Problem set 5

Exercises to hand in: 3.2, 3.18

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
- Close the preview document
- Go to the Files tab of your RStudio and find the file that ends in .pdf (for this assignment, probably problemset5.pdf)
- Upload the PDF file to Gradescope

3.2 Breakfast cereals

a. How many calories would you predict for a breakfast cereal that had 1 gram of fiber and 11 grams of sugar per serving?

CaloriesHat =109.3 + 1*Sugar-3.7 *Fiber

```
109.3+1.0*11-3.7*1
```

```
[1] 116.6
```

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

There will be about 116.6 calories found a breakfast cereal that has 1 gram of fiber and 11 gram of sugar per serving.

b. Compute the residual for Frosted Flakes and explain what this value means.

```
110-116.6
```

```
[1] -6.6
```

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

The residual for the foster flakes is -6.6, and this means that the observed calories are less than the predicted value.

3.18 Real estate near Rails-to-Trails

```
data("RailsTrails")
```

a. Fit the simple linear regression model

```
m1=lm(Adj2007~Distance, data=RailsTrails)
summary(m1)
```

```
Call:
lm(formula = Adj2007 ~ Distance, data = RailsTrails)

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
```

b. Fit a regression model with both explanatory variables

```
m2=lm(Adj2007 ~ Distance + SquareFeet, data=RailsTrails)
summary(m2)
```

```
Call:
lm(formula = Adj2007 ~ Distance + SquareFeet, data = RailsTrails)

Residuals:
     Min       1Q   Median       3Q      Max
-138.835  -32.621   -1.903   27.369  145.504

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  109.742     20.057   5.472 3.25e-07 ***
Distance     -16.486      5.942  -2.775  0.00659 **
SquareFeet   150.780      9.998  15.080  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 51.34 on 101 degrees of freedom
Multiple R-squared:  0.7655,    Adjusted R-squared:  0.7608
F-statistic: 164.8 on 2 and 101 DF,  p-value: < 2.2e-16
```

The addition of square Feet has changed the estimate between distance and Adj2007. Because the adjusted R square for the linear model that was between Adj2007 and distance was 0.2299, after adding the Square Feet to the model the value increased to 0.7608. There is also a huge difference in the estimated coefficients for distance. In the first model, the estimated coefficient was about -54.427, but in the second model, the value changed to -16.486. This means that in the first model we are losing about $54,000 for every mile, but in the second model, we are only losing $16,000. Therefore adding Square Feet to the model makes difference, and is the best model to use.

c. Find 95% confidence intervals

```
confint(m1)
```

```
              2.5 %    97.5 %
```

```
(Intercept) 360.3317 416.07588
Distance    -73.5859 -35.26851
```

```
confint(m2)
```

```
                2.5 %     97.5 %
(Intercept)  69.95460 149.530197
Distance    -28.27307  -4.698861
SquareFeet  130.94601 170.614247
```

The confident interval for the Adj2007(Response Variable, that in the simple linear regression model(m1)) will be in the range of 41.42 and 47.32. And the distance(explanatory variable,(m1)) will be in the range from -9.74 to -6.87. However for a regression model with both explanatory variables (m2), the confidence interval for the Adj2007 would be from -4.43 to -4.26, and the distance will be from 0.35 to 0.4(rounded value).

d. Give a prediction

```
109.742-(16.486*0.5)+(150.780*1.5)
```

```
[1] 327.669
```

```
# use R as a calculator or the predict() function without an interval =
```

We are predicting the price for a particular house that has 1500 square feet of space and is 0.5 miles from a bike trail to be about $327.669.