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
```

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

#t.test(Year*StateControl, FatalityRate)

anova(modelYS, modelYSYS)
```

```
## Analysis of Variance Table
##
## Model 1: FatalityRate ~ Year + StateControl
## Model 2: FatalityRate ~ Year + StateControl + Year * StateControl
##   Res.Df     RSS Df Sum of Sq      F    Pr(>F)
## 1     18 0.25434
## 2     17 0.03061  1   0.22373 124.26 3.082e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

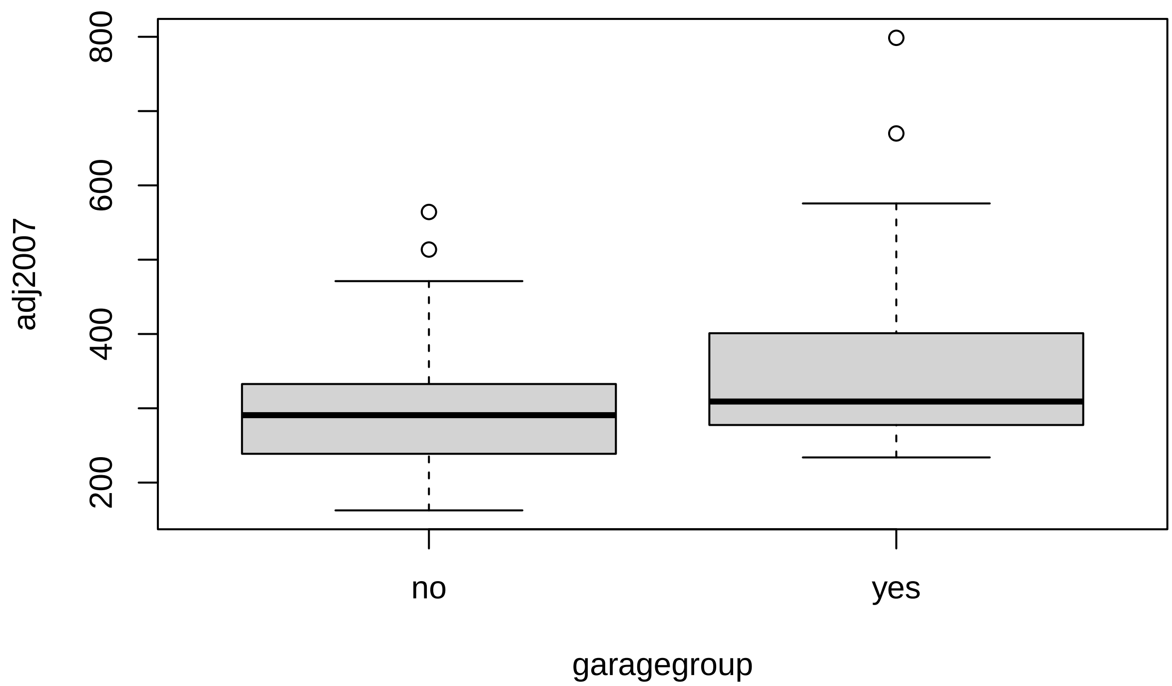
a) H_0: $B_2 = B_3$; H_1: at least one of B_2 or B_3 is not equal to 0. There is a significant difference between the slopes of the two lines because the F-statistic is 62.939 with a p-value of 1.386e-08 which is less than alpha which gives us evidence that there is a significant difference between the slope and intercept of the two lines.

b) H_0: $B_3 = 0$; H_1: B_3 does not equal 0. The F-statistic is 124.26 with a p-value of 3.082e-09 which is less than alpha which indicates that we reject the null hypothesis and conclude that B_3 does not equal 0. There sufficient evidence to suggest there is difference between slopes. The t-test produced a t-value of 11.15 with a p-value of 3.08e-09 which is similar to the nested F-test.

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
modelDGDG = lm(adj2007 ~ distance + garagegroup + distance*garagegroup)
anova(modelD, modelDGDG)
```

```
## Analysis of Variance Table
##
## Model 1: adj2007 ~ distance
## Model 2: adj2007 ~ distance + garagegroup + distance * garagegroup
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1    102 865718
## 2    100 827301  2     38417 2.3218 0.1034
```

a) The t-value of -2.7145 and p-value of .007896 being less than alpha indicates there is a significant difference between the price of a home based on if it has garage spaces or not. The boxplot also indicates price of homes with garage space are higher.

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

c) For every additional unit of distance, the home price decreases approximately on the average by 51.025 thousands of dollars 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 thousands of dollars after adjusting/controlling for distance.

d) The rates are $\widehat{adj2007} =$ 407.945 - 56.18$distance$ and $\widehat{adj2007} =$ 359.083 - 46.3$distance$. The difference in rates is statistically significant because the p-value of Distance:garagegroupyes is greater than alpha and is not statistically significant and cannot be considered in the model. Thus, the first rate of -56.18$distance$ is not accurate because that equation considers Distance:garagegroupyes in the equation.

e) H_0: B_2 and $B_3 = 0$; H_1: B_2 or B_3 does not equal 0. The F statistic is 2.3218 and the p-value is 0.1034. Thus we fail to reject the null hypothesis and can say that terms involving garage space do not add significantly to the model of price on distance.

3.49 Real estate near Rails to Trails: transformation

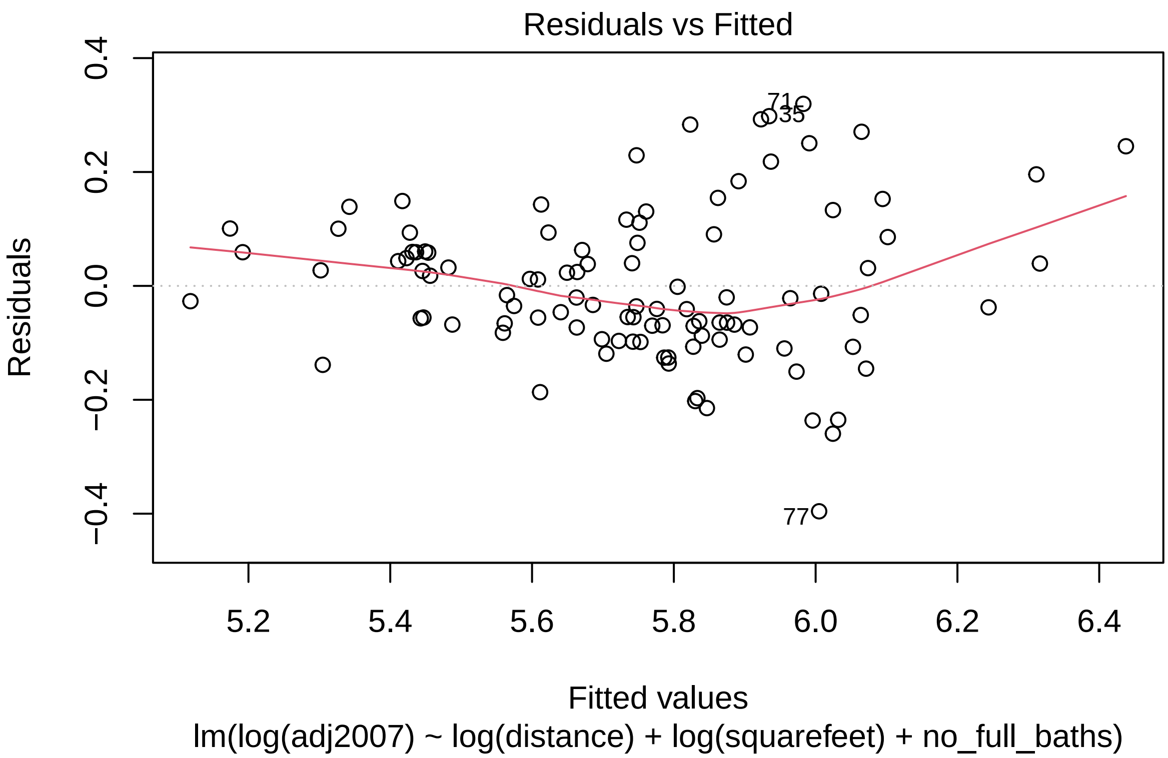
```
rail = read.csv("C:/Users/adhri/OneDrive/Documents/R/App_Reg_and_Time_Series/chpt3/datasets/RailsTrails
attach(rail)
```

```
## 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
```

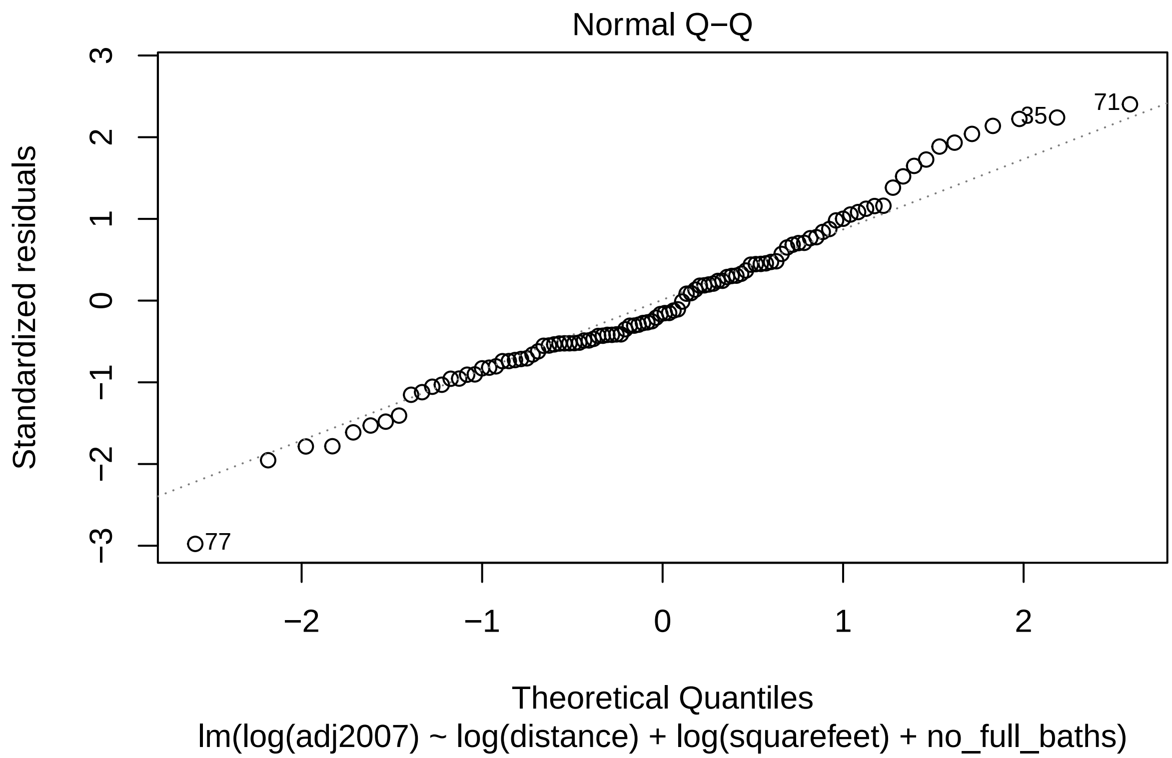
```r
# a
modelDSN = lm(log(adj2007) ~ log(distance) + log(squarefeet) + no_full_baths)
summary(modelDSN)
```

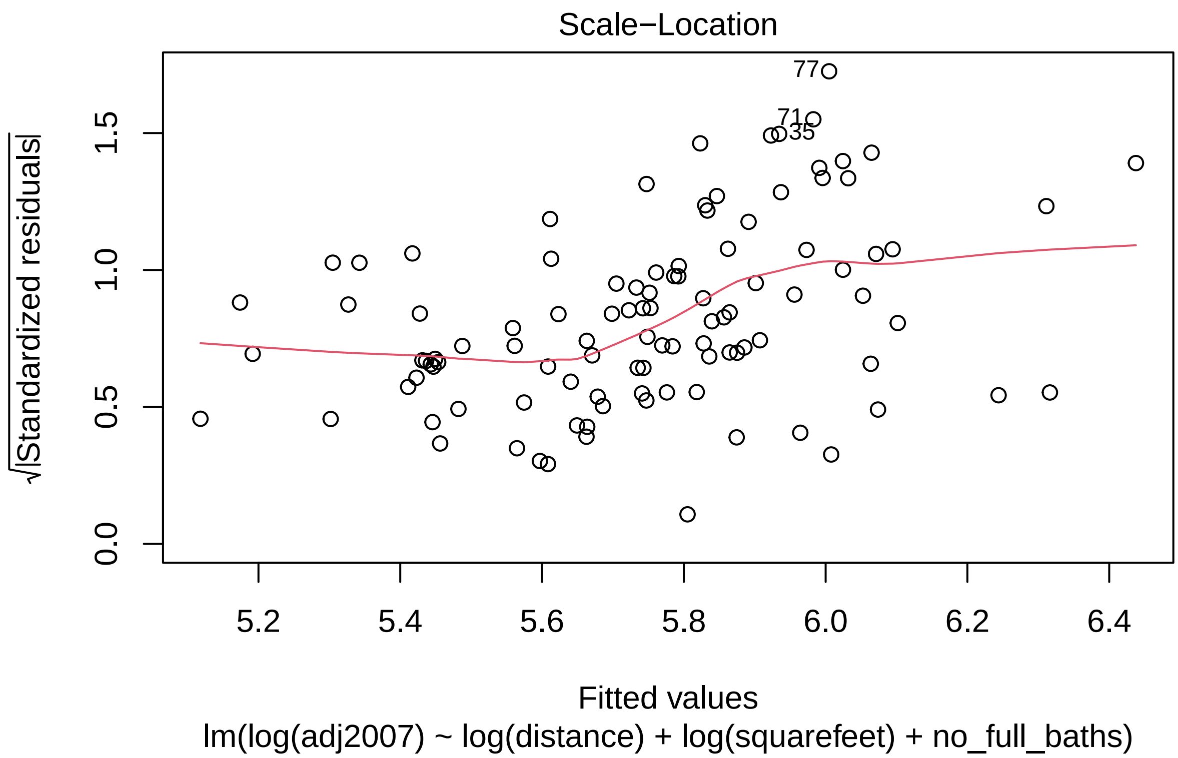
```
##
## Call:
## lm(formula = log(adj2007) ~ log(distance) + log(squarefeet) +
##     no_full_baths)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.39580 -0.07536 -0.02103  0.07813  0.31959
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      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
```

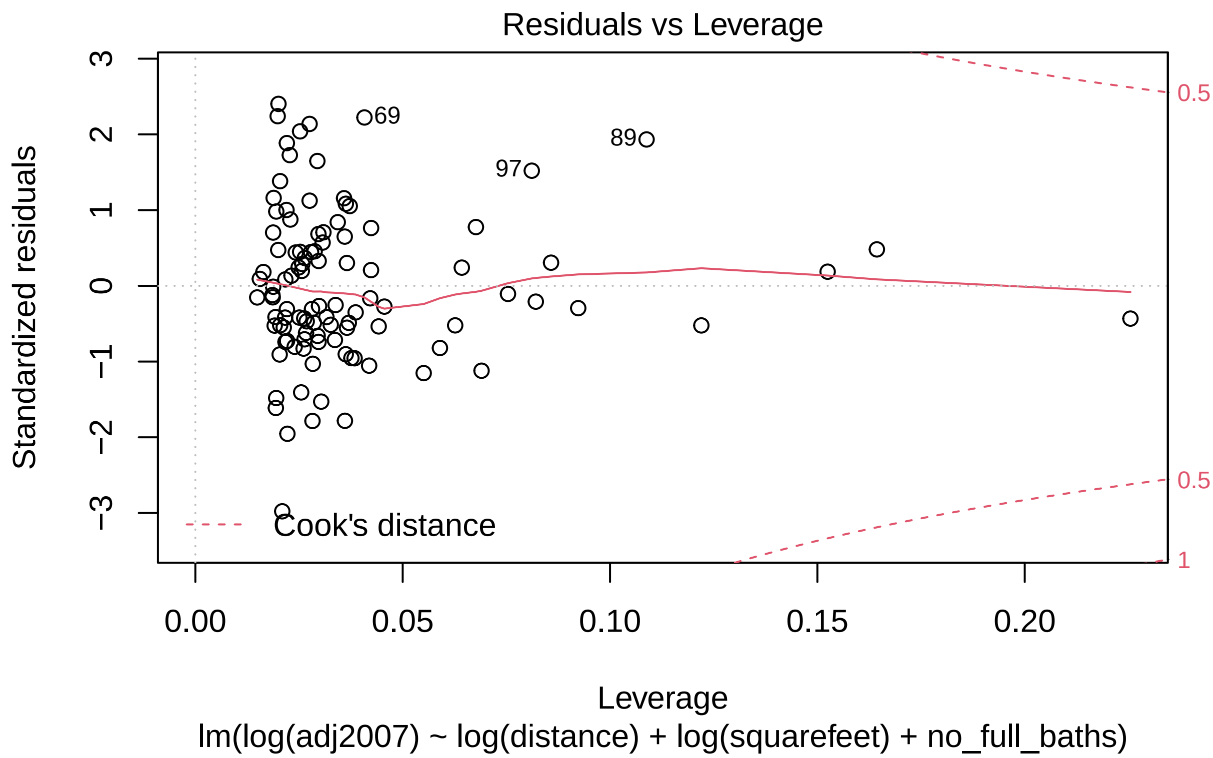
```r
# b
plot(modelDSN)
```



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
−2
−1
0
1
2


Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
5.2
5.4
5.6
5.8
6.0
6.2
6.4
77
71
35
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) $\widehat{log(adj2007)}$ = 5.41777 + -0.04883$\widehat{log(distance)}$ + 0.59328$\widehat{log(squarefeet)}$ + .05667$\widehat{no_full_baths}$. All of the rates are statistically significant. Residual standard errors are low which indicates a good fit. 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. The R-squared value also changes from .7834 to .8007. $\widehat{log(adj2007)} = 5.545207 + 0.355179\widehat{log(squarefeet)} + 0.172022\widehat{log(squarefeet) : no_full_baths}$.

d) H_0: B_5 or B_6 or B_7 or B_8 = 0; H_1: B_5 or B_6 or B_7 or B_8 does not equal 0. 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/FirstYearGl
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
```

```
# 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 direction 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
modelPM = lm(porsche$Price ~ porsche$Mileage)
modelJM = lm(jaguar$Price ~ jaguar$Mileage)
summary(modelPM)
```

```
##
## Call:
## lm(formula = porsche$Price ~ porsche$Mileage)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -19.3077  -4.0470  -0.3945   3.8374  12.6758
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)     71.09045    2.36986    30.0  < 2e-16 ***
## porsche$Mileage -0.58940    0.05665   -10.4 3.98e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.17 on 28 degrees of freedom
## Multiple R-squared:  0.7945, Adjusted R-squared:  0.7872
## F-statistic: 108.3 on 1 and 28 DF,  p-value: 3.982e-11
```

```
summary(modelJM)
```

```
##
## Call:
## lm(formula = jaguar$Price ~ jaguar$Mileage)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -24.3271  -6.6043  -0.8191   8.6924  18.8120
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    54.22746    4.03456  13.441 9.81e-14 ***
## jaguar$Mileage -0.62030    0.09763  -6.353 7.11e-07 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.94 on 28 degrees of freedom
## Multiple R-squared:  0.5904, Adjusted R-squared:  0.5758
## F-statistic: 40.36 on 1 and 28 DF,  p-value: 7.111e-07
```

```
# models
modelJ = lm(Price ~ Mileage + Porsche + Mileage*Porsche)
summary(modelJ)
```

```
##
## Call:
## lm(formula = Price ~ Mileage + Porsche + Mileage * Porsche)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -24.3271  -5.6200  -0.4819   5.0278  18.8120
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)     54.22746    3.41097  15.898  < 2e-16 ***
## Mileage         -0.62030    0.08254  -7.515 4.88e-10 ***
## Porsche         16.86299    4.58044   3.682 0.000523 ***
## Mileage:Porsche  0.03090    0.11024   0.280 0.780302
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.25 on 56 degrees of freedom
## Multiple R-squared:  0.7648, Adjusted R-squared:  0.7522
## F-statistic: 60.68 on 3 and 56 DF,  p-value: < 2.2e-16
```

```
modelP = lm(Price ~ Mileage + Porsche)
summary(modelP)
```

```
##
## Call:
## lm(formula = Price ~ Mileage + Porsche)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -23.8420  -5.6835  -0.5041   5.4209  19.3490
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 53.60555    2.56951  20.862  < 2e-16 ***
## Mileage     -0.60298    0.05427 -11.111 6.89e-16 ***
## Porsche     17.95833    2.36951   7.579 3.45e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.174 on 57 degrees of freedom
## Multiple R-squared:  0.7644, Adjusted R-squared:  0.7562
## F-statistic: 92.48 on 2 and 57 DF,  p-value: < 2.2e-16
```

```r
modelM = lm(Price ~ Mileage)
summary(modelM)
```

```
##
## Call:
## lm(formula = Price ~ Mileage)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -33.088 -10.372   2.868   9.233  20.868
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 62.92843    3.16885  19.858  < 2e-16 ***
## Mileage     -0.61269    0.07621  -8.039 5.26e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.89 on 58 degrees of freedom
## Multiple R-squared:  0.527,  Adjusted R-squared:  0.5189
## F-statistic: 64.63 on 1 and 58 DF,  p-value: 5.263e-11
```

```r
# nested F test
anova(modelM, modelJ) # mileage vs full
```

```
## Analysis of Variance Table
##
## Model 1: Price ~ Mileage
## Model 2: Price ~ Mileage + Porsche + Mileage * Porsche
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     58 9632.6
## 2     56 4791.0  2    4841.5 28.295 3.215e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova(modelP, modelJ) # mileage and porsche vs full
```

```
## Analysis of Variance Table
##
## Model 1: Price ~ Mileage + Porsche
## Model 2: Price ~ Mileage + Porsche + Mileage * Porsche
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     57 4797.8
## 2     56 4791.0  1    6.7205 0.0786 0.7803
```

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 71.09045, whereas Jaguars' intercept is at 54.22746. Porsches also depreciate in prices less quickly because for every additional unit of mileage, the price of a Porsche decreases by -0.58940. In Jaguars, for every additional unit of mileage, the price of the Jaguar decreases by -0.62030. The p-value of the reduced model of just mileage is less than .05, so we reject null hypothesis and conclude that the type of car has statistically significant difference by introducing the interaction term or Porsche term. The 2nd anova is interaction between mileage and type of car. We fail to reject null because the p-value is large, so the interaction term does not produce a statistically significant difference. This tells us that we can't say one car depreciates more than the other because the interaction between Mileage and Porsche is not statistically significant.