Motor Trends: Effect of car transmission type on milage

rcblt@github.com

Executive Summary

In this brief report we analyzed the dataset `mtcars` from the 1974 Motor Trends magazine to evaluate the effect of the automatic or manual transmission on milage. In particular we were interested to evaluate which transmission is better for milage and quantify that difference. The dataset inclued data for 32 different car models from US and non-US makers.

From an initial exploratory data analysis we showed that there seems to be a significant difference in milage between cars with automatic and manual transmission (Fig. 1). However, a closer look at other variables considered in the dataset showed at least one confounder: car weight (Fig. 2). Car weight appeared to correlate with both milage and transmission type (Fig. 3).

In order to understand the effect of transmission type to milage we started from a multivariable linear model comprising all other variables as predictors and repeatedly removed the least significant one until the mimimal set of significant predictors was obtained.

We ended up with a model that used weight, quarter mile time and transmission as predictors, all being very significant as reported by a simple ANOVA test.

Using the model we could estimate that, keeping all the other variables constant, switching from automatic to manual transmission would yeld a decrease in milage of `4.3` mpg, $CI_{95\%} =$ (`2.2, 6.39`).

Results

Linear models considering car weight, quarter mile time and transmission type.

```
## lm(formula = mpg ~ wt + qsec + am - 1, data = mtcars)
```

Fitting a linear model without intercept to predict milage from car weight (wt), seconds to complete a quarter mile (qsec) and transmission type (am) we derived that the expected increase in mpg between manual and automatic transmission is `4.3` mpg with a 95% confidence interval (`2.2, 6.39`), keeping the other variables constant. All three predictors showed to be very statistically significant with an adjusted $R^2 = 0.9858$.

```
##       Estimate Std. Error   t value     Pr(>|t|)
## wt   -3.185455  0.4827586 -6.598442 3.128844e-07
## qsec  1.599823  0.1021276 15.664944 1.091522e-15
## am    4.299519  1.0241147  4.198279 2.329423e-04
```

The linear model with the same predictor variables and the intercept resulted to have the transmission type barely significant and the intercept itself not significant as well as a lower $R^2 = 0.8336$ (data not shown). Finally the standard R diagnostic plots of the linear model showed no evident bias in the residual plots (Fig. 5) and a distribution rather close to the normal (Fig. 6). However, the standardized residual plot showed that a high variability not explained by the chosen predictors remained (Fig. 7).

Discussion

The initial exploratory data analysis revealed that differences in milage between automatic and manual transmission could be actually explained by confounding variables that correlated with both milage and

transmission type. P-values proved that differences are generally highly significant (Fig. 1 and 2). However, since no formal checks of the condition under which a T-test statistics can be used have been performed, one should consider T-test results more as qualitative than quantitative results. Common sense and exploratory analysis indicated that weight is likely to be one of the confounders between milage and transmission. A heavier car will probably consume more than lighter ones. At the same time heavier, bigger and more expensive cars with larger engines are more likely to be equipped with automatic transmission than smaller cars (Fig. 1, 2 and 3).

In addition to weight other variables could explain the difference in milage. In fact, the time to complete a quarter mile (qsec) resulted to summarise quite well the effect of number of cylinders (cyl), displacement (disp), rear axle ratio (rat), gross engine power (hp). Fig. 4 shows that there is a clear difference in milage for cars with or without automatic transmission as well as a relation between milage and both weight and seconds to quarter mile.

Methods

We applied the commonly used methods for the initial exploratory data analysis of the `mtcars` dataset. A box-plot of the distribution of milage values for different transmission and weight (Fig. 1 and 2). We used simple T-test for hypothesis testing to assess the statistical significance of the differences shown by the plots. However, no formal check of the distribution of data under test have been performed. Then we plotted milage as a function of weight and transmission type to confirm there is a correlation between the two predictor variables (Fig. 3).

In order to select the best linear model we applied an iterative approach starting from a model with all variables as predictors of milage. Then, we repeatedly removed the predictore with the worst (higher) P-value -as reported in the model estimated coefficients- from the list of predictors and recomputed the model. During this phase we always included the intercept in the model. We continued to remove variables until we obtained a minimal set of statistically significant predictors. Finally, since the resulting model had the intercept value not statistically significant we tested the same model without the intercept. This gave an even better model with higher R^2 and better P-values for all predictors. In particular the P-value for `am` improved a lot. Diagnostics of the resulting linear model was done using the standard R plots of residuals versus fitted values, Q-Q, and leverage (Fig. 5, 6, 7 and 8).

Finally, we plotted milage as function of time to quarter mile (x axis) and weight (color code) grouping by transmission type. Then we added the lines representing the linear model of milage against time to quarter mile for automatic and manual transmission cars (Fig. 4).

Conclusions

Based on an initial exploratory analysis we identified that (at least) car weight is a confounder when considering the correlation between transmission type and milage. We defined and checked different linear models to predict milage from the other variables in the dataset.

Applying an interative process to reduce the number of predictor to the minimum set we obtained a model that used weight, time to quarter mile and transmission type. We estimated that on average cars with automatic transmission has a milage of `4.3` mpg higher than manual, keeping all other variable constant.

Finally, since it appears that many variables are confounders between milage and transmission, in order to obtain more reliable results one should consider to collect data from a sample of car data where the same make and model is measured both with and without automatic transmission. This would greatly minimizethe counfounding factor of other variables.

References

The complete R-Markdown source code and plots of this report are available on rcblt GitHub page

Appendix

Fig. 1: Milage vs. Transmission

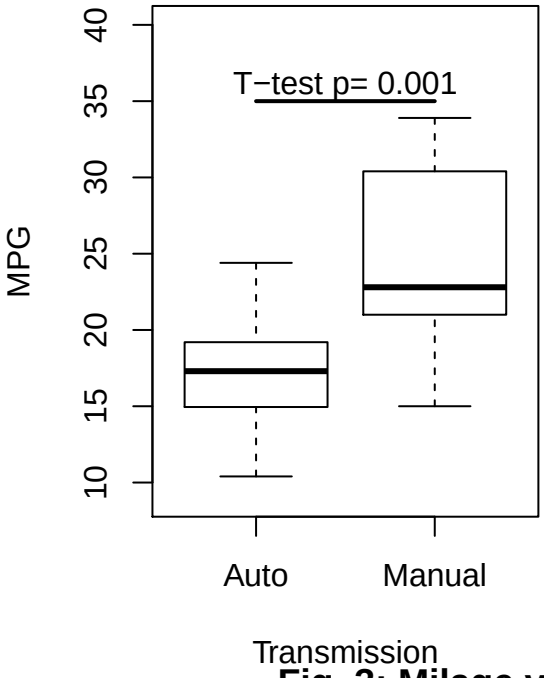

Fig. 2: Weight vs. Transmission

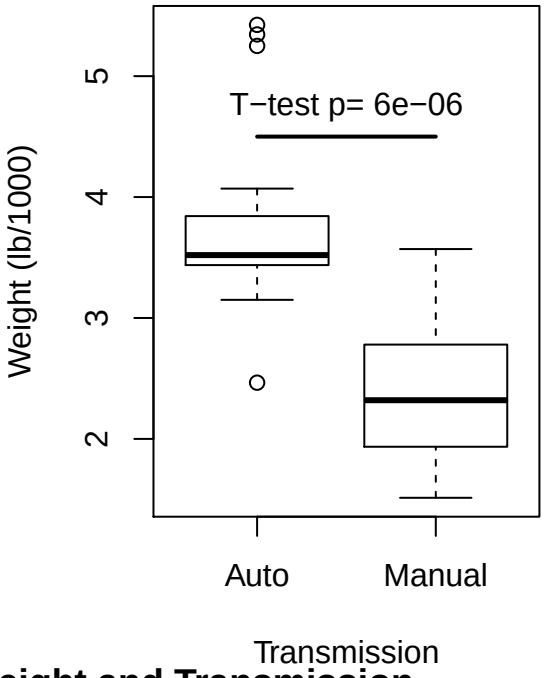

Fig. 3: Milage vs. Weight and Transmission

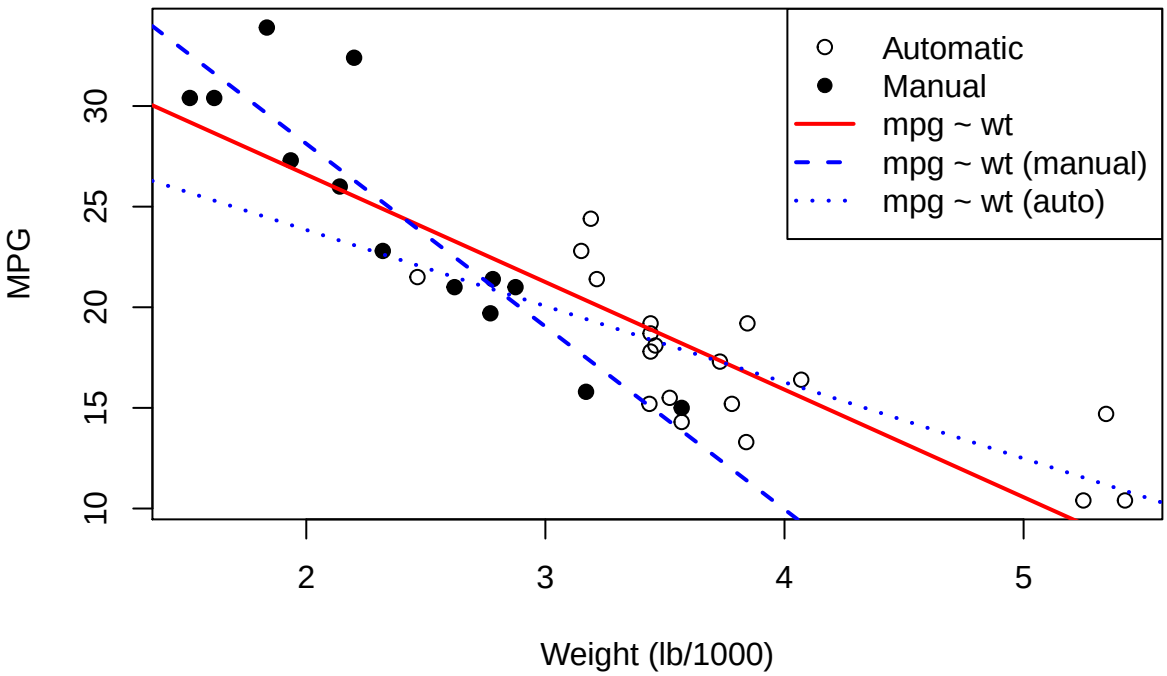

Fig. 4: Milage vs. Qsec and Weight

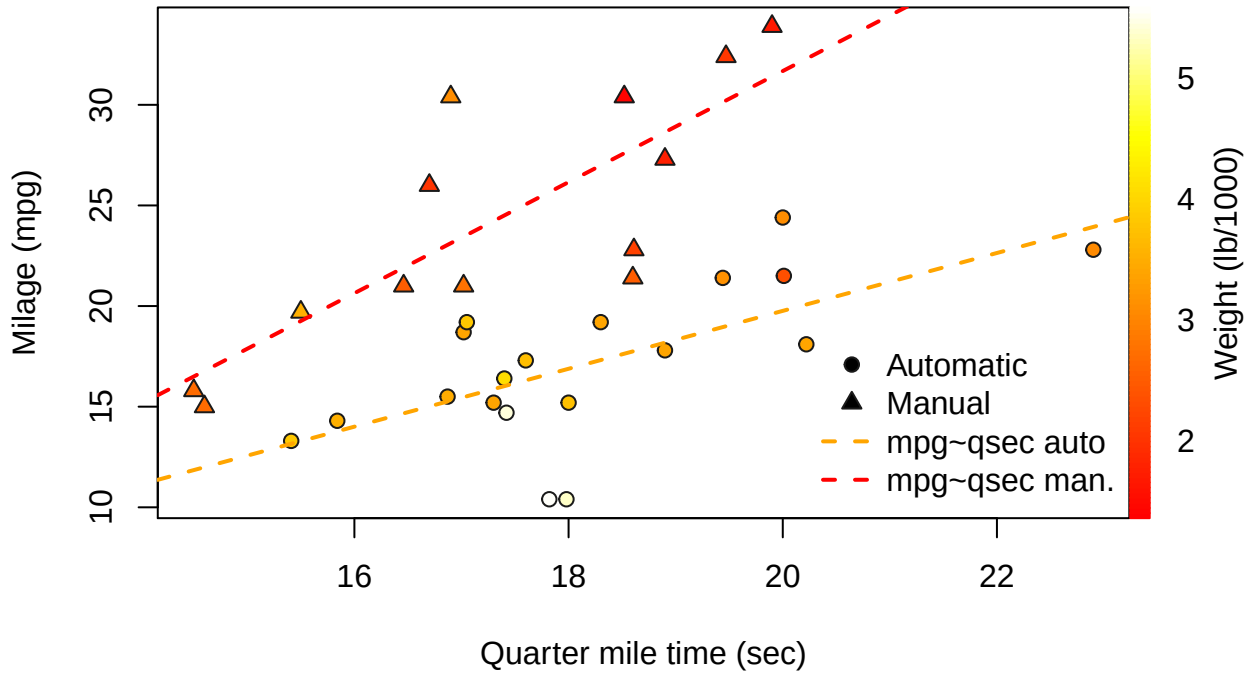

Fig. 5: Residual vs fitted

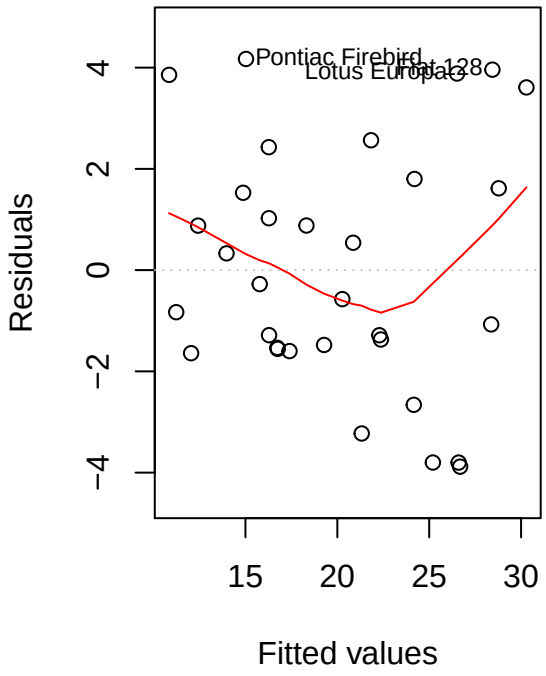

Fig. 6: Normal Q−Q

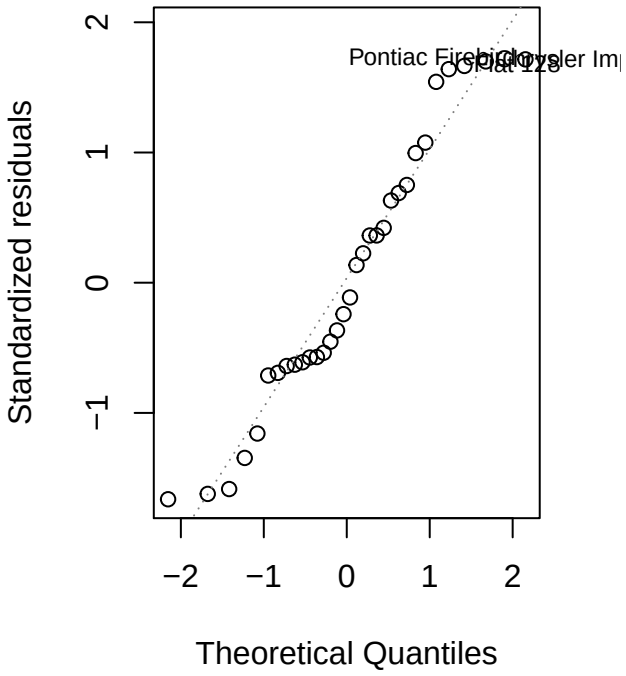

Fig. 7: Scale location

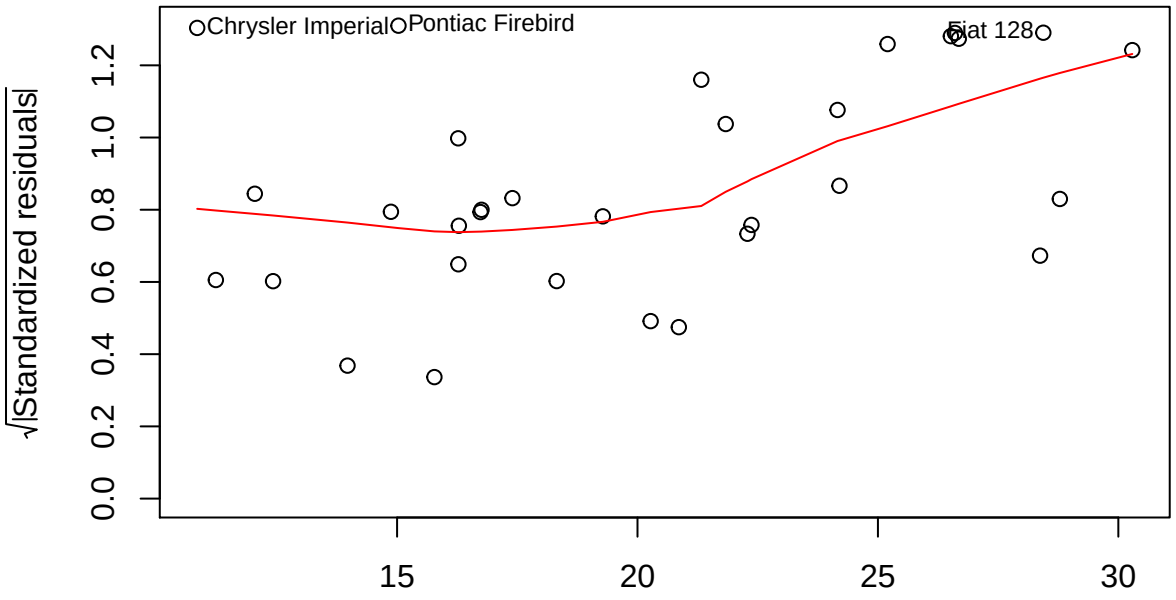

Fig. 8: Residuals vs leverage