     14.052  27.626  < 2e-16 ***
## distance     -54.427      9.659  -5.635 1.56e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 92.13 on 102 degrees of freedom
## Multiple R-squared:  0.2374, Adjusted R-squared:  0.2299
## F-statistic: 31.75 on 1 and 102 DF,  p-value: 1.562e-07
```

```
# c
modelDG = lm(adj2007 ~ distance + garagegroup)
summary(modelDG)
```

```
##
## Call:
## lm(formula = adj2007 ~ distance + garagegroup)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -167.88  -51.55  -21.88   36.79  427.49
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)     365.103     17.661  20.673   <2e-16 ***
## distance        -51.025      9.638  -5.294    7e-07 ***
## garagegroupyes   37.892     18.032   2.101   0.0381 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 90.62 on 101 degrees of freedom
## Multiple R-squared:  0.2693, Adjusted R-squared:  0.2549
## F-statistic: 18.62 on 2 and 101 DF,  p-value: 1.311e-07
```

```
# d
modeldg = lm(adj2007 ~ distance*garagegroup)
summary(modeldg)
```

```
##
## Call:
## lm(formula = adj2007 ~ distance * garagegroup)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -162.46  -51.65  -17.22   30.04  425.76
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)              359.083     21.295  16.862  < 2e-16 ***
## distance                 -46.302     13.391  -3.458 0.000802 ***
## garagegroupyes            48.862     28.108   1.738 0.085222 .
```

```
## distance:garagegroupyes   -9.878     19.366  -0.510 0.611125
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 90.96 on 100 degrees of freedom
## Multiple R-squared:  0.2712, Adjusted R-squared:  0.2494
## F-statistic: 12.41 on 3 and 100 DF,  p-value: 5.785e-07
```

```
# e
anova(modelD, modelDG)
```

```
## Analysis of Variance Table
##
## Model 1: adj2007 ~ distance
## Model 2: adj2007 ~ distance + garagegroup
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1    102 865718
## 2    101 829453  1     36265 4.4159 0.03809 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

a) The p-value of .007896 indicates there is a significant difference between the price of a home based on if it has garage spaces or not.

b) For every additional unit of distance, the home price decreases approximately on the average by 54.427.

c) For every additional unit of distance, the home price decreases approximately on the average by 51.025 after adjusting/controlling for garagegroup. For every additional yes on if there are garage spaces, the home price increases approximately on the average by 37.892 after adjusting/controlling for distance.

d) The distance variable has a p-value < .05, where as garagegroupyes, and distance:garagegroupyes are not. For every additional unit of distance, the home price decreases approximately on the average by 46.302.

e) The p-value is .03809 which is less than .05 and indicates that garagegroup is statistically significant and is useful for predicting the selling price.

3.49 Real estate near Rails to Trails: transformation

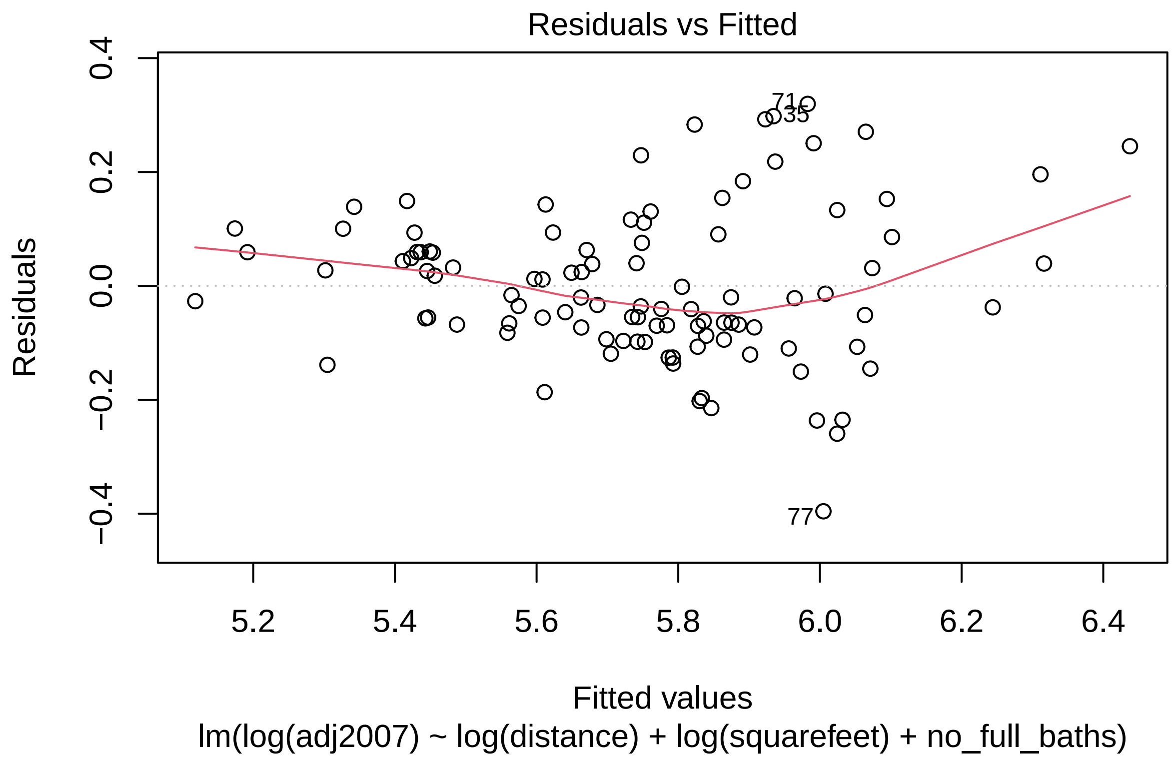
```
rail = read.csv("C:/Users/adhri/OneDrive/Documents/R/App_Reg_and_Time_Series/chpt3/datasets/RailsTrails
attach(rail)
```

```
## The following objects are masked from rail (pos = 3):
##
##     acre, acregroup, adj1998, adj2007, adj2011, bedgroup, bedrooms,
##     bikescore, diff2014, distance, distgroup, garage_spaces,
##     garagegroup, housenum, latitude, longitude, no_full_baths,
##     no_half_baths, no_rooms, pctchange, price1998, price2007,
##     price2011, price2014, sfgroup, squarefeet, streetname, streetno,
##     walkscore, zip
```

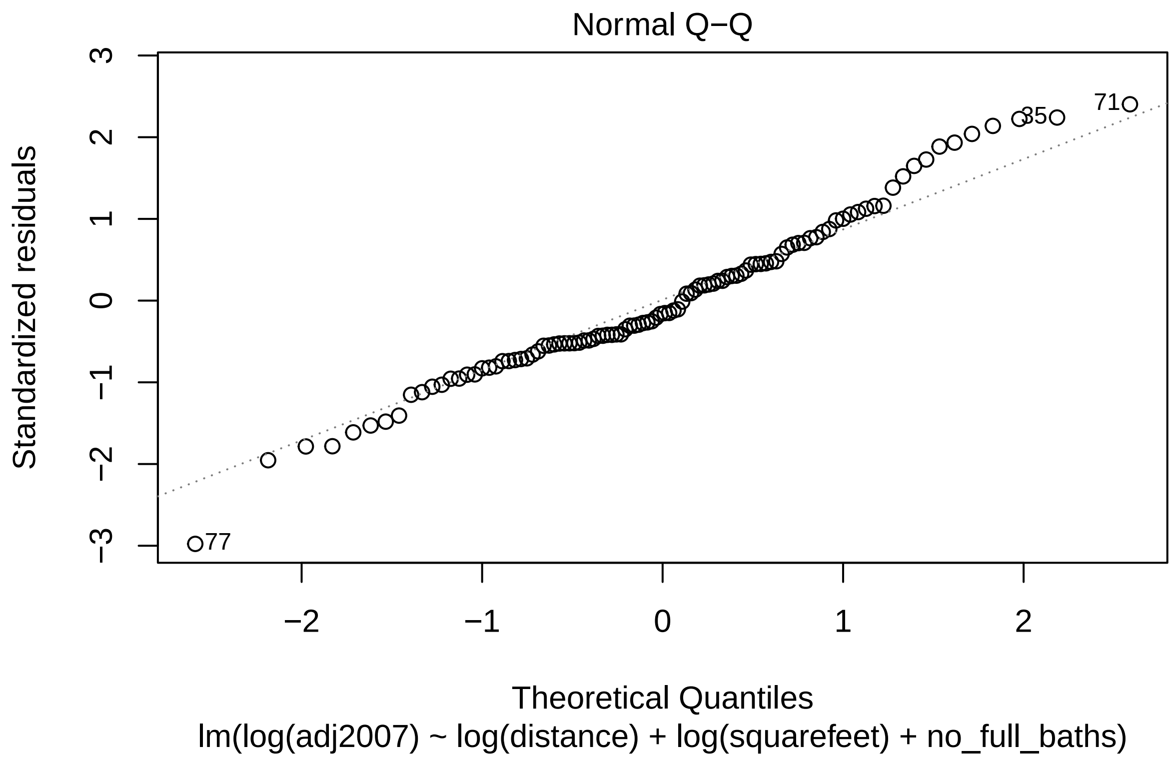
```r
# a
modelDSN = lm(log(adj2007) ~ log(distance) + log(squarefeet) + no_full_baths)
summary(modelDSN)
```

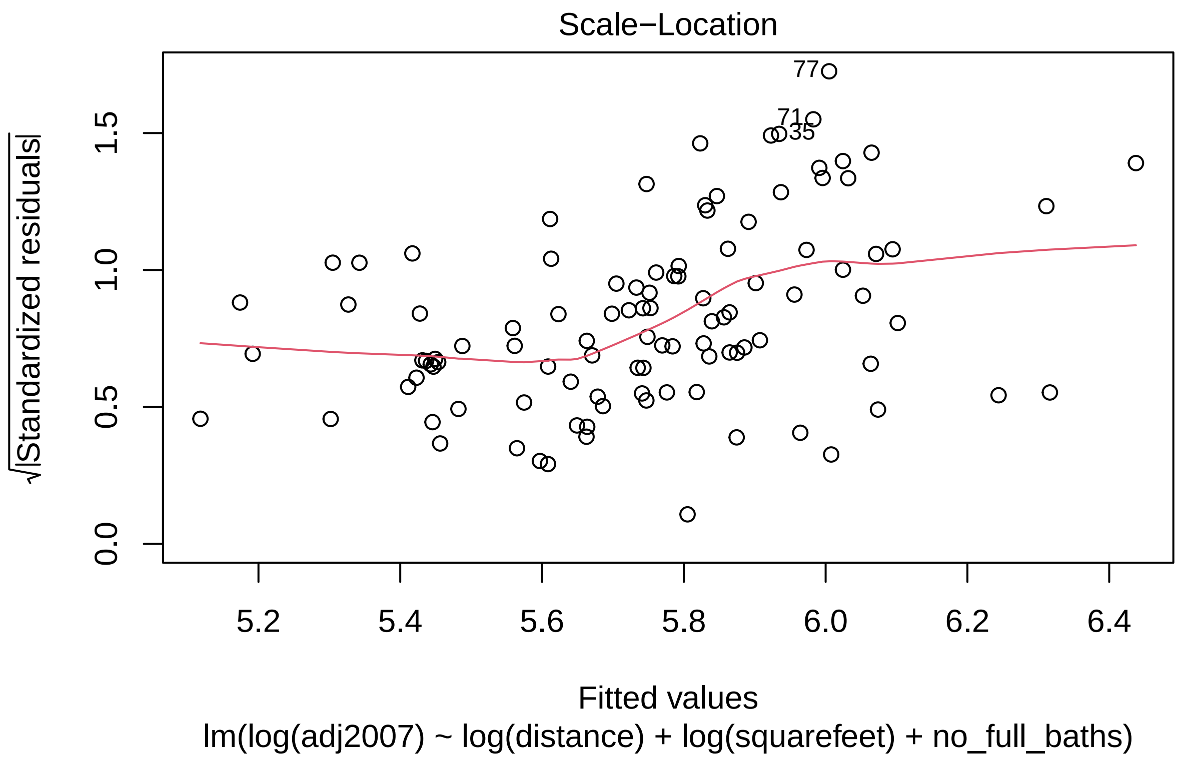
```
##
## Call:
## lm(formula = log(adj2007) ~ log(distance) + log(squarefeet) +
##     no_full_baths)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.39580 -0.07536 -0.02103  0.07813  0.31959
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      5.41777    0.03368 160.870  < 2e-16 ***
## log(distance)   -0.04883    0.01245  -3.922 0.000161 ***
## log(squarefeet)  0.59328    0.04567  12.991  < 2e-16 ***
## no_full_baths    0.05667    0.02500   2.267 0.025548 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1344 on 100 degrees of freedom
## Multiple R-squared:  0.7834, Adjusted R-squared:  0.7769
## F-statistic: 120.6 on 3 and 100 DF,  p-value: < 2.2e-16
```

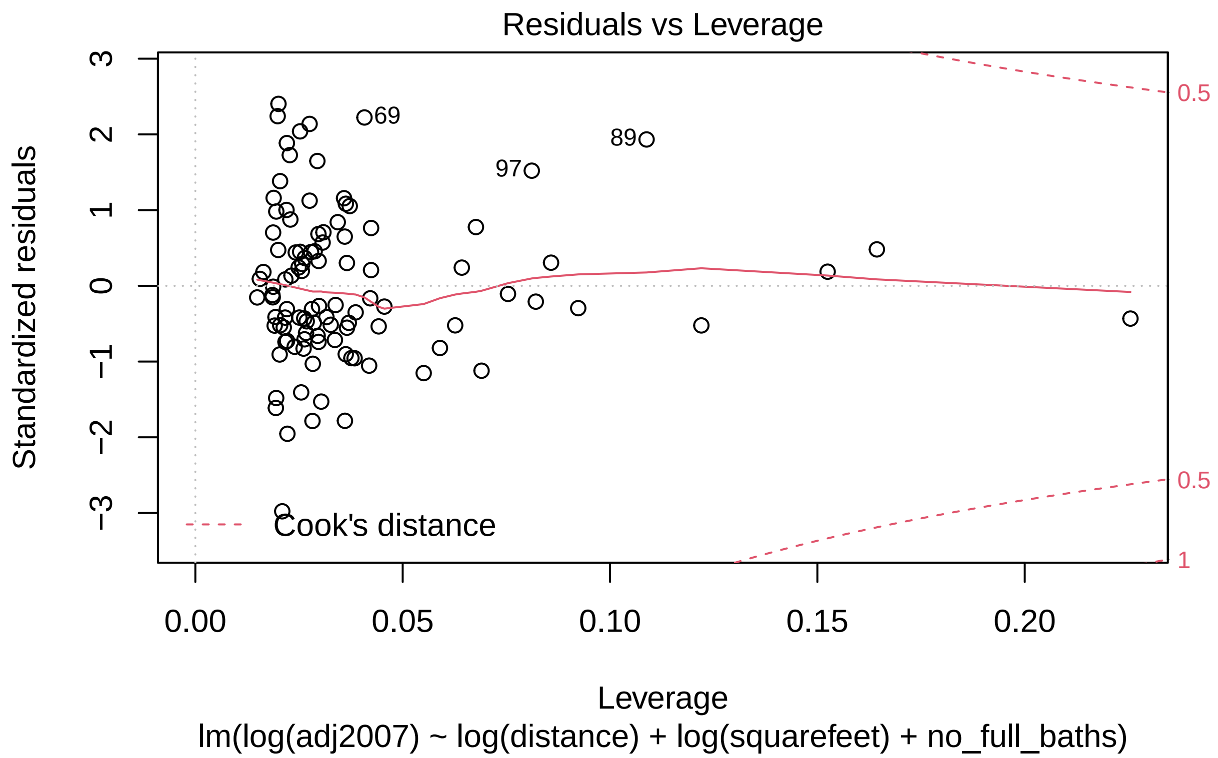
```r
# b
plot(modelDSN)
```



Residuals vs Fitted
Residuals
0.4
0.2
0.0
−0.2
−0.4
71
35
77
5.2
5.4
5.6
5.8
6.0
6.2
6.4
Fitted values
lm(log(adj2007) ~ log(distance) + log(squarefeet) + no_full_baths)


Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(log(adj2007) ~ log(distance) + log(squarefeet) + no_full_baths)
77
35
71
−3
−2
−1
0
1
2
3


Scale−Location
77
71
35
0.0
0.5
1.0
1.5
√|Standardized residuals|
5.2
5.4
5.6
5.8
6.0
6.2
6.4
Fitted values
lm(log(adj2007) ~ log(distance) + log(squarefeet) + no_full_baths)




```r
# c
modeldsn = lm(log(adj2007) ~ log(distance)*log(squarefeet) + log(distance)*no_full_baths + log(squarefe
summary(modeldsn)
```

```
##
## Call:
## lm(formula = log(adj2007) ~ log(distance) * log(squarefeet) +
##     log(distance) * no_full_baths + log(squarefeet) * no_full_baths +
##     log(distance) * log(squarefeet) * no_full_baths)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.39103 -0.07478 -0.00479  0.06668  0.32790
##
## Coefficients:
##                                              Estimate Std. Error t value
## (Intercept)                                  5.545207   0.058168  95.331
## log(distance)                               -0.040887   0.045200  -0.905
## log(squarefeet)                              0.355179   0.102008   3.482
## no_full_baths                               -0.048636   0.047595  -1.022
## log(distance):log(squarefeet)               -0.024984   0.083870  -0.298
## log(distance):no_full_baths                 -0.009463   0.034035  -0.278
## log(squarefeet):no_full_baths                0.172022   0.064910   2.650
## log(distance):log(squarefeet):no_full_baths  0.018293   0.054586   0.335
##                                             Pr(>|t|)
## (Intercept)                                  < 2e-16 ***
```

```
## log(distance)                               0.367955
## log(squarefeet)                             0.000751 ***
## no_full_baths                               0.309413
## log(distance):log(squarefeet)               0.766428
## log(distance):no_full_baths                 0.781580
## log(squarefeet):no_full_baths               0.009410 **
## log(distance):log(squarefeet):no_full_baths 0.738263
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1316 on 96 degrees of freedom
## Multiple R-squared:  0.8007, Adjusted R-squared:  0.7861
## F-statistic: 55.09 on 7 and 96 DF,  p-value: < 2.2e-16
```

```
# d
anova(modelDSN, modeldsn)
```

```
## Analysis of Variance Table
##
## Model 1: log(adj2007) ~ log(distance) + log(squarefeet) + no_full_baths
## Model 2: log(adj2007) ~ log(distance) * log(squarefeet) + log(distance) *
##     no_full_baths + log(squarefeet) * no_full_baths + log(distance) *
##     log(squarefeet) * no_full_baths
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1    100 1.8051
## 2     96 1.6614  4   0.14373 2.0763 0.08986 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

a) $log(\widehat{adj}2007) = 5.41777 +$ -$0.04883log(\widehat{distance}) + 0.59328log(\widehat{square}feet) + .05667no\_\widehat{full}\_baths$. All of the rates are statistically significant. For every additional unit of distance, the log of home price decreases approximately on the average by .04883 after adjusting/controlling for log(squarefeet) and number of full bathrooms. For every additional unit of the log of squarefootage, the home price increases approximately on the average by .59328 after adjusting/controlling for log(distance) and number of full baths. For every additional full bathroom, the home price increases approximately on the average by .05667 after adjusting/controlling for log(distance) and log(squarefeet).

b) The residual plot passes the linearity test because the data seems evenly scattered around zero and there is not a huge curve in the data show in the Residuals vs Fitted graph. There is also no thickening so the data passes the equal variance condition. The Q-Q plot is nearly straight so it also passes the normality test.

c) The only rates that were statistically significant were the log(distance) and log(squarefeet):no_full_baths interaction rate. This is quite different from part A because all rates in part A were statistically significant. $log(\widehat{adj}2007) = 5.545207 + 0.355179log(\widehat{square}feet) + 0.172022log(square\widehat{feet}) : no_full_baths$

d) The p-value from the nested F-test is .08986 which is not statistically significant. Thus this indicates that the more complex model from part B did not add significantly to the simple model from part A. The R-squared value also barely changes from .7834 to .8007 which is another indicator that the more complex model is not necessarily beneficial.

3.53 First-year GPA

```
first = read.csv("C:/Users/adhri/OneDrive/Documents/R/App_Reg_and_Time_Series/chpt3/datasets/FirstYearG
attach(first)

modelF = lm(GPA ~ HSGPA + SATV + HU + White)
summary(modelF)
```

```
##
## Call:
## lm(formula = GPA ~ HSGPA + SATV + HU + White)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.06370 -0.26286  0.02436  0.27338  0.87190
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.6409767  0.2787933   2.299  0.02246 *
## HSGPA       0.4761952  0.0710947   6.698 1.83e-10 ***
## SATV        0.0007372  0.0003417   2.157  0.03209 *
## HU          0.0150566  0.0036383   4.138 5.03e-05 ***
## White       0.2121164  0.0686196   3.091  0.00226 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3824 on 214 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.3375, Adjusted R-squared:  0.3251
## F-statistic: 27.25 on 4 and 214 DF,  p-value: < 2.2e-16
```

```
# b
#Confidence Interval for a Prediction
newF = data.frame(HSGPA = 3.2, SATV = 600, HU = 10, White = 1)
predict(modelF, newF, se.fit=T, interval = "prediction", level = .95)
```

```
## $fit
##        fit      lwr      upr
## 1 2.969829 2.212739 3.726919
##
## $se.fit
## [1] 0.03607865
##
## $df
## [1] 214
##
## $residual.scale
## [1] 0.3823948
```

```
# c
modelFS = lm(GPA ~ HSGPA + SATV + HU + White + SS)
summary(modelFS)
```

```
##
## Call:
## lm(formula = GPA ~ HSGPA + SATV + HU + White + SS)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.08660 -0.25827  0.04326  0.25822  0.87954
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.5684876  0.2827454   2.011  0.04563 *
## HSGPA       0.4739983  0.0709413   6.682 2.03e-10 ***
## SATV        0.0007481  0.0003410   2.194  0.02932 *
## HU          0.0167447  0.0038183   4.385 1.82e-05 ***
## White       0.2060408  0.0685881   3.004  0.00298 **
## SS          0.0077474  0.0054401   1.424  0.15587
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3815 on 213 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.3437, Adjusted R-squared:  0.3283
## F-statistic: 22.31 on 5 and 213 DF,  p-value: < 2.2e-16
```

```r
# d
#Confidence Interval for a Prediction
newFS = data.frame(HSGPA = 3.2, SATV = 600, HU = 10, White = 1, SS = 10)
predict(modelFS, newFS, se.fit=T, interval = "prediction", level = .95)
```

```
## $fit
##       fit      lwr      upr
## 1 2.98512 2.229526 3.740715
##
## $se.fit
## [1] 0.0375597
##
## $df
## [1] 213
##
## $residual.scale
## [1] 0.3814795
```

a) The predicted GPA is 2.969829.

b) The 95% prediction interval for the GPA of this student is 2.212739 to 3.726919 dollars.

c) The predicted GPA would be 2.98512. The 95% prediction interval for the GPA of this student would be 2.229526 to 3.740715 dollars.

3.54 Combining explanatory variables

a) $Y = X_2 + 3$. There is a positive association between X_2 and Y.

b) Y = -.5X_1 + 2X_2 + 1. No they are not in the direction because the signs of X_1 and X_2 are different. The sign of X_1 in its original equation was positive, whereas the sign of X_1 in the multivariate equation is negative. The sign of X_2 is the same in both the simple and multivariate equations.

3.55 Porsche versus Jaguar prices

```
cars = read.csv("C:/Users/adhri/OneDrive/Documents/R/App_Reg_and_Time_Series/chpt3/datasets/PorscheJagua
attach(cars)

# splitting data based on porsche or jaguar
porsche = subset(cars,Porsche %in% "1")
jaguar = subset(cars,Porsche %in% "0")


# models
modelPMA = lm(porsche$Price ~ porsche$Mileage + porsche$Age)
modelJMA = lm(jaguar$Price ~ jaguar$Mileage + jaguar$Age)
summary(modelPMA)
```

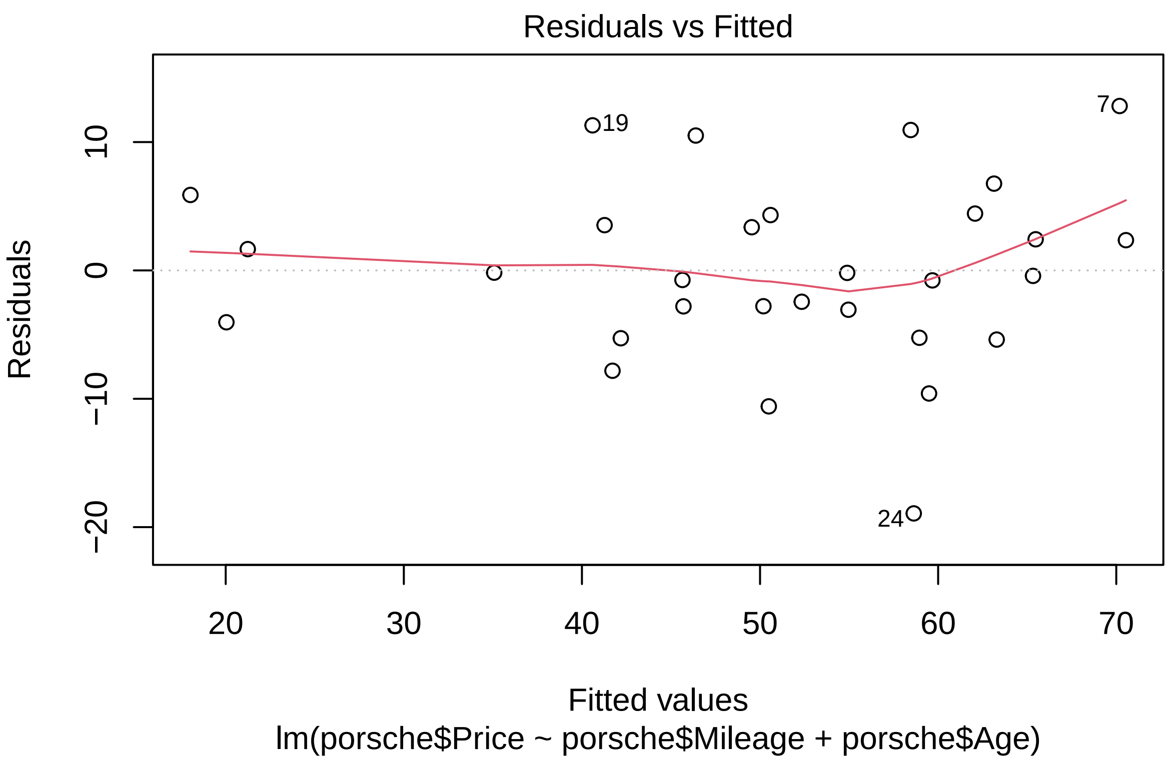
```
##
## Call:
## lm(formula = porsche$Price ~ porsche$Mileage + porsche$Age)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -18.930  -3.795  -0.309   4.116  12.811
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       70.9192     2.4835  28.556  < 2e-16 ***
## porsche$Mileage  -0.5613     0.1141  -4.921 3.76e-05 ***
## porsche$Age      -0.1302     0.4568  -0.285    0.778
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.291 on 27 degrees of freedom
## Multiple R-squared:  0.7951, Adjusted R-squared:  0.7799
## F-statistic: 52.39 on 2 and 27 DF,  p-value: 5.073e-10
```

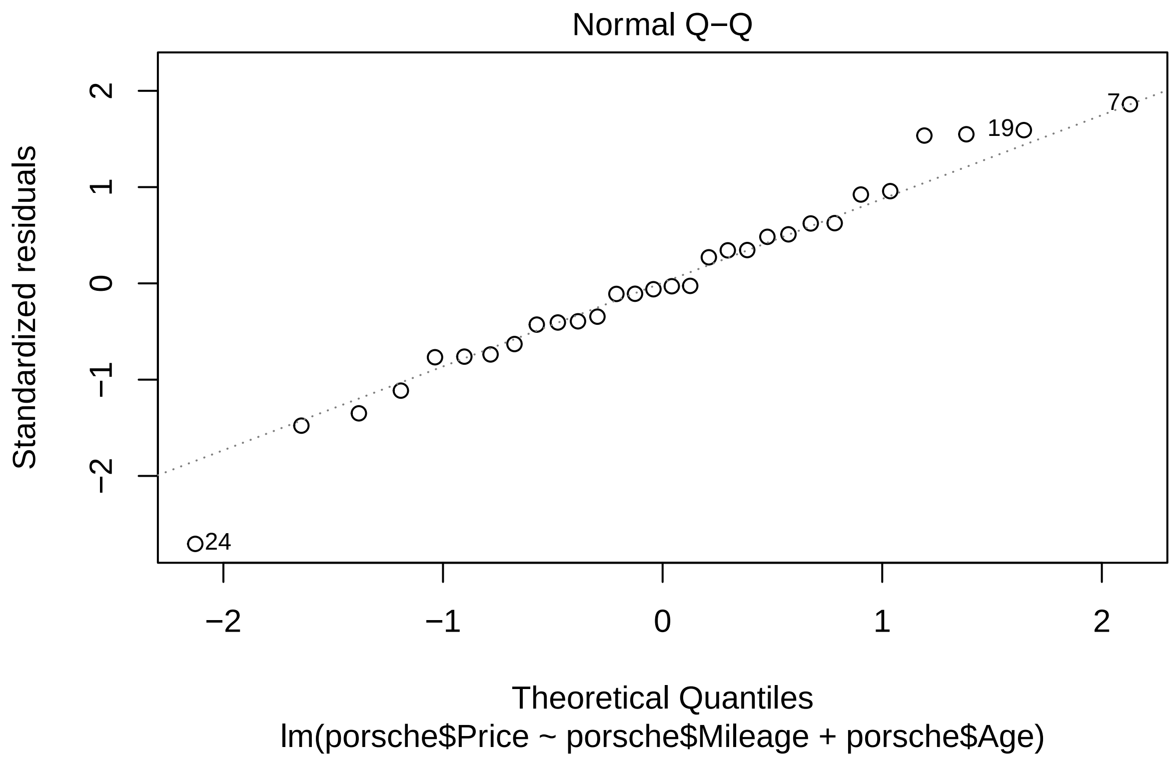
```
summary(modelJMA)
```

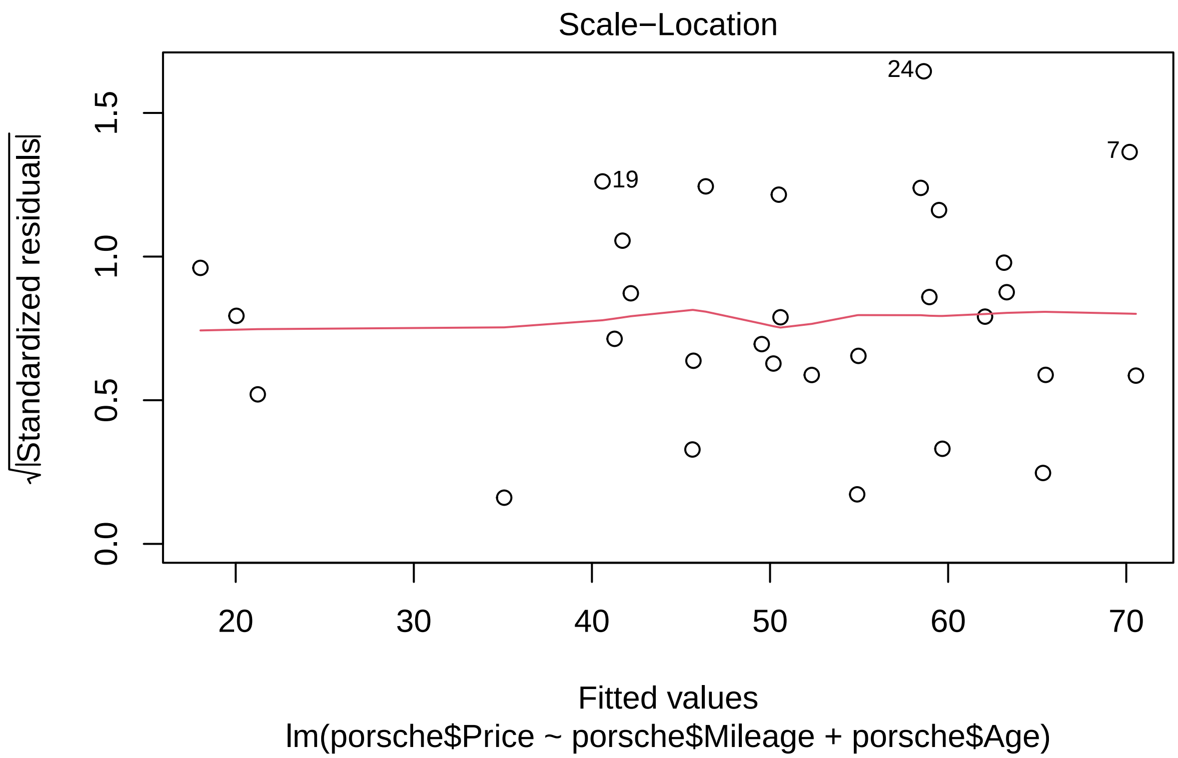
```
##
## Call:
## lm(formula = jaguar$Price ~ jaguar$Mileage + jaguar$Age)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -19.3836  -7.3345   0.6728   7.8354  18.3551
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
```

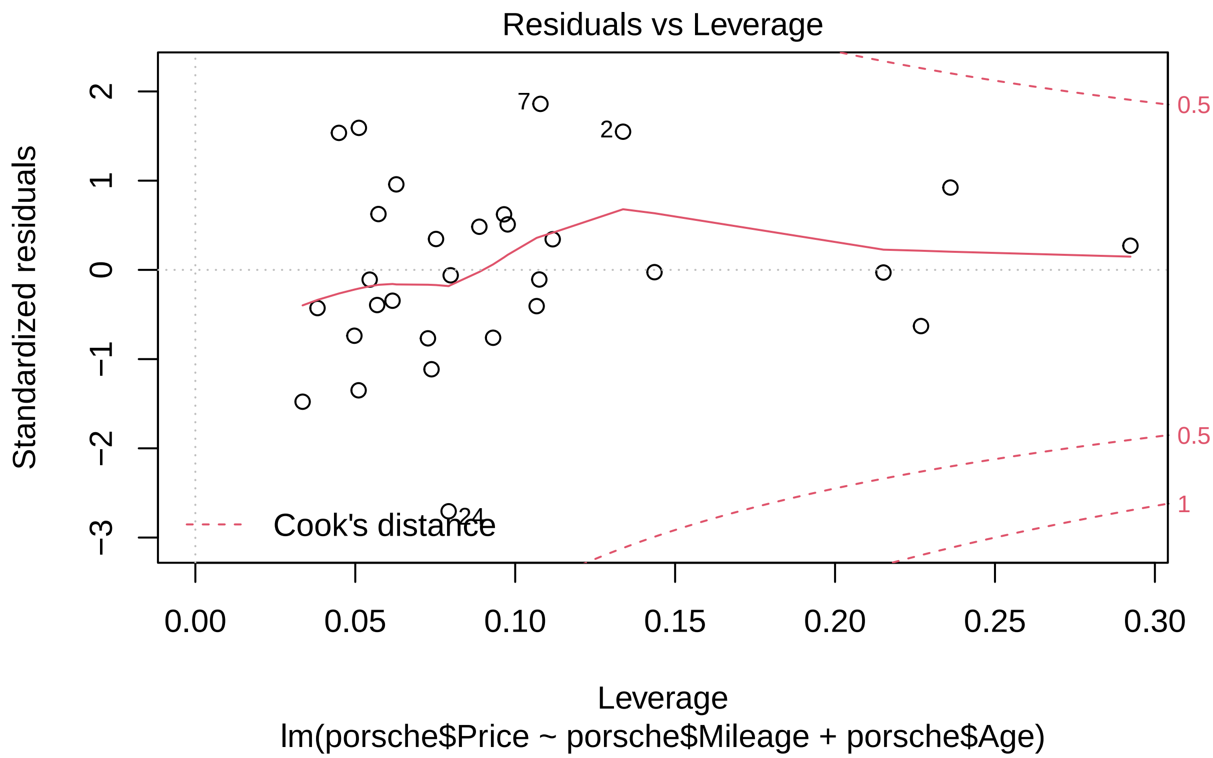
```
## (Intercept)     57.8865     4.6015  12.580 8.37e-13 ***
## jaguar$Mileage  -0.4394     0.1513  -2.904  0.00727 **
## jaguar$Age      -2.0442     1.3284  -1.539  0.13550
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.68 on 27 degrees of freedom
## Multiple R-squared:  0.6235, Adjusted R-squared:  0.5956
## F-statistic: 22.35 on 2 and 27 DF,  p-value: 1.878e-06
```

```r
# plots
plot(modelPMA)
```

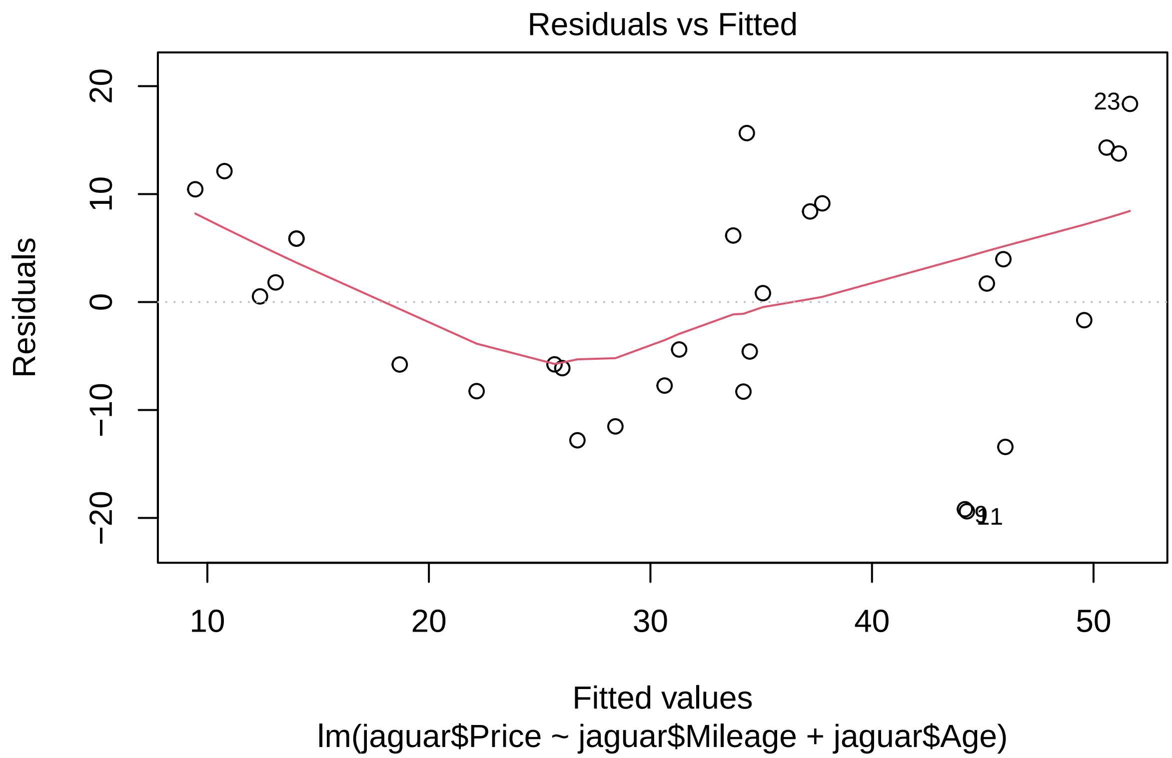


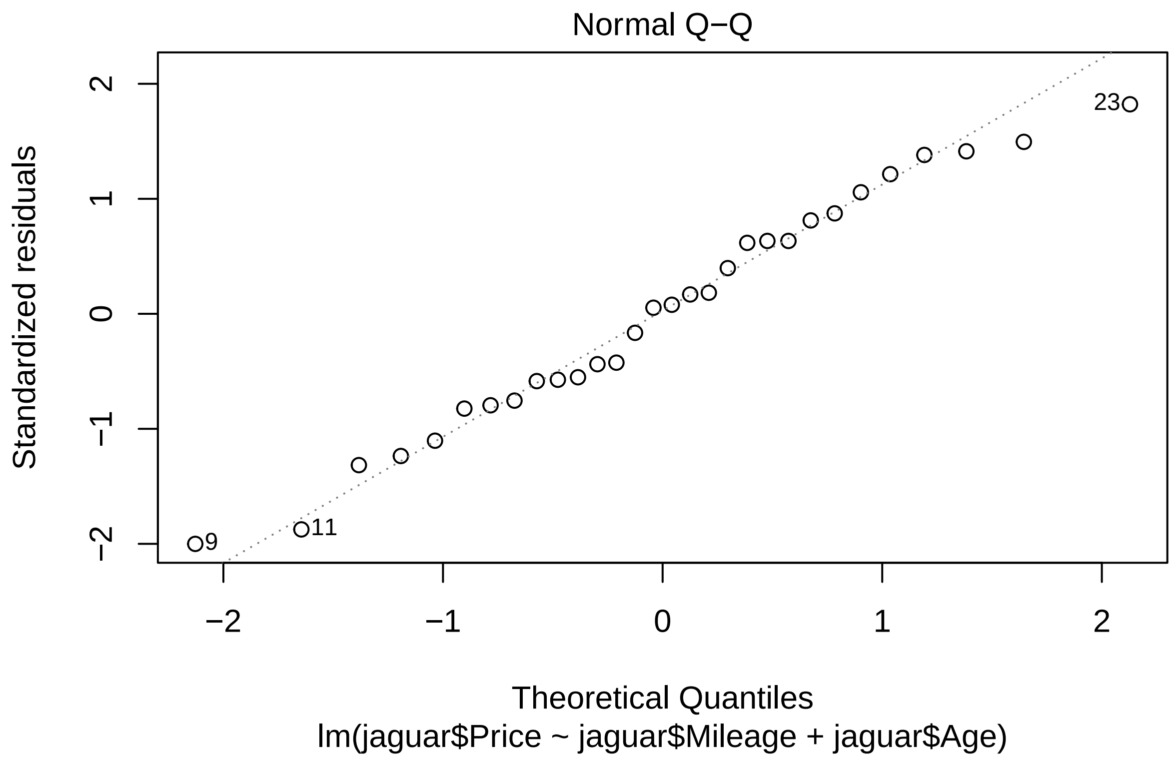

Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(porsche$Price ~ porsche$Mileage + porsche$Age)
2
1
0
−1
−2
−2
−1
0
1
2
7
19
24


Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
20
30
40
50
60
70
Fitted values
lm(porsche$Price ~ porsche$Mileage + porsche$Age)
24
7
19

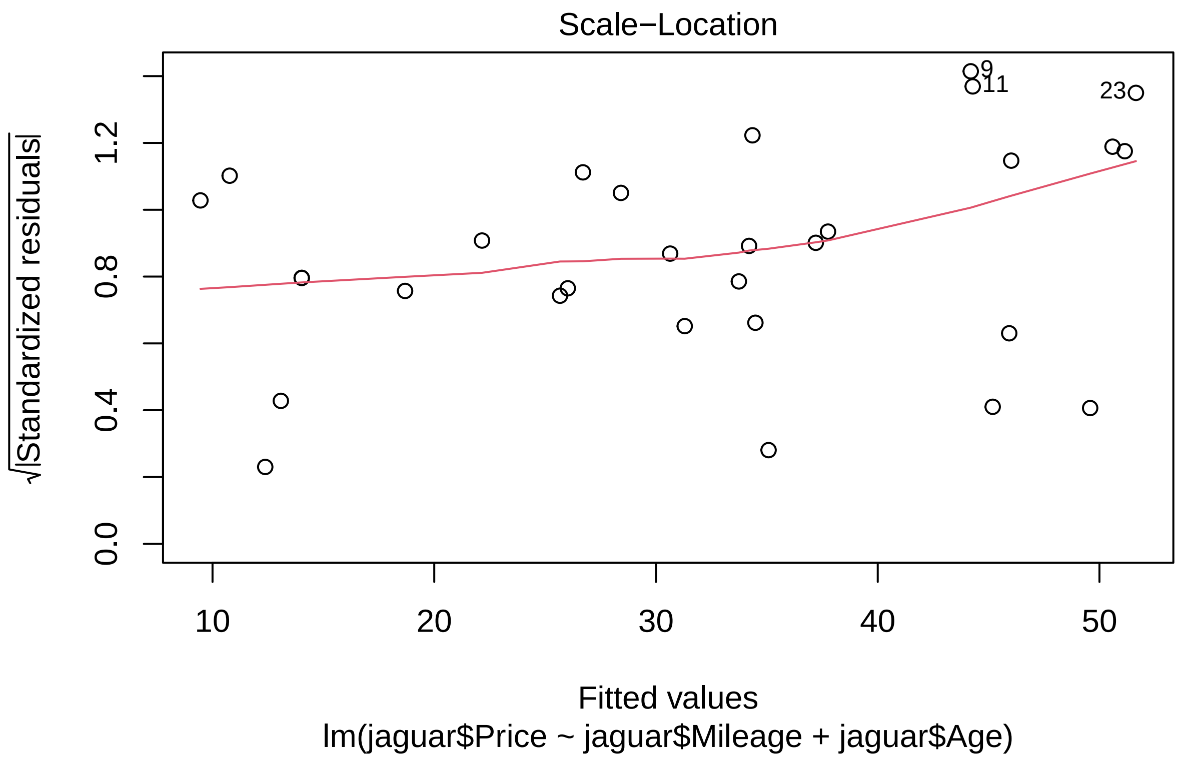


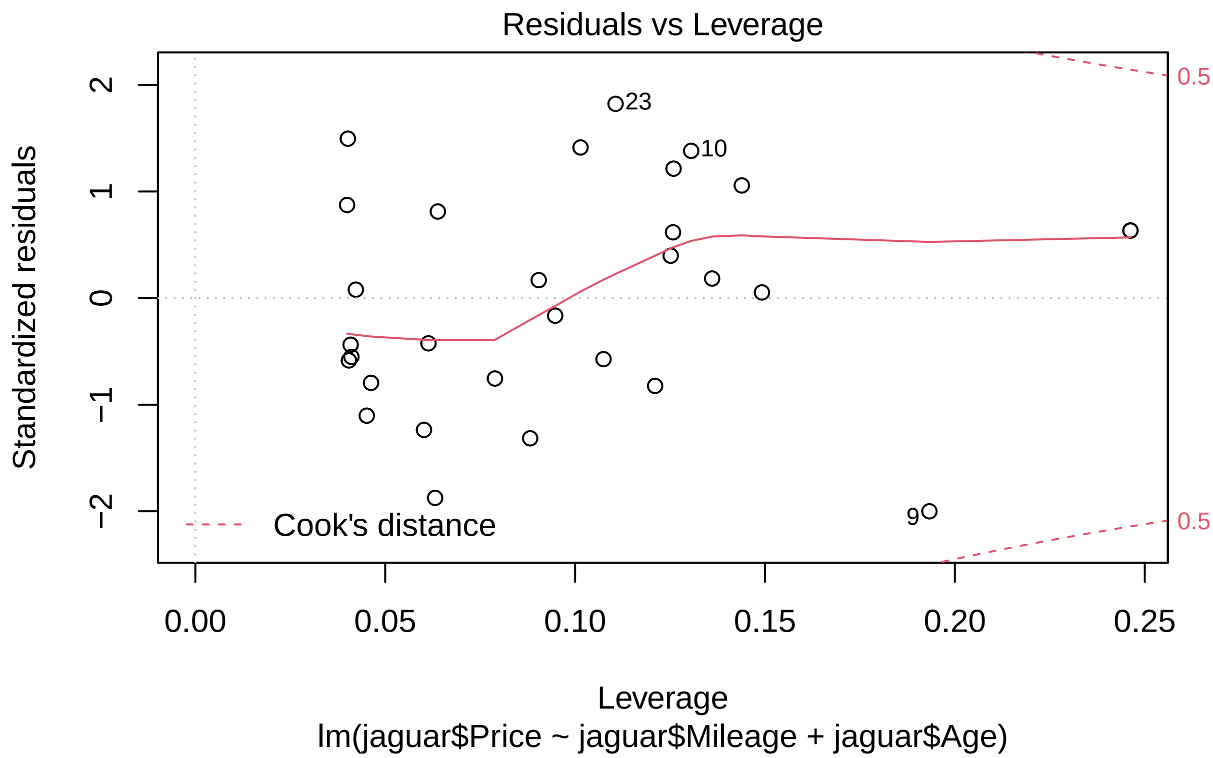

```
plot(modelJMA)
```


Residuals vs Fitted
Residuals
20
10
0
−10
−20
10
20
30
40
50
Fitted values
lm(jaguar$Price ~ jaguar$Mileage + jaguar$Age)
23
9
11



Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(jaguar$Price ~ jaguar$Mileage + jaguar$Age)
−2
−1
0
1
2
9
11
23


Scale−Location
√|Standardized residuals|
0.0
0.4
0.8
1.2
10
20
30
40
50
9
11
23
Fitted values
lm(jaguar$Price ~ jaguar$Mileage + jaguar$Age)




```
# nested F test
modelPM = lm(porsche$Price ~ porsche$Mileage)
modelJM = lm(jaguar$Price ~ jaguar$Mileage)
anova(modelPM, modelPMA)
```

```
## Analysis of Variance Table
##
## Model 1: porsche$Price ~ porsche$Mileage
## Model 2: porsche$Price ~ porsche$Mileage + porsche$Age
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     28 1439.6
## 2     27 1435.2  1    4.3194 0.0813 0.7778
```

```
anova(modelJM, modelJMA)
```

```
## Analysis of Variance Table
##
## Model 1: jaguar$Price ~ jaguar$Mileage
## Model 2: jaguar$Price ~ jaguar$Mileage + jaguar$Age
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     28 3351.5
## 2     27 3081.3  1    270.22 2.3678 0.1355
```

Both porsche and jaguar residual plots passed the linearity and equal variance tests. The residual plot for jaguar had a slight curve to it, but it still passes linearity. They both also pass the normality condition because their Q-Q plots appear to be straight. In both multiple linear regression models, the car Age variable had p-values of over .05 and were thus not statistically significant. Both Mileage variables were statistically significant. If we reduce the models to just mileage as the explanatory variables and run a nested F-test, both p-value are above .05 thus indicating that Age is not a beneficial variable in both models. The nature of the price versus mileage is similar for the two types of cars because in both cases, every additional mile reduces the price of the car by a similar value. Porsches tend to be naturally more expensive, starting at an intercept of 70.9192, wherease Jaguars' intercept is at 57.8865. Porsches also depreciate in prices more quickly because for every additional unit of mileage, the price of a Porsche decreases by -0.5613. In Jaguars, for every additional unit of mileage, the price of the Jaguar decreases by -0.4394.