Problem set 4

Exercises to hand in: 2.12, 2.16, 2.22, 2.46, 2.61

For this (and all other homework assignments), I am expecting a mixture of text, code, and output.

As you work, I suggested rendering your document frequently to see if you encounter errors. You need to upload the rendered PDF document of your finished homework. This means you should:

- Render your document and look at the preview to make sure it looks good
- Go to the Files tab of your RStudio and find the file that ends in .pdf (for this assignment, probably problemset4.pdf)
- Upload the PDF file to Gradescope

2.12 Inference for slope, again

a. Test the hypothesis that beta1 is 0

```
5.3/2.8
```

```
[1] 1.892857
```

```
2*pt(1.892857, df = (82-2))
```

```
[1] 1.938007
```

```
# use R as a calculator
```

b. Construct a 95% confidence interval for beta1

```
5.3+1.892857*2.8
```

```
[1] 10.6
```

```
5.3-1.892857*2.8
```

```
[1] 4e-07
```

```
# use R as a calculator
```

2.16 Textbook prices

```
data(TextPrices)
```

a. Perform a significance test

```
m3<-lm(Price~Pages,data=TextPrices)
summary(m3)
```

```
Call:
lm(formula = Price ~ Pages, data = TextPrices)

Residuals:
    Min      1Q  Median      3Q     Max
-65.475 -12.324  -0.584  15.304  72.991

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -3.42231   10.46374  -0.327    0.746
Pages        0.14733    0.01925   7.653 2.45e-08 ***
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 29.76 on 28 degrees of freedom
Multiple R-squared:  0.6766,    Adjusted R-squared:  0.665
F-statistic: 58.57 on 1 and 28 DF,  p-value: 2.452e-08
```

```
anova(m3)
```

```
Analysis of Variance Table

Response: Price
          Df Sum Sq Mean Sq F value    Pr(>F)
Pages      1  51877   51877  58.573 2.452e-08 ***
Residuals 28  24799     886
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
0.14733/0.01925
```

```
[1] 7.653506
```

```
pt(7.653506, df=(30-2))
```

```
[1] 1
```

b. Find and interpret a 95% confidence interval

```
confint(m3)
```

```
                 2.5 %    97.5 %
(Intercept) -24.8563229 18.011694
Pages         0.1078959  0.186761
```

2.22 Partitioning variability, again

```
38+64
```

```
[1] 102
```

```
38/102
```

```
[1] 0.372549
```

```
# use R as a calculator
```

R-Squared =0.372549

2.46 Real estate near Rails to Trails: home size transformation

```
data(RailsTrails)
```

a. Use a simple linear regression model

Adj2007hat= 72.97 +162.53 * SquareFoot

```
rt<-lm(Adj2007 ~ SquareFeet, data=RailsTrails)
162.53 * 1.5
```

```
[1] 243.795
```

```
243.795 +72.97
```

```
[1] 316.765
```

b. Give a 95% prediction interval

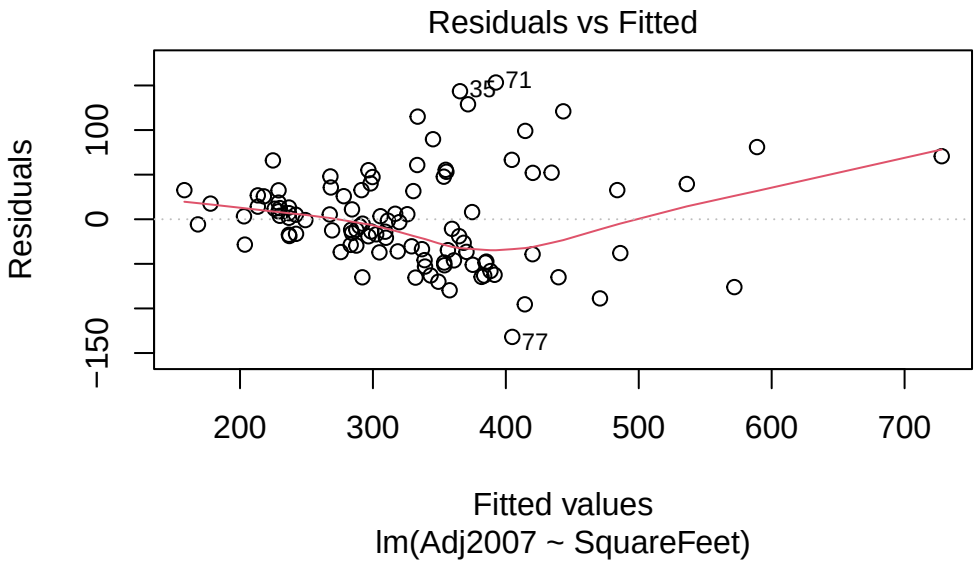
```
predict(rt,newdata=data.frame(SquareFeet =1.5),interval ="prediction")
```

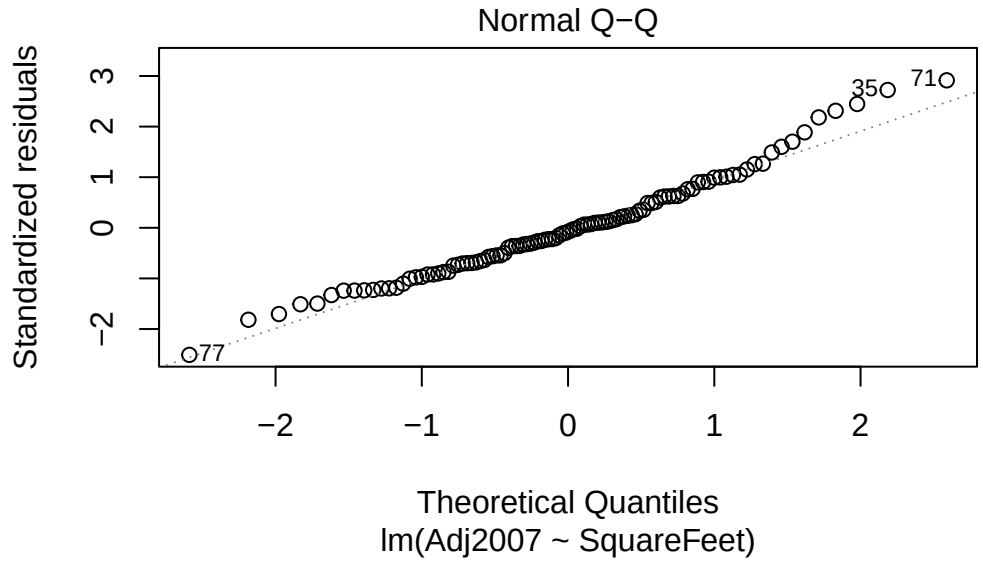
```
       fit      lwr      upr
1 316.7619 211.1232 422.4006
```

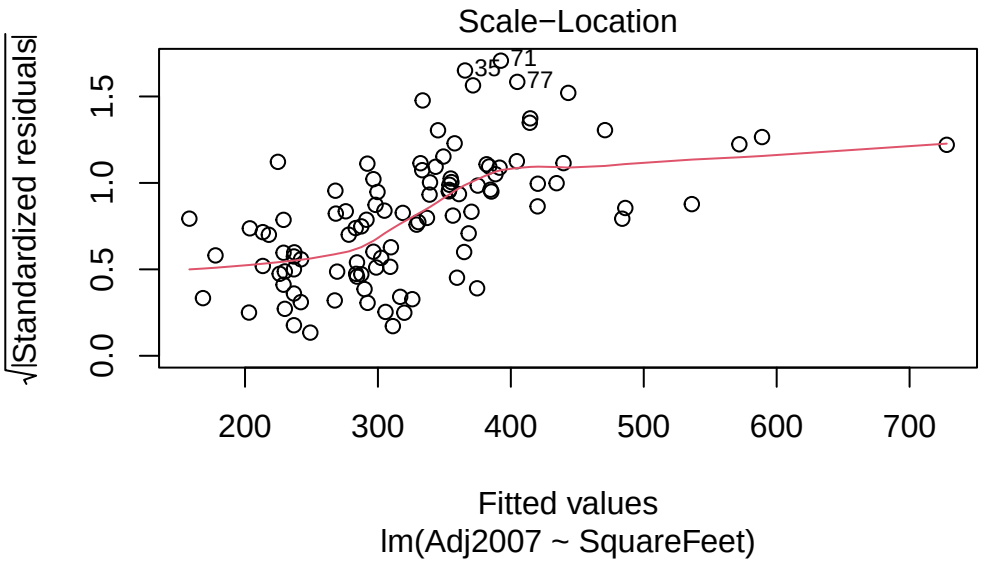
We are predicting that there is a 95% probability that there is a value between 211.1232 and 422.4006

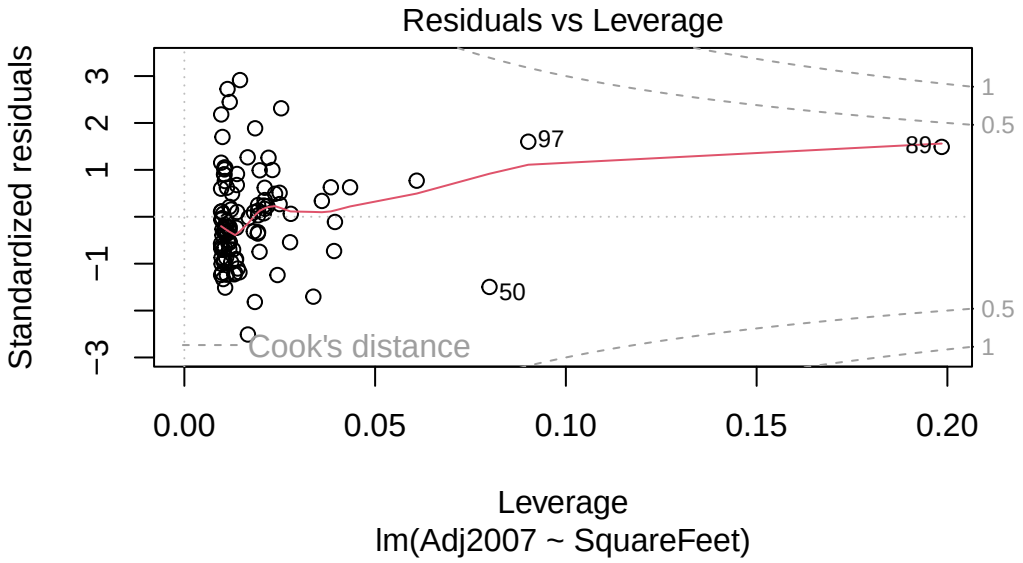
c. Comment on adherence to the model conditions and any effect on answer (b)

```
plot(rt)
```




Normal Q−Q
Standardized residuals
−2
0
1
2
3
77
35
71
−2
−1
0
1
2
Theoretical Quantiles
lm(Adj2007 ~ SquareFeet)


Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
35
71
77
200
300
400
500
600
700
Fitted values
lm(Adj2007 ~ SquareFeet)



The linearity plot and the normality of the graph seems to be normal along with independence and the equality of variance appears to be slightly skewed. There is no outliers spotted from the graph.

d. Redo the regression using log

```
RailsTrails<- RailsTrails %>%
  mutate(transformed_SquareFeet= log(SquareFeet),transformed_Adj2007=log(Adj2007))
rr<- lm(transformed_Adj2007~transformed_SquareFeet, data = RailsTrails)
```

e. Redo the prediction interval

We are predicting that there is 95% probability that there is a value between 5.469522 and 6.051882.

```
predict(rr, newdata=data.frame(transformed_SquareFeet =log(1.5)),interval ="prediction")
```

```
       fit      lwr      upr
1 5.760702 5.469522 6.051882
```

2.61 Gate count

(No dataset)

a. Find the equation of the least squares line for predicting the gate count from enrollment.

GateCountHat= 247235 + 88.75701 * Enrollment

```
0.701 *(254116/2007)
```

```
[1] 88.75701
```

```
247235-88.75701*2009
```

```
[1] 68922.17
```

```
# use R as a calculator
```

b. What percentage of the variation in the gate counts is explained by enrollments?

0.8% of gate counts are explained by enrollments.

```
2009 +247235
```

```
[1] 249244
```

```
2009/249244
```

```
[1] 0.008060375
```

```
# use R as a calculator
```

c. Predict the number of persons who will use the library at a small liberal arts college with an enrollment of 1445.

We predict that 375488.9 persons will use the libary at a small liberal arts college with an enrollment of 1445.

```
247235+88.75701*1445
```

```
[1] 375488.9
```

```
# use R as a calculator
```

d. One of the reporting colleges has an enrollment of 2200 and a gate count of 130,000. Find the value of the residual for this college.

We predict that 442500.4 grate count with an enrollment of 2200, when the actual gate count was 130000. This makes the residual for this college -312500.4.

```
247235 + (88.75701 *2200)
```

```
[1] 442500.4
```

```
130000-442500.4
```

```
[1] -312500.4
```

```
# use R as a calculator
```