Motor Trend Regression Project

JJD

10/23/2021

Summary

You work for Motor Trend, a magazine about the automobile industry. Looking at a data set of a collection of cars, they are interested in exploring the relationship between a set of variables and miles per gallon (MPG) (outcome). They are particularly interested in the following two questions:

“Is an automatic or manual transmission better for MPG?”

“Quantify the MPG difference between automatic and manual transmissions”

Data

The data 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).

A data frame with 32 observations on 11 (numeric) variables.

[, 1] mpg Miles/(US) gallon

[, 2] cyl Number of cylinders

[, 3] disp Displacement (cu.in.)

[, 4] hp Gross horsepower

[, 5] drat Rear axle ratio

[, 6] wt Weight (1000 lbs)

[, 7] qsec 1/4 mile time

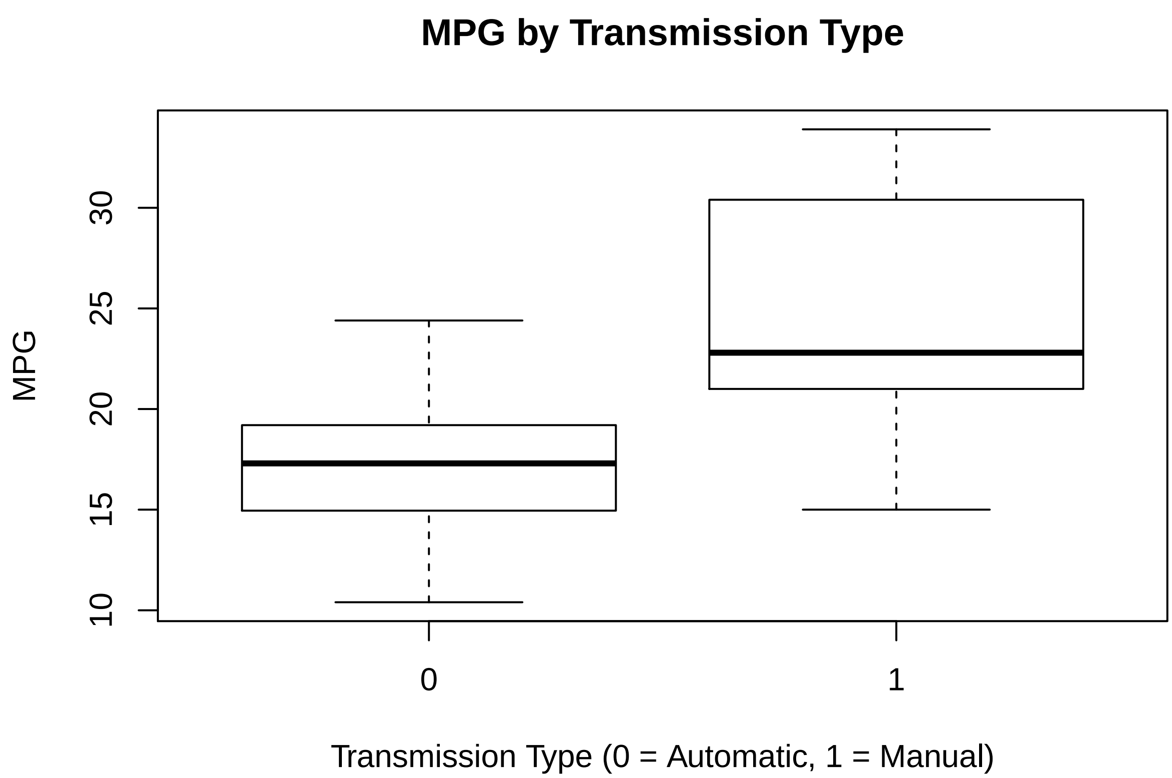
[, 8] vs Engine (0 = V-shaped, 1 = straight)

[, 9] am Transmission (0 = automatic, 1 = manual)

[,10] gear Number of forward gears

[,11] carb Number of carburetors

Exploratory Analysis



At initial glance, manual transmission types have a higher MPG on average compared to automatic transmissions. The significance in difference between manual and automatic transmissions need to be examined using a t-test. The results of the t-test can be found in the appendix.

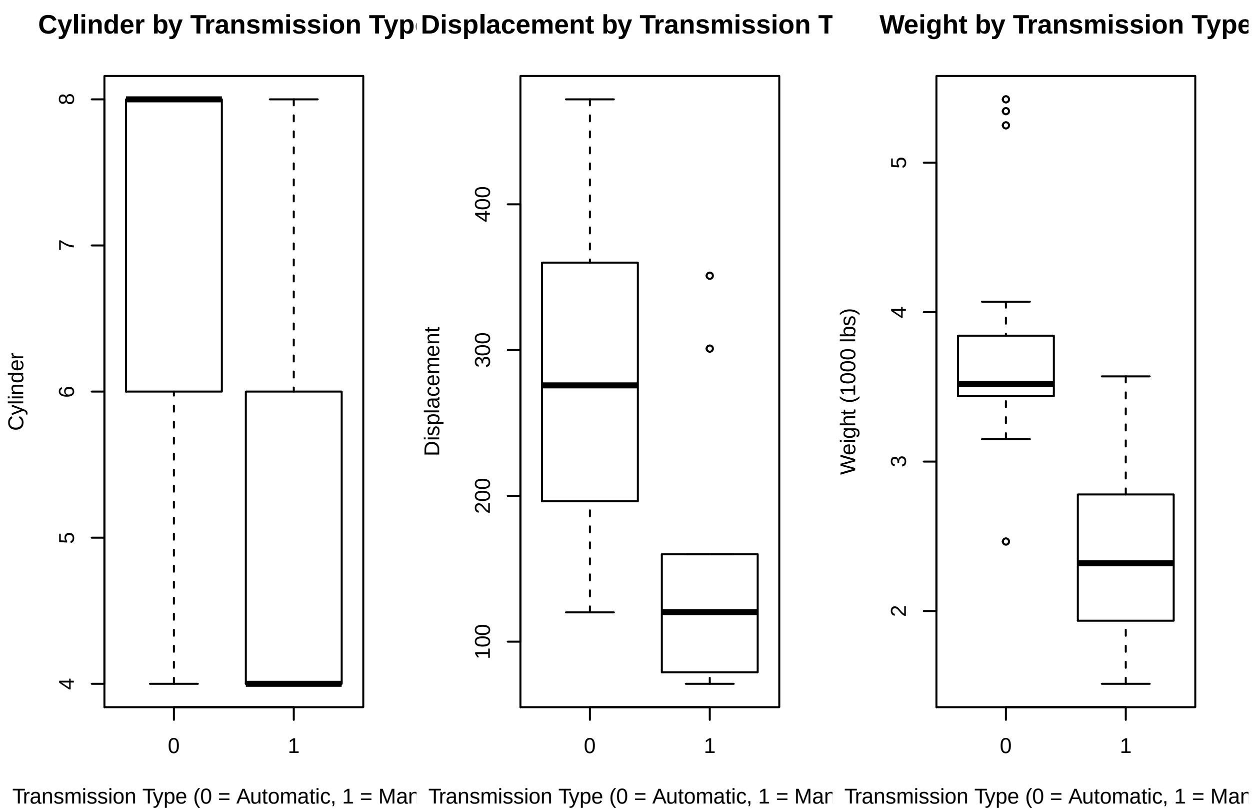
The estimated average of an automatic transmission is 17.15 mpg while the estimated average of a manual transmission is 24.40 mpg. The p-value of the estimation is 0.001374 meaning that there is a significant difference in mpgs between the two transmissions.

To determine the most effective model, an examination on other variables must be done. Regressing all the variables on mpg, size estimates can be gauged to build out an initial model. Results can be found in the appendix.

From the results, weight has the highest effect on mpg. For every 1,000 lbs on a car there is a -3.71530 miles/gallon decrease. Transmission estimates show that a manual transmission increases a car's mpg by 2.52023. Another variable that should be considered is qsec. Qsec has a positive relationship to mpg and for every quarter mile time, mpg increases 0.82. However, the results are not statistically significant.

A review of variables need to conducted in order to build an accurate model. Using a pairs plot will highlight which variables are correlated to mpg. In figure 1 of the appendinx the pairs plot below shows correlation between mpg and transmission is 0.6. Other variables with higher correlations might be more helpful at building out the model. For example, there is a higher negative correlation to mpg with cylinder (0.852), displacement (0.848), and weight (0.868).

Examining the transmission type by the highly correlated values to mpg, it can be inferenced that manual cars overall have a lower weight, displacement and cylinder, which all could be contributing to the higher average mpg we saw earlier. However, a conclusion can't be drawn without further investigation.



Model Selection

1. Univariate model: mpg = 17.147 + 7.245am

From the univariate model automatic transmissions average a 17.1 mpg while manual transmissions can average 7.24 mpg more than the automatic transmission and shows that the estimate is statistically significant at the 99% level. However, the R^2 is only .36 which explains only 36% of the variance. From this result it is apparent that a multivariate model is needed to further explain what other factors may affect mpg.

2. Step model: mpg = 9.6178 - 3.9165wt + 1.2259qsec + 2.9358am

Using the step() function uses AIC for the right to determine the best fitting model. The best fitting model would be lm(formula = mpg ~ wt + qsec + am, data = mtcars) with an AIC at 61.31. In this model it shows that an 1,000 lb increase can lead to a 3.92 decrease in mpg, while a 1/4 second increase in acceleration can lead to a 1.23 increase in mpg and a manual transmisson can average 2.94 mpg than it's automatic counterpart. Through the step function over lm() model, we see that variables like weight and qsec turns out to be good predictors of mpg. Both estimates for weight and qsec are statistically significant at the 99% level while transmission is only at the 90% still shows it's significant in determining mpg.The combination of these variables are able to explain 84% of variation in mpg.

3. Variable selection: mpg = 40.898313 - 1.784173cyl + 0.007404disp - 3.583425wt + 0.129066am

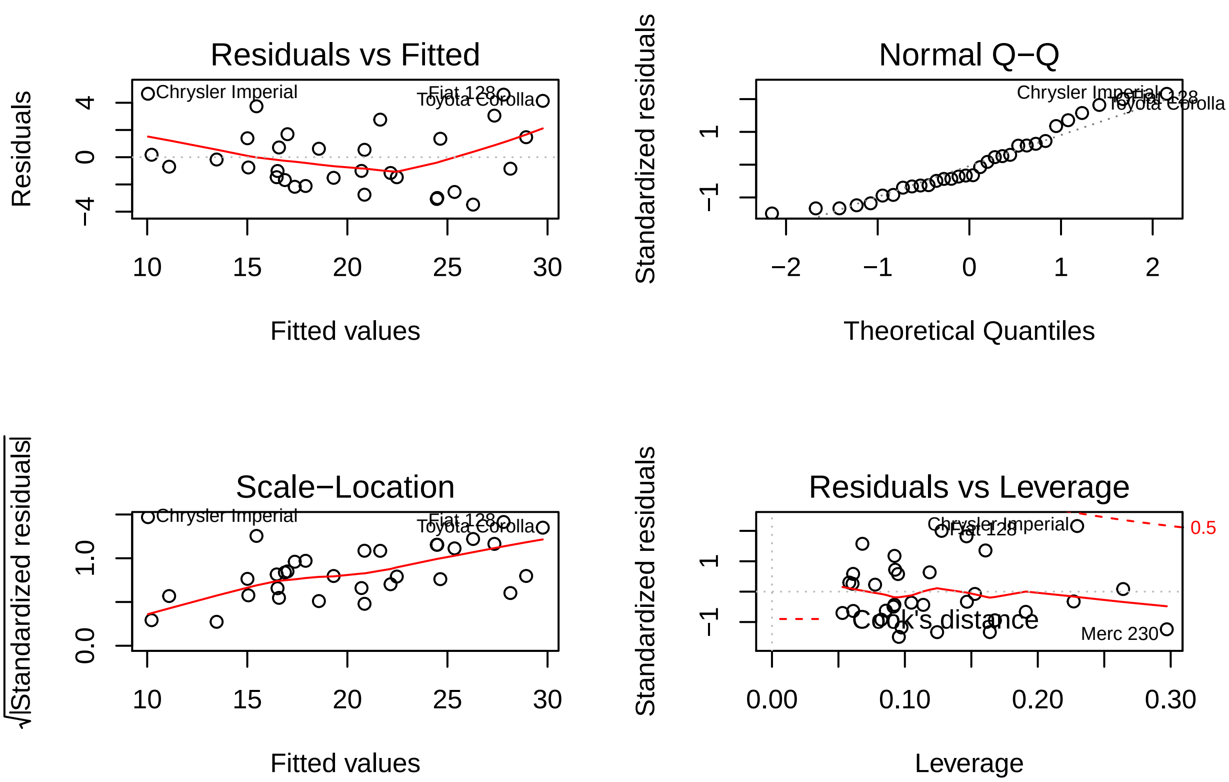
Using the variables that were highly correlated to mpg in this model shows that only cylinder and weight are statistically significant estimates in the model. This model is aligned to the previous model in terms of how weight affects mpg, however there is only a 0.13 increase in mpg on manual transmission cars on automatic. It doesn't make for a good model because the intercept is showing a highstatistical significance and a high estimate, which could mean that there is predictor that's not being accounted for in this model or this model is using the wrong predictors. The model can explain 83% of the variance in mpg.

ANOVA model selection

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ wt + qsec + am
## Model 3: mpg ~ cyl + disp + wt + am
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     30 720.90
## 2     28 169.29  2    551.61 39.521 9.538e-09 ***
## 3     27 188.43  1    -19.14
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Using anova, each model can be tested for level of significance and allow for an accurate estimation on transmission estimate on mpg. In the analysis model 2 (step) shows stronger evidence against the null hypothesis. In this case model2 would be the best fitting model.

Residual Analysis



Using the model determined through the step function will be used to asses residuals and check for heteroskedacity.

Conclusion

It can be concluded that manual transmission is better on mpg than automatic transmission and quantified by a 2.94 increase in mpg. However, in another iteration a deeper exploration on what other variables may impact mpg and if there is multicollinearity in the independent variables.

Appendix

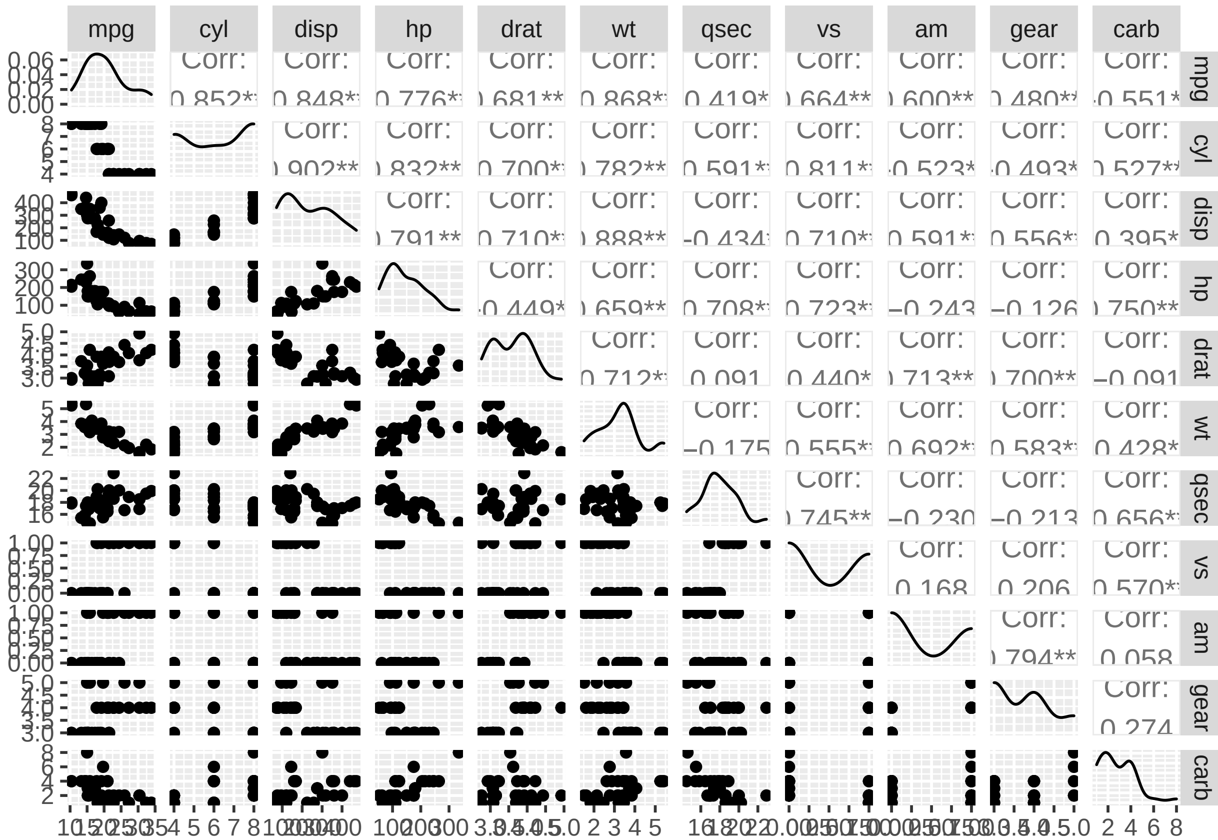
T-test

```
##
##  Welch Two Sample t-test
##
## data:  mtcars[mtcars$am == 0, ]$mpg and mtcars[mtcars$am == 1, ]$mpg
## t = -3.7671, df = 18.332, p-value = 0.001374
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.280194  -3.209684
## sample estimates:
## mean of x mean of y
##  17.14737  24.39231
```

MPG regressed on all variables

```
##
## Call:
## lm(formula = mpg ~ ., data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4506 -1.6044 -0.1196  1.2193  4.6271
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 12.30337   18.71788   0.657   0.5181
## cyl         -0.11144    1.04502  -0.107   0.9161
## disp         0.01334    0.01786   0.747   0.4635
## hp          -0.02148    0.02177  -0.987   0.3350
## drat         0.78711    1.63537   0.481   0.6353
## wt          -3.71530    1.89441  -1.961   0.0633 .
## qsec         0.82104    0.73084   1.123   0.2739
## vs           0.31776    2.10451   0.151   0.8814
## am           2.52023    2.05665   1.225   0.2340
## gear         0.65541    1.49326   0.439   0.6652
## carb        -0.19942    0.82875  -0.241   0.8122
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.65 on 21 degrees of freedom
## Multiple R-squared:  0.869,  Adjusted R-squared:  0.8066
## F-statistic: 13.93 on 10 and 21 DF,  p-value: 3.793e-07
```

Figure 1: Pairs Plot



mpg
cyl
disp
hp
drat
wt
qsec
vs
am
gear
carb