Firstname Lastname

Date

Dr. Professor

Econ 0000 - Section 0000

In this project, data is being analyzed to report back to the EPA. They are interested in determining what factors in a car contribute to fuel efficiency because car emissions and fuel consumption can be reduced by making cars more efficient. This analysis is important because it will guide policy towards making more efficient vehicles to reduce carbon emissions.

This dataset is from a 1974 edition of “Motor Trends” magazine. In the dataset, there are thirty-two different types of cars observed and ten factors that measure performance. For this model, we are interested in what factors will best predict fuel efficiency (miles per gallon) out of the ten in the dataset. The factors are listed below as well as the means of the dependent and independent variables of the suggested linear model.

**Variables:**

Names of cars = car\_models

Miles per gallon = mpg

Type of transmission (0=automatic, 1= manual) = am

Weight (1000lbs) = wt

Number of cylinders = cyl

Rear axle ratio= drat

V Shaped (0) vs Straight (1) engine = vs

Number of gears = gear

Number of carburetors = carb

Horsepower = hp

Quarter mile time (seconds) = qsec

**Means for IV and DV:**

Mean\_mpg = 20.090625

Mean\_hp = 146.6875

The suggested model is a decent model. There are significant p values for the B0 and B1 values. Within the confidence intervals for the coefficients, zero is not a possible value. The R-squared is high enough at 0.5892 to indicate good inferential quality. The coefficient for hp is very statistically significant and slightly negative, indicating a negative relationship between mpg and hp. Intuitively, it makes sense that horsepower would be a negative factor in fuel efficiency, since sports cars and trucks typically have low fuel efficiency and high horsepower.

I played around with adding different variables to form a new model, but I found that adding only weight to the previous model resulted in the most significant increase in explaining the variance in the model. Many of the variables would have caused multicollinearity (like quarter mile time) or are irrelevant (like the number of gears). I decided to do a log log transformation to tighten the data. The adjusted r-squared value rose to 0.8748. The slope coefficients, loghp and logwt became very statistically significant. The residual standard error dropped dramatically from 3.863 in the suggested model to 0.1054 in the new model.

I think that the second model is better for describing the fuel efficiency of a given car. Intuitively, it seems like horsepower and weight are the biggest factors in the fuel efficiency of the car; this is backed up by experimenting with log transformations to adjust for nonlinearity in hp which was evident when viewing a scatterplot of hp as the IV and mpg as the DV. The p-values for the coefficients on the second model are significant and the R-squared is significantly higher. The residual standard error is also much smaller than the original simple model, meaning it fits the data much better than the suggested model. I would trust the second model’s prediction because it more comprehensively explains the factors behind fuel efficiency and it is a better model objectively - based on SER, R-squared, and P-values.

To summarize, a model was created to predict the factors that contribute to fuel efficiency. Though the first model was good, adding wt and a log log transformation significantly increased the quality of the model. For the EPA, it is suggested that they introduce regulation to incentivize the production of lighter less powerful cars to increase fuel-efficiency.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Dependent Variable: mpg** | |  |  |  |  |  | |
| **Independent Variables: hp** | |  |  |  |  |  | |
| **Regression Statistics** | |  |  |  |  |  | |
|  | **R2** | **Adj. R2** | **Std.Err.Reg.** | **# Observations** | **# Missing** |  | |
|  | 0.6 | 0.5892 | 3.863 | 30 | 2 |  | |
| **Summary Table** | |  |  |  |  |  | |
| **Variable** | **Coeff** | **Std.Err.** | **t-Stat.** | **P-value** | **Lower95%** | **Upper95%** | |
| Intercept | 30.0989 | 1.634 | 18.421 | 10x2^-16 | 26.762 | 33.44 | |
| hp | -0.07 | 0.01012 | -6.742 | 10x1.79^-7 | -0.0889 | -0.048 | |
| **Dependent Variable: logmpg** | |  |  |  |  |  |
| **Independent Variables: loghp and logwt** | |  |  |  |  |  |
| **Regression Statistics** | |  |  |  |  |  |
|  | **R2** | **Adj. R2** | **Std.Err.Reg.** | **# Observations** | **# Missing** |  |
|  | 0.8829 | 0.8748 | 0.1054 | 29 | 3 |  |
| **Summary Table** | |  |  |  |  |  |
| **Variable** | **Coeff** | **Std.Err.** | **t-Stat.** | **P-value** | **Lower95%** | **Upper95%** |
| Intercept | 4.8346 | 0.2244 | 21.545 | < 2e-16 | 4.3757 | 5.2936 |
| loghp | -0.2553 | 0.05840 | -4.372 | 0.000145 | -0.3747 | -0.1358 |
| logwt | -0.5622 | 0.08742 | -6.432 | 4.9e-07 | -0.7410 | -0.3834 |