Seth Emery
CS5565-05
HW2 (Linear Regression)
Feb 12, 2024

**2. Use the summarize() function to print the results. Comment on the output. For instance:**

A) **Is there a relationship between the predictors and the response? Use the p-values to answer this question.**
From the summary, it's clear that the predictors and the response variable have a relationship, because some of the predictors have p-values that indicate a significant contribution to the response variable (< .05). It can be said that the predictors as a whole have a relationship with the response variable and in plain words this means that as the predictors (Cylinders, Displacement, Horsepower, Weight, Acceleration, Year, Origin) change in value, there is a change in the response variable (MPG). Note that not all the predictors have a statistically significant relationship with the response variable, but because some do, the predictors as a whole have a relationship to the response variable.

B) **Which predictors appear to have a statistically significant relationship to the response?**
The predictors Displacement, Weight, Year, and Origin have p-values below .05, and therefore have a statistically significant relationship with the response variable MPG.

The predictors Cylinders, Horsepower, and Acceleration have p-values above .05, and therefore *do not* have a statistically significant relationship with the response variable MPG.

C) **What does the coefficient for the year variable suggest?**
The coefficient for the year variable (.7508) suggests that for each year increase there is a .7508 increase in MPG with all else being equal. This might suggest that newer vehicles are more fuel efficient than older ones. It could be caused by advances in technology, or standards for fuel economy have improved other the years.

**3. Fit some models with interactions as described in the lab. Comment on the output - are any interactions statistically significant?**

A) **Model with interaction between horsepower and weight**
Horsepower and Weight both have negative coefficients without an interaction, which suggests that the higher the Horsepower or Weight, the lower the MPG. So the more powerful or heavy the car, the lower the MPG. When I create a model that looks at the interaction between Horsepower and Weight and their effect on MPG, there is a positive coefficient. The positive coefficient suggests that the negative impact of Horsepower on MPG becomes less of a factor at higher vehicle Weight, and looking at it from the other way, the negative impact of Weight on MPG is mitigated by higher Horsepower to some

extent. The interaction indicates that the relationship between Horsepower (or Weight) and MPG depends on the level of the other variable.

One conclusion you may draw is that for heavier vehicles, increasing horsepower might not reduce fuel efficiency as much as it would for lighter vehicles. It is an interesting interaction. Looking at the statistical significance of all the variables, Horsepower, Weight, and the interaction between the two, all P-values are (< .05) and therefore have statistical significance in regards to their effects on MPG.

**B) Model with interaction between Cylinders and Acceleration that takes into account all predictors**

The first thing to note is that for this model I used all the predictors and the interaction between cylinders and acceleration instead of only focusing on the specific two variables and their interaction like in 3A. This might provide more context and be a better model.

Cylinders and Acceleration both have positive coefficients independently which means that more cylinders or more acceleration in a car is associated with increases in MPG or more fuel efficiency. When I create a model that looks at the interaction between Cylinders and Acceleration and their effect on MPG, there is a negative coefficient. The interaction suggests that as you increase the number of cylinders, the ability of acceleration to counteract the negative effect on MPG diminishes. Or looking at it the other way, as you increase the acceleration of a car, having more cylinders doesn't necessarily equate to the same decrease in MPG as it would in a car with slower acceleration. Essentially, fast acceleration can soften the blow of what having more cylinders does to a car's MPG.

One conclusion you may draw is that for cars with more cylinders, higher acceleration doesn't lead to the same reduction in fuel efficiency as it might in cars with fewer cylinders. This is another interesting interaction. When examining the statistical significance of all the variables, Cylinders, Acceleration, and the interaction between them, all P-values are (< .05), indicating that they are statistically significant concerning their impact on MPG. This interaction is interesting because the conventional downsides of having more cylinders on fuel efficiency change based on how quickly the vehicle can accelerate.

**C) Model with interaction between Cylinders and Acceleration**

‘cylinders’ has a negative coefficient which means that more cylinders in a car is associated with decreases in MPG. ‘acceleration’ has a positive coefficient which means that more acceleration is associated with increases in MPG. When I create a model that looks at the interaction between Cylinders and Acceleration and their effect on MPG, there is a negative coefficient. The interaction suggests that as you increase the number of cylinders, the ability of acceleration to counteract the negative effect on MPG diminishes. Or looking at it the other way, as you increase the acceleration of a car, having more cylinders doesn't necessarily equate to the same decrease in MPG as it would in a car with slower acceleration. Essentially, fast acceleration can soften the blow of what having more cylinders does to a car's MPG. This is the same conclusion made from the previous model that included all predictors.

Just like in the previous model, one conclusion you may draw is that for cars with more cylinders, higher acceleration doesn't lead to the same reduction in fuel efficiency as it might in cars with fewer cylinders. This is another interesting interaction.

When examining the statistical significance of all the variables in the model, ‘cylinders’ and ‘acceleration’ have P-values (< .05), indicating that they are statistically significant concerning their impact on MPG. This ‘cylinder’ and ‘acceleration’ interaction has a P-value (> .05) being .055. This means that this is not statistically significant and the evidence isn’t strong enough to reject the null hypothesis. It’s interesting that when you view this interaction not in the context of a model including all predictors, the P-value raises enough to not qualify as statistically significant.

Sources:
https://islp.readthedocs.io/en/latest/datasets/Auto.html
https://www.statsmodels.org/0.6.1/examples/notebooks/generated/interactions_anova.html