Does Transmission Type Affect Fuel Economy?

Brian Schousek September 2014

Executive summary

Common wisdom holds that manual transmissions achieve better fuel efficiency than automatic transmissions. Data from Motor Trend for 32 1973-74 model automobiles (from the mtcars data set) is analyzed here to test that idea. In the end, it is shown that there is no demonstrable statistical difference between vehicles with automatic and manual transmissions.

First model

We first examine a naive model taking into account only the difference between automatic and manual transmissions. A Welch Two sample t-test tells us that the means of mileage based on transmission are not equal, with manual transmissions having a mean value of 24.4 mpg and automatics having 17.1 mpg with a p-value of 0.0014 which seems pretty convincing. See Figure 1 in the appendix for a boxplot of the data.

Exploratory Analysis

Taking a look at the list of cars included, however, reveals a wide variety of vehicle types. They range from a four cylinder Toyota Corona to a Ferrari Dino. The differences between these many vehicles go far beyond the type of transmission. Figure 2 in the appendix shows a corrgram of all the possible predictors. As can be seen, mileage has very strong correlations (>0.70) with cylinders,displacement, horsepower and weight. Correlation with transmission type is strong at 0.60. Other parameters with correlations greater than the transmission type include rear axle ratio (drat) and engine configuration (vs, or V vs straight.) Let us consider two more models: one with only the very strong coefficients and the other including all coefficients with correlation at least as strong as the transmission type.

Second Model: Very Strong correlations

Fitting a model of mileage versus cylinders, displacement, horsepower and weight yields a fit with an adjusted R-squared of 0.8305 but an examination of the output shows that only coefficients for weight and factor levels for 4 cylinder and 6 cylinder reach significance. Noting that displacement and cylinder count are strongly physically related the model was refined to include interaction between cylinder count and displacement. In this refined model all coefficients but horspower achieved significance and the adjusted R squared value increases to 0.8679. Residuals for the fitted refined model have a mean of 0 and standard deviation of 1.9.

Third Model: Strong correlations

A third model was fitted which includes the second model predictors (including cylinder/displacement interaction) and adding transmission type, axle ratio, and engine configuration. The adjusted R squared value for this model was actually worse than the second model, at 0.8518 compared to the previous 0.8679. The magnitude and distribution of overall residuals remain about the same as well with a mean of 0 and standard deviation of 1.9.

Final model: Very strong correlations + transmission

Since adding the other strong correlation terms seems to have reduced the overall R-squared value, and didn't improve the residuals, we choose for a final model just the very strong predictors with transmission type added. This leaves us with an adjusted R-squared of 0.8626 which is slightly lower than that of the refined second model.. Residuals have a mean of 0 and standard deviation of 1.9.

Residuals and Diagnostics

A plot of residuals per predictor is shown as Figure 3 in the appendix. Little evidence of heteroscedasticity is present and the residuals appear comfortably linear. Diagnostic plots in Figure 4 further highlight these observations, and the Normal QQ plot indicates that the residuals are comfortably normal. The Hornet 4 does have more leverage on the final results than other models. An ANOVA of the model in the table below show all predictors except transmission type and horsepower having significance.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
cyl	2	824.78	412.39	82.60	0.0000
disp	1	57.64	57.64	11.55	0.0025
hp	1	18.50	18.50	3.71	0.0667
wt	1	64.99	64.99	13.02	0.0015
am	1	9.72	9.72	1.95	0.1763
cyl:disp	2	35.58	17.79	3.56	0.0449
Residuals	23	114.83	4.99		

Table 1: Anova of Final Model

Inference and coefficient interpretation

The coefficients of the final model can be seen in table 2a below. Examining the transmission coeffecient (ammanual) from the fit shows that holding all else equal, we could expect that a car with an automatic transmission would have mileage 0.4096 less than one with a manual transmission. This is directionally similar to what was observed in the very first model which took only transmission type into account. The magnitude of the difference is much less when other confounding variables are taken into account however. Furthermore, the box portion of the boxplot from the first model (Figure 1) does not overlap across transmission type, whereas the box portion of the partial residual boxplot for transmission in the final model (figure 3) does overlap a great deal.

Examining confidence intervals for the coefficients in table 2b below helps explain this overlap and sheds light on final conclusions. It is apparent that the 95% confidence intervals for both transmission type, weight and horsepower include zero. Based on this uncertainty, it could be possible that vehicles with automatic transmissions actually have better mileage on average than those with manual transmissions.

	Estimate	Std. Error	t value	Pr(>|t|)		2.5 %	97.5 %
(Intercept)	41.7553	4.2306	9.87	0.0000	(Intercept)	33.00	50.51
cyl6	-14.9168	5.7030	-2.62	0.0155	cyl6	-26.71	-3.12
cyl8	-14.2478	5.6471	-2.52	0.0190	cyl8	-25.93	-2.57
disp	-0.0847	0.0356	-2.38	0.0260	disp	-0.16	-0.01
hp	-0.0198	0.0137	-1.45	0.1612	hp	-0.05	0.01
wt	-2.1223	1.1582	-1.83	0.0799	wt	-4.52	0.27
ammanual	0.4096	1.5546	0.26	0.7945	ammanual	-2.81	3.63
cyl6:disp	0.0943	0.0387	2.44	0.0230	cyl6:disp	0.01	0.17
cyl8:disp	0.0851	0.0337	2.53	0.0188	cyl8:disp	0.02	0.15

(a) Table of Coeffecents   (b) Table of ANOVA outuput

Table 2: Final Model

Conclusion

Since the confidence interval on the transmission type coefficient for a complete linear model based on Motor Trend 1974 data includes 0, it can not be conclusively stated that a vehicle with manual transmission has better fuel mileage than one with an automatic transmission.
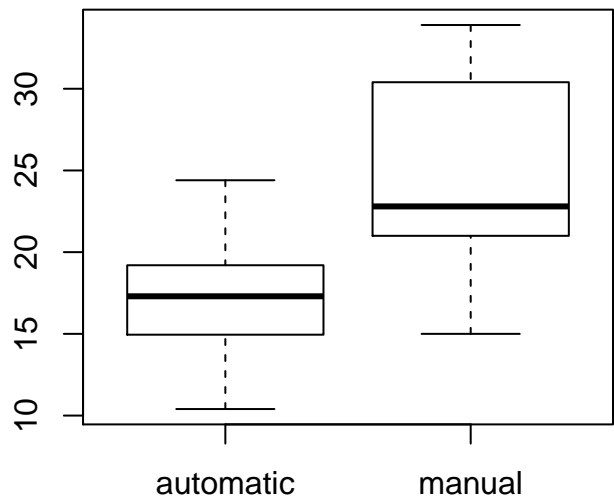
Appendix

Figure 1: Single Term Model
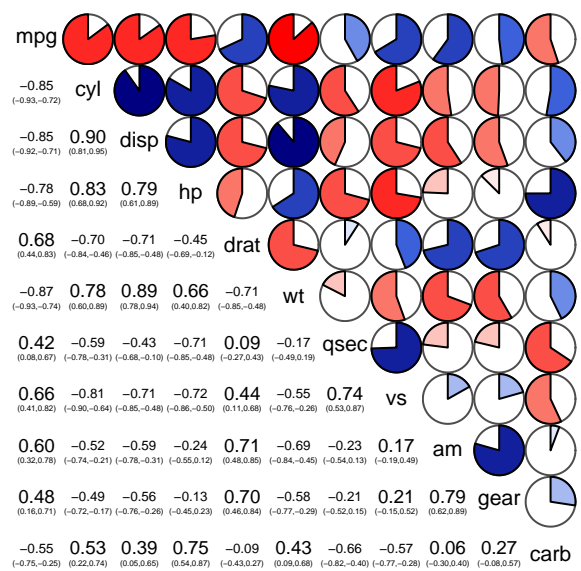

Figure 2: Correlation of Terms
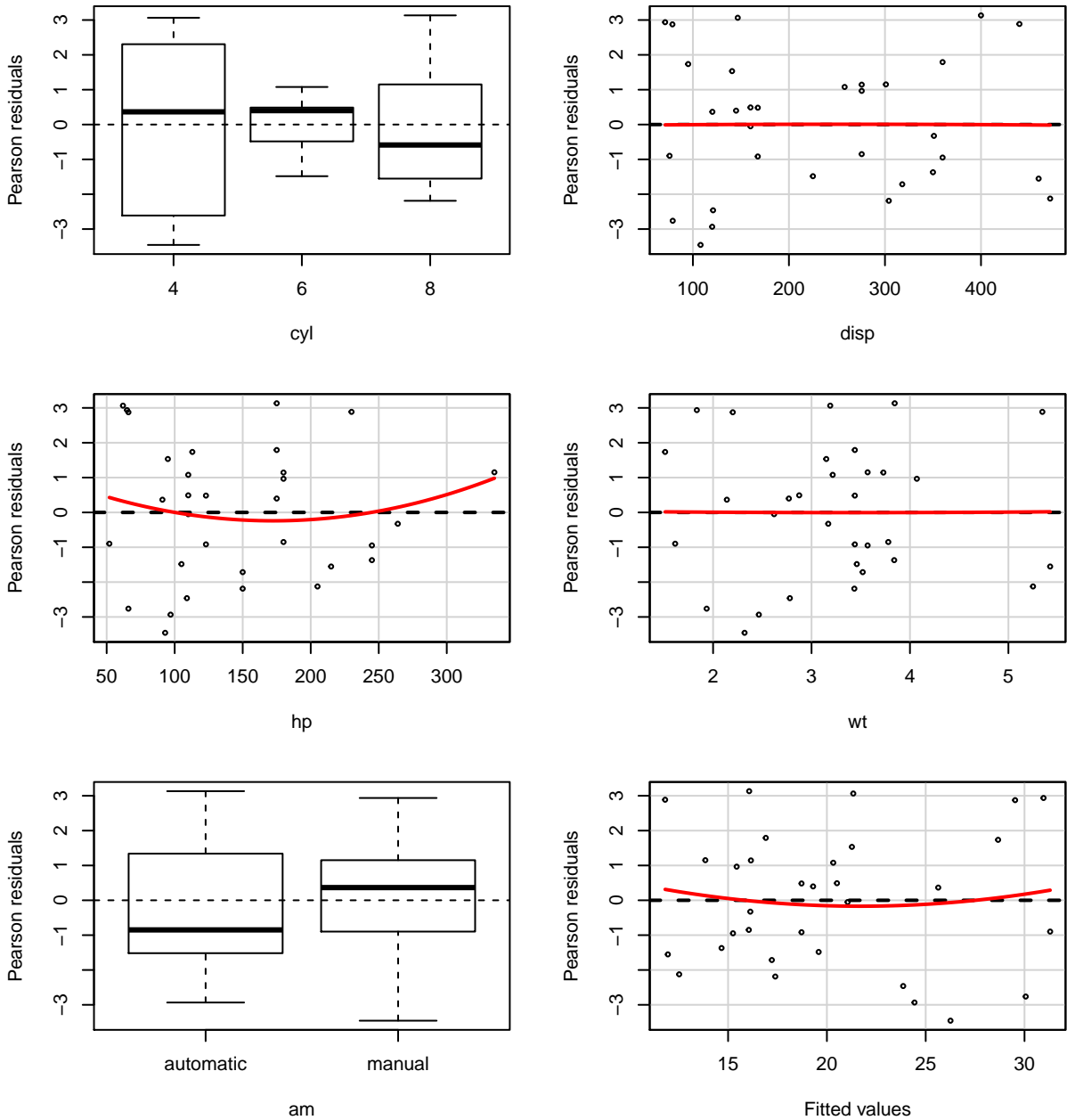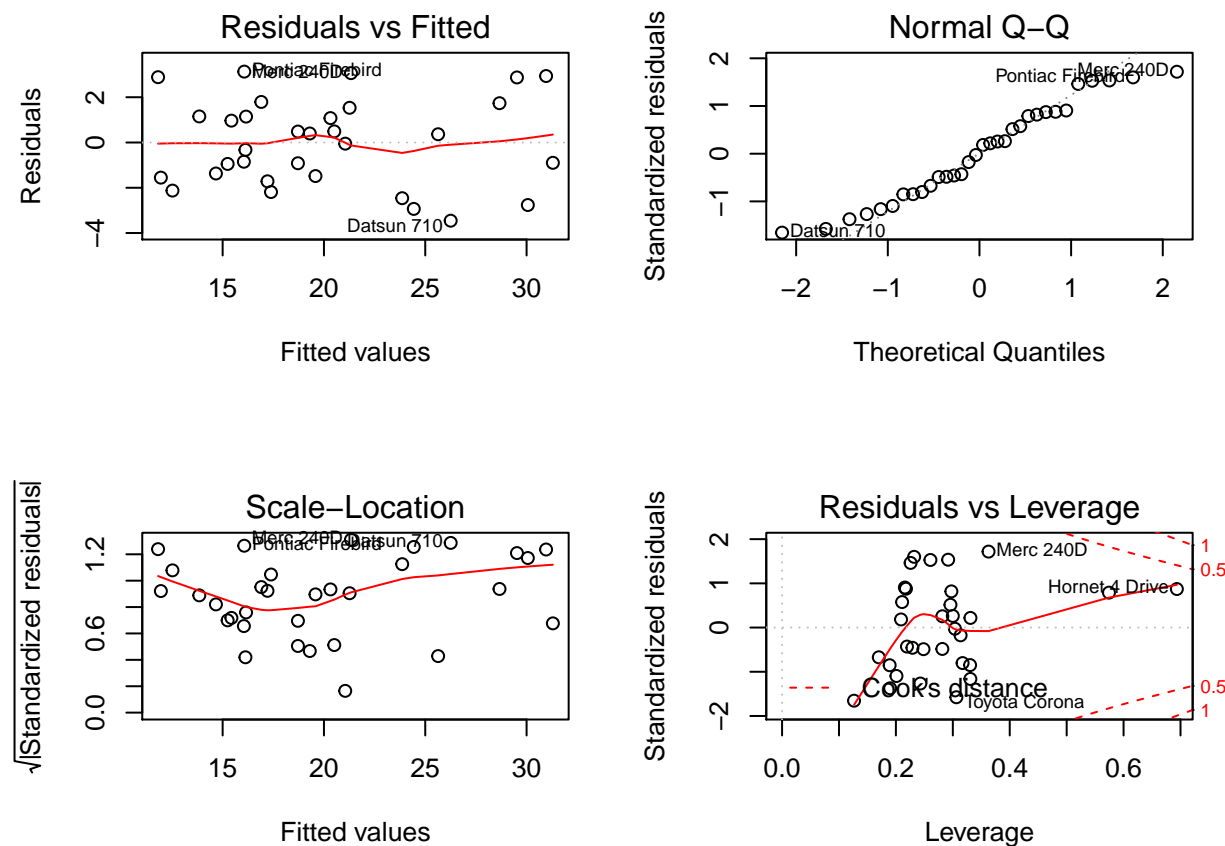
Figure 3: Residual Plots

Figure 4: Diagnostics

Computing Environment

These analyses were done using the following versions of R, the operating system, and add-on packages:

```
## R version 3.1.1 (2014-07-10)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
##
## locale:
## [1] LC_COLLATE=English_United States.1252
## [2] LC_CTYPE=English_United States.1252
## [3] LC_MONETARY=English_United States.1252
## [4] LC_NUMERIC=C
## [5] LC_TIME=English_United States.1252
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] xtable_1.7-4 car_2.0-21   corrgram_1.6
##
## loaded via a namespace (and not attached):
##  [1] cluster_1.15.3   colorspace_1.2-4 digest_0.6.4    evaluate_0.5.5
##  [5] formatR_1.0      gclus_1.3.1      grid_3.1.1      htmltools_0.2.6
##  [9] knitr_1.6        MASS_7.3-34      nnet_7.3-8      rmarkdown_0.2.68
## [13] seriation_1.0-13 stringr_0.6.2    tools_3.1.1    TSP_1.0-9
## [17] yaml_2.1.13
```

Source Availability

R Markdown source for this report can be found at https://github.com/bschousek/regression.git

This is a class project for Regression Models, a Johns Hopkins Bloomberg School of Public Health Data Science Course through Coursera