Analysis of how Miles Per Gallon is effected by car characteristics

Chris Shaw

12 June 2016

Executive Summary

Motor Trend magazine have requested an analysis of the relationship between a set of car variables and the predicted miles per gallon (MPG). The following questions are of particular interest:

- Is an automatic or manual transmission better for MPG?
- Quantify the MPG difference between automatic and manual transmissions.

The analysis is based on a data set of a collection of cars, summarised in the appendix.

We conclude that manual transmission is better for MPG, especially when considered alongside vehicle weight and horsepower. For a car of average weight and horsepower, we expect an increase of 2.1 MPG by moving transmission to manual. On the other hand, MPG decreases by 0.3 for each additional 100lb of weight and by 0.4 for each addtional 10 horsepower.

Exploratory analysis

Summarising the data set gives an insight into how the mean MPG changes when compared to different variables. At a high level, the overall means for manual and automatic are as follows:

Type	Num Cars	Mean mpg	Std dev
Automatic	19	17.147	3.834
Manual	13	24.392	6.167

If we group the vehicles by horsepower quantiles (1 being the <25% quantile and 4 being >75%), it can be seen that there is a sizeable increase in the mean in each category when going from automatic to manual:

Type	Horsepower quantile	Num Cars	Mean mpg	Std dev
Automatic	1	2	23.600	1.131
Automatic	2	5	19.600	1.768
Automatic	3	7	16.786	1.663
Automatic	4	5	12.620	2.090
Manual	1	6	28.800	4.186
Manual	2	4	23.450	4.637
Manual	3	1	19.700	NA
Manual	4	2	15.400	0.566

However, when the cars are grouped by number of cylinders, there is a much more modest increase for 6 and 8 cylinder cars moving from automatic to manual.

Type	Number of cylinders	Num Cars	Mean mpg	Std dev
Automatic	4	3	22.900	1.453
Automatic	6	4	19.125	1.632
Automatic	8	12	15.050	2.774
Manual	4	8	28.075	4.484
Manual	6	3	20.567	0.751
Manual	8	2	15.400	0.566

We need to build a more sophisticated model to understand how each variable contributes to changes in MPG.

Regression model

The first task was to transform the data so that the weight, horsepower, quarter mile time, displacement and rear axle ratio of centred around the mean. This makes interpreting the co-efficients in the model easier.

We fitted a number of models of the form $mpg \sim am + var$ where var was each of the 9 other variables in turn. The most promising combinations which left the smallest residual mean square are:

Formula	Residual.MSS
mpg ~ am + wt	9.597231
mpg ~ am + cyl	9.446274
mpg ~ am + hp	8.463424

Further analysis showed that *cyl* didn't really add more to the model than *hp*, which is to be expected since those variables are somewhat dependent on each other. We were able to improve the residuals by including a third term. The final selected model was $mpg \sim am + wt + hp$ which produced a residual mean square of 6.439. The table of co-efficients and the confidence interval are:

	2.5 %	97.5 %	Model_coef
(Intercept)	17.7756947	20.7125408	19.2441178
am1	-0.7357587	4.9031790	2.0837101
wt	-4.7323235	-1.0248273	-2.8785754
hp	-0.0571545	-0.0178029	-0.0374787

The Model_coef value for the Intercept shows that the mean MPG for a manual car of average weight and horsepower is -0.06. 19.24. The co-efficients for the other terms in the model show the relative increase or decrease per unit increase in the associated variable.

The confidence interval for weight and horsepower confirms the negative adjustment to MPG 95% of the time. However, it can be seen that the lower bound for the transmission interval could be negative. This needs further investigation.

The plots in the appendix show the diagnostics for this selected model. The QQ plot is along the expected line, indicating that the errors are guassian. Furthermore, the residuals plot shows the Cook distance is less than 0.5 for all data points

Appendix

Dataset

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160.0	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160.0	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258.0	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360.0	175	3.15	3.440	17.02	0	0	3	2
Valiant	18.1	6	225.0	105	2.76	3.460	20.22	1	0	3	1
Duster 360	14.3	8	360.0	245	3.21	3.570	15.84	0	0	3	4
Merc 240D	24.4	4	146.7	62	3.69	3.190	20.00	1	0	4	2
Merc 230	22.8	4	140.8	95	3.92	3.150	22.90	1	0	4	2
Merc 280	19.2	6	167.6	123	3.92	3.440	18.30	1	0	4	4
Merc 280C	17.8	6	167.6	123	3.92	3.440	18.90	1	0	4	4
Merc 450SE	16.4	8	275.8	180	3.07	4.070	17.40	0	0	3	3
Merc 450SL	17.3	8	275.8	180	3.07	3.730	17.60	0	0	3	3
Merc 450SLC	15.2	8	275.8	180	3.07	3.780	18.00	0	0	3	3
Cadillac Fleetwood	10.4	8	472.0	205	2.93	5.250	17.98	0	0	3	4
Lincoln Continental	10.4	8	460.0	215	3.00	5.424	17.82	0	0	3	4
Chrysler Imperial	14.7	8	440.0	230	3.23	5.345	17.42	0	0	3	4
Fiat 128	32.4	4	78.7	66	4.08	2.200	19.47	1	1	4	1
Honda Civic	30.4	4	75.7	52	4.93	1.615	18.52	1	1	4	2
Toyota Corolla	33.9	4	71.1	65	4.22	1.835	19.90	1	1	4	1
Toyota Corona	21.5	4	120.1	97	3.70	2.465	20.01	1	0	3	1
Dodge Challenger	15.5	8	318.0	150	2.76	3.520	16.87	0	0	3	2
AMC Javelin	15.2	8	304.0	150	3.15	3.435	17.30	0	0	3	2
Camaro Z28	13.3	8	350.0	245	3.73	3.840	15.41	0	0	3	4
Pontiac Firebird	19.2	8	400.0	175	3.08	3.845	17.05	0	0	3	2
Fiat X1-9	27.3	4	79.0	66	4.08	1.935	18.90	1	1	4	1
Porsche 914-2	26.0	4	120.3	91	4.43	2.140	16.70	0	1	5	2
Lotus Europa	30.4	4	95.1	113	3.77	1.513	16.90	1	1	5	2
Ford Pantera L	15.8	8	351.0	264	4.22	3.170	14.50	0	1	5	4
Ferrari Dino	19.7	6	145.0	175	3.62	2.770	15.50	0	1	5	6
Maserati Bora	15.0	8	301.0	335	3.54	3.570	14.60	0	1	5	8
Volvo 142E	21.4	4	121.0	109	4.11	2.780	18.60	1	1	4	2

Residual plots

plot(fit_am_wt_hp)