**Shay Snyder**

**Data Analytics**

**Exercise 5**

**Problem #1:**

A close up of a piece of paper

Description automatically generated

**Problem #2:**

A close up of a piece of paper

Description automatically generated

**Problem #3:**

A screenshot of a cell phone

Description automatically generated

**Problem #4:**

*Simple Model*

A screenshot of a cell phone

Description automatically generated

F-statistic = 99.59 on 1 and 198 DF

P-value = < 2.2e-16

**Problem #5**

*Simple Model*

*A screenshot of a cell phone

Description automatically generated*

The low p-value of our regression model means we can use the F-statistic to further evaluate our model to determine if the number of adverts is a significant variable. Given the low p-value and the high F-statistic, we can reject the null hypothesis given its application on 1 and 198 DF. In conclusion, these values suggest that the adverts attribute is quite significant.

**Problem #6**

*Simple Model*

*A screenshot of a cell phone

Description automatically generated*

Coefficient for Intercept: 134.1

Coefficient for adverts: 0.09612

Line of Best Fit: y = 0.09612x + 134.1

**Problem #7**

According to our model, approximately 13,110 records will be sold if $135,000 is

invested into adverts.

Y = 0.09612x + 134.1

x = 135000

y = 0.09612(135000) + 134.1

y = approx: 13,110

**Problem #8**

*Complex Model*

*A screenshot of a cell phone

Description automatically generated*

F-statistic: 129.5 on 3 and 196 DF

P-value: = < 2.2e-16

**Problem #9**

Given the fact that higher R-squared values are better because more variance is explained by the model, we know the complex (multiple) regression model is better than the simple model. Thus, adding the other attributes allowed the more complex model to better fit our data.

Simple Model R-squared (adjusted): 0.3313

Complex Model R-squared (adjusted): 0.6595