An inquiry into the influence of automatic vs manual transmission on fuel economy of cars in USA, 1974

Eoin Sharkey

March 4 - 22 , 2015

Executive Summary

It is America, 1974. Motorists are reeling from the impact of the oil-price shock. The price of gas has increased by over 40% in the previous year after decades of low or no increases[1]. So the fuel economy of cars, in miles per gallon (mpg) has become a hot topic for US motorists, notably the aficionados who subscribe to the leading popular technical publication *Motor Trends*. [1] [2]

In this brief paper we will address the question :

`Which is better for fuel economy; automatic or manual, in the presence of these other factors_.`

Noting that the manual cars in the cohort studied have mean mpg significantly greater (7 mpg) than the automatic cars, we will show that manual transmission offers a small (~0.8 mpg) fuel economy over automatic for a hypothetical car of average weight and performance (measured by QSec) - however the error associated with the prediction means any difference may not be statistically significant. We will offer some reasons for the difference between the observed and modelled cases and make suggestions for further study.

[1] Especially interesting is whether motorists need to move from automatic to manual transmission cars to achieve superior fuel economy. Clouding the issue is the number of other car design factors (weight, displacement and so on) which may impact fuel economy either independently of, or interacting with the transmission type.

[2] Newton's 2nd Law of Motion :

$$\vec{F} = m\vec{a}$$

Data Collection

For this analysis we used the mtcars data, part of standard R distribution[2]. Data was loaded and processed using the R programming language [3]. This consists of 32 records, each for 1 make of car.

For each record 10 observations including fuel economy (miles per gallon or mpg), weight (wt in units of 1,000 lb), automatic (0) or manual (1) transmission (am), Qsec (seconds to travel a quarter-mile), displacement (disp in units of cubic inches) and so on.

	mpg	cyl	disp	hp	drat	wt
Mazda RX4	21.0	6	160	110	4	2.6
Mazda RX4 Wag	21.0	6	160	110	4	2.9
Datsun 710	22.8	4	108	93	4	2.3

mpg	fuel economy (miles per gallon)
am	automatic(1) or manual (1)
qsec	Seconds to quarter-mile
disp	Displacement(in^3)
hp	horsepower
drat	drive ratio

Figure 1: Key fields

Exploratory Analysis

Exploratory analysis was by examining tables and plots of the downloaded data. Inspection using complete.cases() and the summary() command within r reveals no missing or anomalous data. We noted high-levels of correlation between the variables which suggests multivariable models possibly with interaction terms may be appropriate. Observing that weight (wt) was a highly influential variable we constructed simple linear models for the different classes of transmission.

```
## Loading required package: ggplot2
```

We noted that the mean value for mpg was widely different for manual (24.4) vs automatic (17.1), and confirmed the difference (7.2) was significant with Student's T-test (95%CI : 3.21, 11.28)

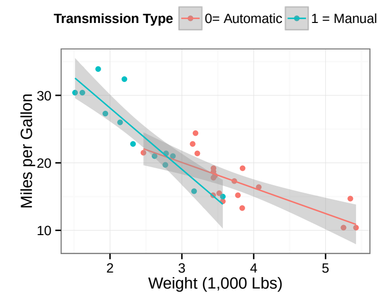


Figure 2: Initial linear models

Statistical Modelling Methods

To relate fuel-economy with other vehicle characteristics we constructed a series of linear regression models, which were tested using the linear-model and ANOVA functions of [2]. The goal of a 'parsimonious' model was pursued using the simple technique outlined by King[4].

Models built are reported in the Appendix where standard parameters including adj-R^2, RSS and the p-values for constant and co-efficient terms are given.

Model-building and evaluation

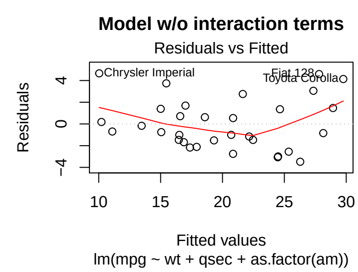
Starting with a full model (lm(mpg~.)), terms were successively removed using commands of the form (update(prevLM, .~. -LST)) where LST is the 'least significant term' as measured by the (significance of) the t-value.

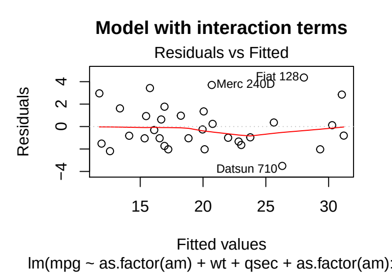
This resulted in a parsimonious model with 3 predictor variables. Inspection of residuals (see Table 3) revealed a parabolic shape characteristic of an interaction term being required.

Therefore models with interaction terms were developed and after inspection of model differences using ANOVA, a final model with a single interaction term was chosen - see sidebar:

Co-efficient mean values and 95% CI are given in the table below. Also shown are predicted fit values (and 95% CI limits) for a hypothetical car with mean values for weight and QSec performance.

Our model shows that a car with manual transmission has 14.08 better mpg than an automatic,that every additional 1,000 Lb weight





$MPG_i = \beta_0 + \beta_1 WT_i + \beta_2 QSec_i + \beta_3 AM_i + \beta_4 WT_i * AM_i + \epsilon_i$

where WT_i is weight , $Qsec_i$ is seconds to travel quarter-mile, AM_i is automatic vs Manual transmission $AM_i * WT_i$ is the interaction between those terms and ϵ_i represent error and we assume $\epsilon_i \sim N(0, \sigma^2)$

Figure 3: Final model equation

reduces fuel economy by an average -2.94 mpg and that each additional second per quartermile adds 1.02 to mpg *holding all other variables constant*, and also that a car with manual transmission has a steeper response curve to weight - with the slope increasing from -2.94 by an additional -4.14 mpg for a manual car.

	2.5% CI	97.5% CI	Mean
(Intercept)	-2.38	21.83	9.72
as.factor(am)1	7.03	21.13	14.08
wt	-4.30	-1.57	-2.94
qsec	0.50	1.53	1.02
as.factor(am)1:wt	-6.60	-1.69	-4.14

	fit	lwr	upr
Automatic	18.43	17.15	19.71
Manual	19.18	16.91	21.46

Table 1: Using the model to predict mpg for cars with mean values of weight (wt)and QSec (qsec) for manual and automatic transmission

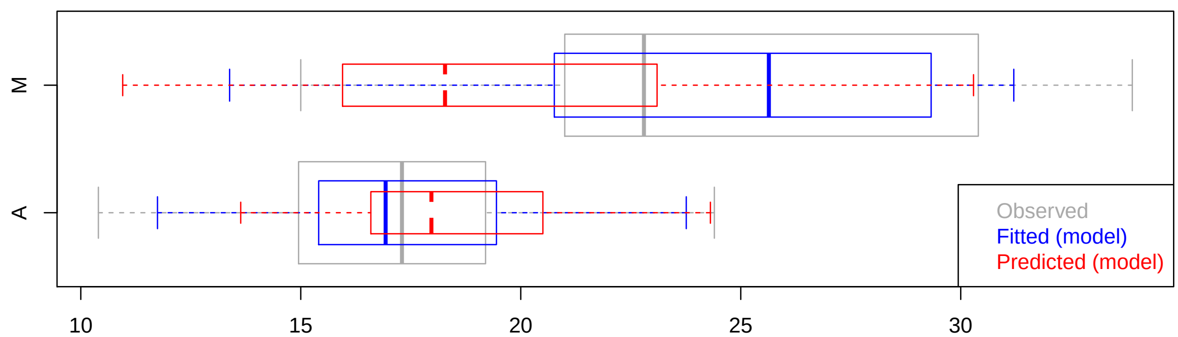


Figure 4: Distribution of mpg for observed and modelled(fitted) values for Manualand Automatic cars

Results for our 'average' car show, that actually there is little difference in the mpg between automatic and manual, furthermore the wide (and overlapping) error bands mean the results are likely not to be statistically significant.

Building on this interesting result, we create some additional data; a set of 18 hypothetical cars, 9 auto, 9 manual where the weight of the cars runs from the 10% quantile to the 90% in even steps and the qsec performance runs in the opposite direction : in essence we are designing cars which range from light, but slow to heavy and fast.

Conclusion

How do we explain the paradox that our data plainly shows a significant benefit in mpg for manual *vice* automatic cars, but our model despite being carefully constructed to maximise adj-R^2 and minimise RSS while keeping only significant terms shows only a small, possibly not statistically significant, difference. Either

a) We failed to build a good predictive model, either because we omitted variables which might have been important during the model tuning phase or because they are not available in the in original data-set (unknown unknowns)

b) Our model is sound, and the effects of (automatic vs manual) tranmsission alone are rather small, and what really contributes to fuel economy are other car design parameters such as weight and the 'tuning' of the car for performance or economy (represented by Qsec in our model). In this case we assert that manual vs automatic transmission may simply reflect a choice by the vehicle manufacturer, and we will note that the manual cars in the data are almost all non-US made, while the automatic cars are overwhelmingly US made.

These results can only be applied to the sample data since we have no means of knowing if they are representative of the population of cars available generally - future research might look at this shortcoming. We should be careful about over-fitting any model to such a limited data-set, and future work might include model-building with some form of cross-validation technique applied to guard against this.

Appendix

References

[1]:"First oil-price shock" as referenced in `http://en.wikipedia.org/wiki/1973_oil_crisis`, accesssed on 12/Mar 2015.

[2]: R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL `http://www.R-project.org/`.

[3]: Henderson and Velleman (1981), Building multiple regression models interactively. Biometrics, 37, 391-411 (not accessed)

[4]: "Tutorial on Multiple Regression", author William B. King as published : `http://ww2.coastal.edu/kingw/statistics/R-tutorials/multregr.html`, accessed on 17/Mar 2015.

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Cadillac Fleetwood	10.4	8	472	205	2.93	5.25	17.98	0	0	3	4
Lincoln Continental	10.4	8	460	215	3.00	5.42	17.82	0	0	3	4
Camaro Z28	13.3	8	350	245	3.73	3.84	15.41	0	0	3	4
Duster 360	14.3	8	360	245	3.21	3.57	15.84	0	0	3	4
Chrysler Imperial	14.7	8	440	230	3.23	5.34	17.42	0	0	3	4
Maserati Bora	15.0	8	301	335	3.54	3.57	14.60	0	1	5	8
Merc 450SLC	15.2	8	276	180	3.07	3.78	18.00	0	0	3	3
AMC Javelin	15.2	8	304	150	3.15	3.44	17.30	0	0	3	2
Dodge Challenger	15.5	8	318	150	2.76	3.52	16.87	0	0	3	2
Ford Pantera L	15.8	8	351	264	4.22	3.17	14.50	0	1	5	4
Merc 450SE	16.4	8	276	180	3.07	4.07	17.40	0	0	3	3
Merc 450SL	17.3	8	276	180	3.07	3.73	17.60	0	0	3	3
Merc 280C	17.8	6	168	123	3.92	3.44	18.90	1	0	4	4
Valiant	18.1	6	225	105	2.76	3.46	20.22	1	0	3	1
Hornet Sportabout	18.7	8	360	175	3.15	3.44	17.02	0	0	3	2
Merc 280	19.2	6	168	123	3.92	3.44	18.30	1	0	4	4
Pontiac Firebird	19.2	8	400	175	3.08	3.85	17.05	0	0	3	2
Ferrari Dino	19.7	6	145	175	3.62	2.77	15.50	0	1	5	6
Mazda RX4	21.0	6	160	110	3.90	2.62	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160	110	3.90	2.88	17.02	0	1	4	4
Hornet 4 Drive	21.4	6	258	110	3.08	3.21	19.44	1	0	3	1
Volvo 142E	21.4	4	121	109	4.11	2.78	18.60	1	1	4	2
Toyota Corona	21.5	4	120	97	3.70	2.46	20.01	1	0	3	1
Datsun 710	22.8	4	108	93	3.85	2.32	18.61	1	1	4	1
Merc 230	22.8	4	141	95	3.92	3.15	22.90	1	0	4	2
Merc 240D	24.4	4	147	62	3.69	3.19	20.00	1	0	4	2
Porsche 914-2	26.0	4	120	91	4.43	2.14	16.70	0	1	5	2
Fiat X1-9	27.3	4	79	66	4.08	1.94	18.90	1	1	4	1
Honda Civic	30.4	4	76	52	4.93	1.61	18.52	1	1	4	2
Lotus Europa	30.4	4	95	113	3.77	1.51	16.90	1	1	5	2
Fiat 128	32.4	4	79	66	4.08	2.20	19.47	1	1	4	1
Toyota Corolla	33.9	4	71	65	4.22	1.83	19.90	1	1	4	1

Table 2: The full mtcars data-set

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	22.00	139.02				
2	28.00	169.29	-6.00	-30.26	0.80	0.58
3	27.00	117.28	1.00	52.01	8.23	0.01

Table 3: Plot of Fitted vs Residuals and Comparison of models, (1) = Intermediate Model, (2) = Parsimonious model with no interaction terms, (3) = Final model with a single interaction term

Table 4: Comparison of Models (1) = intermediate model with 8 predictors, (2) = Select parsimonious model with 3 predictors, (3) = Parsimonious model with most significant interaction term added. Numbers in brackets are the Standard Error for each co-efficient.

	Dependent variable:		
	mpg		
	(1)	(2)	(3)
as.factor(cyl)6	-1.458		
	(1.983)		
as.factor(cyl)8	0.484		
	(3.910)		
disp	0.007		
	(0.014)		
hp	-0.029		
	(0.017)		
drat	0.588		
	(1.503)		
wt	-3.155^{**}	-3.917^{***}	-2.937^{***}
	(1.420)	(0.711)	(0.666)
qsec	0.523	1.226^{***}	1.017^{***}
	(0.690)	(0.289)	(0.252)
as.factor(vs)1	1.238		
	(2.106)		
as.factor(am)1:wt			-4.141^{***}
			(1.197)
as.factor(am)1	3.001	2.936^{**}	14.079^{***}
	(1.853)	(1.411)	(3.435)
Constant	19.866	9.618	9.723
	(14.801)	(6.960)	(5.899)
Observations	32	32	32
R^2	0.877	0.850	0.896
Adjusted R^2	0.826	0.834	0.880
Residual Std. Error	2.514 (df = 22)	2.459 (df = 28)	2.084 (df = 27)
F Statistic	17.355^{***} (df = 9; 22)	52.750^{***} (df = 3; 28)	58.061^{***} (df = 4; 27)
Note:			*p<0.1; **p<0.05; ***p<0.01