Christopher Androne

Regression Analysis Project

Summer 2013

Introduction

One of the ways cars have historically been measured is by their rate of acceleration. In particular, the two most popular acceleration metrics have been 0-60 miles per hour and the time and trap speed through the quarter mile. Of these two metrics, the speed of a car at the end of the quarter mile is most indicative power, as a 0-60 time is often limited by traction and shift points. However, what characteristics of a car contribute to the quickest quarter mile trap speed? To answer this question, I will regress quarter mile trap speed against the following characteristics:

- Engine Displacement (in cubic centimeters) – chosen because a bigger engine should produce more power.
- Horsepower – one measure of power
- Torque – one measure of force
- Weight
- Weight per Horsepower
- Turbo Engine (to standardize this data, yes = 1, no = 0) – turbo engines may have certain advantages in creating power.
- Transmission Type (to standardize this data, Manual = 1, Automatic = 0) – it was once believed that a manual is faster than an automatic. Is this still true?
- EPA City Fuel Economy - Since combustion is the product of air, fuel, and spark, it would seem a more powerful engine would use more fuel.

To find the necessary data, I used data from various pages on www.caranddriver.com. I selected 29 cars that are somewhat interesting to me, with an additional goal of representing fast, slow, and normal cars. So my listing contains supercars such as the Bugatti Veyron, and economy cars such as my current ride, the Mazda 3.

Data

The variables for my analysis are as follows:

Y – Quarter Mile Trap Speed

X_1 – Displacement

X_2 – Horsepower

X_3 – Torque

X_4 – Weight

X_5 – LB/HP

X_6 – Turbo

X_7 – Transmission

X_8 – EPA City MPG

The automotive data was found at www.caranddriver.com. I can be seen below:

Car	Qtr Trap Speed	Displacement	HP	Torque	Weight	LB/HP	Turbo	Transmission	EPA City
MB GL 450	97	4663	362	400	5855	16.17	1	0	14
Mazda 3	91	2488	184	185	3094	16.82	0	0	28
Porsche Cayman	102	2706	275	213	3083	11.21	0	1	20
Dodge Caravan	86	3605	283	260	4694	16.59	0	0	17
Nissan GTR	124	3799	545	463	3885	7.13	1	0	16
BMW 320	92	1997	180	200	3337	18.54	1	0	24
BMW 135	107	2979	320	317	3335	10.42	1	1	20
Subaru BRZ	94	1998	200	151	2757	13.79	0	1	22
Subaru BRZ Auto	92	1998	200	151	2814	14.07	0	0	25
Bugatti	142	7998	1001	922	4486	4.48	1	1	7
VW GTI	99	1984	217	258	3232	14.89	1	1	23
Ford Focus ST	100	1999	247	266	3057	12.38	1	1	23
BMW M6	119	4395	560	500	4496	8.03	1	1	14
Jaguar XKR	119	5000	550	502	4144	7.53	1	0	15
MB SL63	124	5461	557	664	4128	7.41	1	0	16
911 Carrera S	114	3800	400	325	3427	8.57	0	1	19
BMW X5 M	113	4395	555	500	5289	9.53	1	0	12
Grand Cherokee SR8	103	6410	470	465	5272	11.22	0	0	12
MB ML 63	113	5461	550	560	5285	9.61	1	0	14
Porsche Cayenne Turbo	109	4806	500	516	5242	10.48	1	0	15
Ford Fiesta	79	1596	120	112	2566	21.38	0	1	28
Honda Fit	83	1497	117	106	2524	21.57	0	1	27
Mazda 2	81	1498	100	98	2286	22.86	0	1	29
Ford Escape	81	1597	178	184	3716	20.88	1	0	22
Honda CRV	86	2354	185	163	3607	19.50	0	0	22
Hyundai Tuscon	83	2360	176	168	3421	19.44	0	0	21
Kia Sportage	82	2360	176	168	3509	19.94	0	0	21
Mazda CX5	82	1998	155	150	3486	22.49	0	0	25
Toyota RAV4	84	2494	179	172	3501	19.56	0	0	21

Correlation

Using the Correlation Analysis tool in Excel, I gathered the following results:

	Qtr Trap Speed	Displacement	HP	Torque	Weight	LB/HP	Turbo	Transmission	EPA City
Qtr Trap	1								
Displacement CCs	0.8151	1.0000							
HP	0.9366	0.9319	1.0000						
Torque	0.9163	0.9419	0.9799	1.0000					
Weight	0.4862	0.7799	0.6607	0.6990	1.0000				
LB/HP	-0.9560	-0.7672	-0.8673	-0.8364	-0.4825	1.0000			
Turbo	0.6318	0.4348	0.5837	0.6485	0.4886	-0.5575	1.0000		
Transmission	0.1195	-0.1667	-0.0219	-0.0720	-0.4792	-0.1106	-0.0441	1.0000	
EPA City	-0.7790	-0.9160	-0.8981	-0.8860	-0.8522	0.7663	-0.5238	0.1971	1

The table above shows quarter mile trap speed has strong positive correlations with displacement, horsepower and torque. Quarter mile trap speed has strong negative correlations with LB/HP and a weaker negative correlation with EPA City miles per gallon. Lastly, quarter mile traps speed has a weaker positive correlation with Turbo status, weight, and the weakest correlation with transmission type.

As one might expect, the explanatory variables are interrelated. For example, there is a very high positive correlation between torque and horsepower, as well as torque and displacement. Similarly, there is a very high negative correlation between torque and EPA city mpg. This kind of correlation should be expected.

The Model

Regression analysis was completed using the Excel Regression Analysis tool. My goal is to optimize my regression equation. To do this, I will start with a regression equation that includes all 8 explanatory variables. I will then eliminate variables that do not sufficiently contribute to the model. This is done by eliminating variables with the highest P value and lowest absolute value of the t statistic. The P value is a probability of obtaining a test statistic as extreme as observed value, assuming the null hypothesis is true. The t statistic is similar in that it measures the departure of an estimated parameter from its estimated value (divided by the standard error).

I will continue eliminating explanatory variables based on their P value and t statistic until I have maximized the R squared value and the F Ratio. The R squared measures how well the data points fit the regression line. The F Ratio is a ratio of the Regression Mean Square Error to the Residual Mean Square Error.

Model Iteration #1 – All Explanatory Variables

Regression Equation: $Y = \alpha + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_4 + \beta_5X_5 + \beta_6X_6 + \beta_7X_7 + \beta_8X_8$

Regression Statistics	
Multiple R	0.99166211
R Square	0.98339375
Adjusted R Square	0.97675125
Standard Error	2.51114553
Observations	29

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	8	7468.43	933.554	148.046	4.35778E-16
Residual	20	126.12	6.306		
Total	28	7594.55			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	103.813022	12.9454	8.0193	0.0000	76.8095	130.8166	76.8095	130.8166
Displacement	-0.0001295	0.0018	-0.0702	0.9448	-0.0040	0.0037	-0.0040	0.0037
HP	0.0388815	0.0163	2.3871	0.0270	0.0049	0.0729	0.0049	0.0729
Torque	0.0129831	0.0222	0.5847	0.5653	-0.0333	0.0593	-0.0333	0.0593
Weight	-0.0020087	0.0016	-1.2177	0.2375	-0.0054	0.0014	-0.0054	0.0014
LB/HP	-1.5497708	0.1917	-8.0839	0.0000	-1.9497	-1.1499	-1.9497	-1.1499
Turbo	2.8115827	2.4175	1.1630	0.2585	-2.2311	7.8543	-2.2311	7.8543
Transmission	0.1784972	1.2625	0.1414	0.8890	-2.4551	2.8121	-2.4551	2.8121
EPA City	0.3798771	0.3527	1.0771	0.2942	-0.3558	1.1155	-0.3558	1.1155

Iteration #1 gives the following regression equation:

$$Y = 103.81 - 0.0001295X_1 + 0.0389X_2 + 0.01298X_3 - 0.0020X_4 - 1.5498X_5 + 2.812X_6 + 0.1785X_7 + 0.3799X_8$$

The R squared value is 0.97675. Iteration #1 shows displacement has the highest P-value at 0.9448 and the lowest absolute t Statistic of 0.0702. This implies engine displacement has the least effect on quarter mile trap speed. I will remove engine displacement as an explanatory variable for Iteration #2.

Model Iteration #2 – Remove Engine Displacement

Regression Equation: $Y = \alpha + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + \beta_8 X_8$

Regression Statistics	
Multiple R	0.991660054
R Square	0.983389662
Adjusted R Square	0.977852882
Standard Error	2.450928611
Observations	29

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	7	7468.403652	1066.915	177.6104	3.12292E-17
Residual	21	126.1480722	6.007051		
Total	28	7594.551724			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	103.759170	12.6127	8.2266	0.0000	77.5296	129.9887	77.5296	129.9887
HP	0.0390676	0.0157	2.4907	0.0212	0.0064	0.0717	0.0064	0.0717
Torque	0.0118279	0.0145	0.8136	0.4250	-0.0184	0.0421	-0.0184	0.0421
Weight	-0.0020646	0.0014	-1.4644	0.1579	-0.0050	0.0009	-0.0050	0.0009
LB/HP	-1.5480203	0.1855	-8.3441	0.0000	-1.9338	-1.1622	-1.9338	-1.1622
Turbo	2.9518841	1.3255	2.2269	0.0370	0.1953	5.7085	0.1953	5.7085
Transmission	0.1709900	1.2278	0.1393	0.8906	-2.3824	2.7243	-2.3824	2.7243
EPA City	0.3824280	0.3424	1.1170	0.2766	-0.3296	1.0944	-0.3296	1.0944

Iteration #2 gives the following regression equation:

$$Y = 103.76 + 0.0391X_2 + 0.01183X_3 - 0.0021X_4 - 1.5480X_5 + 2.952X_6 + 0.1710X_7 + 0.3824X_8$$

The R squared value increased to 0.97785. Iteration #2 shows transmission type has the highest P-value at 0.8906 and the lowest absolute t Statistic of 0.1393. This implies transmission type has the least effect on quarter mile trap speed in iteration #2. I will now remove transmission type as an explanatory variable for Iteration #3.

Model Iteration #3 – Remove Transmission Type

Regression Equation: $Y = \alpha + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_8 X_8$

Regression Statistics	
Multiple R	0.991652319
R Square	0.983374321
Adjusted R Square	0.978840045
Standard Error	2.395683404
Observations	29

ANOVA

	df	*SS*	*MS*	*F*	*gnificance F*
Regression	6	7468.287	1244.715	216.8757	2.03E-18
Residual	22	126.2646	5.739299		
Total	28	7594.552			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	104.317980	11.6879	8.9253	0.0000	80.0788	128.5572	80.0788	128.5572
HP	0.0391743	0.0153	2.5582	0.0179	0.0074	0.0709	0.0074	0.0709
Torque	0.0117348	0.0142	0.8267	0.4173	-0.0177	0.0412	-0.0177	0.0412
Weight	-0.0021549	0.0012	-1.7610	0.0921	-0.0047	0.0004	-0.0047	0.0004
LB/HP	-1.5501244	0.1807	-8.5766	0.0000	-1.9250	-1.1753	-1.9250	-1.1753
Turbo	2.9728551	1.2873	2.3094	0.0307	0.3032	5.6425	0.3032	5.6425
EPA City	0.3754063	0.3310	1.1341	0.2690	-0.3111	1.0619	-0.3111	1.0619

Iteration #3 gives the following regression equation:

$$Y = 104.32 + 0.0392X_2 + 0.01173X_3 - 0.0022X_4 - 1.5501X_5 + 2.973X_6 + 0.3754X_8$$

The R squared value increased to 0.97884. Iteration #3 shows torque has the highest P-value at 0.4173 and the lowest absolute t Statistic of 0.8267. This implies torque has the least effect on quarter mile trap speed among remaining explanatory variables. I will now remove torque as an explanatory variable for Iteration #4.

Model Iteration #4 – Remove Torque

Regression Equation: $Y = \alpha + \beta_2 X_2 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_8 X_8$

Regression Statistics	
Multiple R	0.991391889
R Square	0.982857877
Adjusted R Square	0.979131329
Standard Error	2.379137022
Observations	29

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	5	7464.365	1492.873	263.7448	1.64364E-19
Residual	23	130.1867	5.660293		
Total	28	7594.552			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	101.154828	10.9675	9.2231	0.0000	78.4668	123.8429	78.4668	123.8429
HP	0.0505280	0.0067	7.5126	0.0000	0.0366	0.0644	0.0366	0.0644
Weight	-0.0017971	0.0011	-1.5811	0.1275	-0.0041	0.0006	-0.0041	0.0006
LB/HP	-1.5376709	0.1789	-8.5968	0.0000	-1.9077	-1.1677	-1.9077	-1.1677
Turbo	3.4116365	1.1647	2.9293	0.0075	1.0024	5.8209	1.0024	5.8209
EPA City	0.4455409	0.3177	1.4022	0.1742	-0.2118	1.1029	-0.2118	1.1029

Iteration #4 gives the following regression equation:

$$Y = 101.15 + 0.0505X_2 - 0.0018X_4 - 1.5377X_5 + 3.411X_6 + 0.4455X_8$$

The R squared value increased to 0.97913. Iteration #4 shows EPA City miles per gallon has the highest P-value at 0.1742 and the lowest absolute t Statistic of 1.40. This implies EPA City mpg has the least effect on quarter mile trap speed among remaining explanatory variables. I will now EPA City mpg as an explanatory variable for Iteration #5.

Model Iteration #5 – Remove EPA City MPG

Regression Equation: $Y = \alpha + \beta_2 X_2 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6$

Regression Statistics	
Multiple R	0.990653
R Square	0.981393
Adjusted R Square	0.978291
Standard Error	2.426551
Observations	29

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	7453.236	1863.309	316.4507	2.20127E-20
Residual	24	141.3156	5.888149		
Total	28	7594.552			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	115.5662	3.90427	29.59995	2.11E-20	107.5081922	123.6242	107.5082	123.6242
HP	0.044697	0.005392	8.290187	1.67E-08	0.033569608	0.055825	0.03357	0.055825
Weight	-0.00309	0.000677	-4.56304	0.000126	-0.004488504	-0.00169	-0.00449	-0.00169
LB/HP	-1.47018	0.1757	-8.36755	1.41E-08	-1.832809821	-1.10755	-1.83281	-1.10755
Turbo	3.84889	1.1445	3.362943	0.002582	1.486757325	6.211023	1.486757	6.211023

Iteration #5 gives the following regression equation:

$$Y = 115.57 + 0.0447X_2 - 0.0031X_4 - 1.470X_5 + 3.849X_6$$

The R squared value decreased to 0.9783. Iteration #5 shows Turbo Status has the highest P-value at 0.002582 and the lowest absolute t Statistic of 3.36. This implies Turbo Status has the least effect on quarter mile trap speed among remaining explanatory variables. However, the R Squared value has decreased. This indicates the regression line in iteration #5 is a poorer fit than the regression line in iteration #4.

Summary

Iteration	Adjusted R Square	Standard Error	Regression F
#1 - 8 explanatory variables	0.9768	2.511	148.0
#2 - 7 explanatory variables - exclude engine displacement	0.9779	2.451	177.6
#3 - 6 explanatory variables - exclude transmission type	0.9788	2.396	177.6
#4 - 5 explanatory variables - exclude torque	0.9791	2.379	263.7
#5 - 4 explanatory variables - exclude city MPG	0.9783	2.511	316.5

In terms of acceleration, the above regression managed to debunk a number of automotive myths.

- One adage in the automotive world is that "there is no substitute for displacement." As engine displacement was the first explanatory variable removed, that old adage does no longer seem to be valid.
- Additionally, it was once believed that a manual transmission would be quicker than an automatic. As transmission type was the second explanatory variable removed, it shows that for this data set, a manual transmission is not necessarily quicker than an automatic transmission.

Iteration #4 proved to be optimal because we had the highest Adjusted R Square, the lowest standard error, and the highest F ratio. Iteration #4 gave us the following regression equation with 5 explanatory variables.

$$Y = 101.15 + 0.0505X_2 - 0.0018X_4 - 1.5377X_5 + 3.411X_6 + 0.4455X_8$$

Our model shows that for my data set, the explanatory variables are horsepower, weight, lb/hp, turbo status, and EPA city miles per gallon.