Regression Course Proejct

jmtaysom

06/19/2015

Executive summary

Although on average the cars with manual transmissions get better gas milage than those with automatic transmissions it is because those cars have smaller engines and are lighter weight than those that have automatic transmissions. There is no statistical significance to the difference gas milage between the two transmission types.

Introduction

The Motor Trend Car Road Tests dataset "was extracted from the 1974 Motor Trend US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973–74 models)." In this study the question of whether or not manual transmission cars get better gas milage than automatic transmission cars and what is the quantity of that difference if it exists.

Analysis

The intial question can be answered by creating a linear regression comparing cars with automatic transmissions and manual transmissions. The linear model has an intercept of 17.1473684 and a slope of 7.2449393 which translates into automatic cars having an average mpg of 17.1473684 and by switching to a manual you will get a 7.2449393 increase in mpg. This increase is deemed statistically significant. It does not look at any of the other differences between the two cars. This linear model only accounts for 35.9798943 percent of the variation between automatic and manual cars based on the r squared value of the model.

In order to see what other factors might affect the gas milage of vehicles a linear regression was created comparing mpg to all other variables. From that the variable that had the highest p value was identified to see how adding it to the original model would change the outcome of the original findings.

```
## (Intercept)         cyl        disp          hp        drat          wt
##  0.51812440  0.91608738  0.46348865  0.33495531  0.63527790  0.06325215
##        qsec          vs          am        gear        carb
##  0.27394127  0.88142347  0.23398971  0.66520643  0.81217871
```

When looking at the p values above the lowest one identified was wt or weight of the car in 1,000 lbs. So a linear model predicting gas milage by transmission type and weight was created and compared to the original model. Now when the effect of switching from an automatic transmission to a manual transmission is -0.0236152. So the effect is quite small an with a p value of .988 is no longer statistically significant.

Now it is import to check if adding weight to the linear regression was statistically significant. Through the use of an Anova Table there are several things that show the adding weight is important in understanding the effects of transmission type. The first is that the p value for adding weight to the model is 1.86741503808364e-07 signifying that it is statistically import to include wt in a linear regression that is predicting the gas milage of a car. Second looking at the residual sum of squares of the original model 720.8965992 compared to the new model 278.3196972 the decrease indicates that the new model has a better fit than the old model.

In order to veriify that this model is not being overly influenced by outliers a plot of Residuals vs Leverage was created (figure 4). In this plot all of the points fall under the Cook's Distance indicating that none of the points are influencing the model excessively.

Summary

The analysis shows that switching from an average automatic car to an average manual car will increase gas milage by 7.2449393 but that is because that the average manual car is lighter and not that manual transmissions are more fuel efficient than automatic cars.

Appendix

Summary of the linear regression predicting gas milage by transmission type

```
##
## Call:
## lm(formula = mpg ~ am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.3923 -3.0923 -0.2974  3.2439  9.5077
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   17.147      1.125  15.247 1.13e-15 ***
## am             7.245      1.764   4.106 0.000285 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.902 on 30 degrees of freedom
## Multiple R-squared:  0.3598, Adjusted R-squared:  0.3385
## F-statistic: 16.86 on 1 and 30 DF,  p-value: 0.000285
```

Figure 1

Summary of the linear regression predicting gas milage by transmission type and weight

```
##
## Call:
## lm(formula = mpg ~ am + wt, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.5295 -2.3619 -0.1317  1.4025  6.8782
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 37.32155    3.05464  12.218 5.84e-13 ***
## am          -0.02362    1.54565  -0.015    0.988
## wt          -5.35281    0.78824  -6.791 1.87e-07 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.098 on 29 degrees of freedom
## Multiple R-squared:  0.7528, Adjusted R-squared:  0.7358
## F-statistic: 44.17 on 2 and 29 DF,  p-value: 1.579e-09
```

Figure 2

Anova Table comparing the models that include transmission type only vs transmission type and weight.

```
wt.anova
```

Figure 3

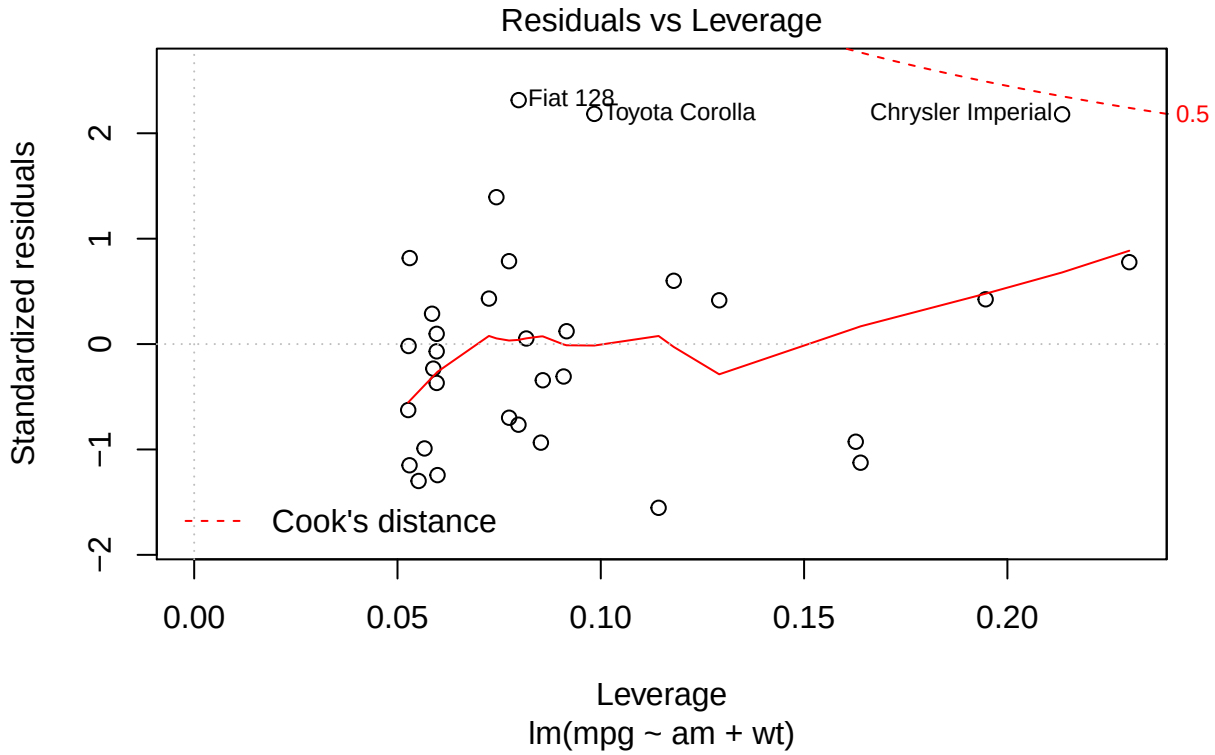

Figure 4