1. Which of the four predictors is the most effective at predicting MPG? You can answer this naively by regressing with each predictor alone and noting the R2 value of the model (variance explained by the model). (<https://en.wikipedia.org/wiki/Coefficient_of_determination>)

Intercept 35.120636

Displacement -0.060051

R2 = 0.648

Intercept 39.935861

Horsepower -0.157845

R2 = 0.606

Intercept 46.216525

Weight -0.007647

R2 = 0.693

Intercept 4.833250

Acceleration 1.197624

R2 = 0.179

From these output, it is proven that Weight, with the largest R2 value of 0.693, would be the most effective at predicting MPG.

1. In the single-predictor model of the best predictor, what is the coefficient on that predictor? How should you interpret it? (<http://www.theanalysisfactor.com/interpreting-regression-coefficients>)

The coefficient of Weight is -0.007647. That means that a decrease of Weight of a car is associated with an increase in the MPG of that car.

1. In the linear regression model containing all four predictors, what is its coefficient of determination? How should you interpret that?

Intercept 45.251140

Displacement -0.006001

Horsepower -0.043608

Weight -0.005281

Acceleration -0.023148

R2 = 0.707

Hence, the coefficient of determination is 0.707. The coefficient is higher when we use multiple features to assess linear regression, therefore this method is a better fit to the data as opposed to using single-predictor model.

1. If we wanted to include the cylinders information from the dataset into our model, could we still use linear regression? If so, how would we do it? Show the code for your approach.

Yes, we can add cylinders information into our model and still use linear regression. Simply add the column into our functions:

sns.pairplot(data, x\_vars=[‘Cylinders’,'Displacement','Horsepower','Weight','Acceleration'], y\_vars='MPG', size=7, aspect=0.7, kind = 'reg')

lm = smf.ols(formula='MPG ~ Cylinders + Displacement + Horsepower + Weight + Acceleration', data=data).fit()