Regression models

Casper Christiansen

Sunday, July 19, 2015

How do I get most miles per gallon (MPG)?

A quantative approach

In this paper we will examine the relationship between mpg and various characteristics in cars. Our goal is to answer whether automatic og manual transmission is better for milage, and also try to investigate exactly how big of a difference there is.

We start off by doing a bit of exploration.

Table 1: Overview of the variables

Statistic	N	Mean	St. Dev.	Min	Max
mpg	32	20.090	6.027	10.400	33.900
cyl	32	6.188	1.786	4	8
disp	32	230.700	123.900	71.100	472.000
hp	32	146.700	68.560	52	335
drat	32	3.597	0.535	2.760	4.930
wt	32	3.217	0.978	1.513	5.424
qsec	32	17.850	1.787	14.500	22.900
vs	32	0.438	0.504	0	1
gear	32	3.688	0.738	3	5
carb	32	2.812	1.615	1	8

We see that we have 11 variables describing the 32 observations. We have a fair amount of variance in all of the variables, and no observations with missing data. We will assume MLR 1 through 6 is satisfied such that

1. The model is linear in parameters
2. We are dealing with a random sample
3. There is no perfect colinerity
4. The mean conditional variance is 0, such that $E[u|X] = E[u] = 0$.
5. There is no heteroskedascitícy
6. The error term is normally distributed.

We will start by examine the model $mpg_i\ \beta_0 + \beta_1 am_i + \beta_2 wt_i + \beta_3 hp_i$. In essence, we are interested in finding out whether tranmission type has any influence on milage. One would expect that the number of horse power has some influence, since more horse power almost always is associated with higher fuel consumption. Holding that constant, the weight of the car is also highly influential, since higher weight would probably mean higher consumption. One could argue that displacement (size of engine) could have an effect, but I think it is reasonable to assume that this is not the case. This is based on the fact that even large engines can have better mileage than smaller engines, if the smaller engine has been designed to output maximum amount of horse power, fx by fitting a turbo, which the data set does not include. Hence, *hp* is used.
The above mentioned is model 1 in the regression table. We start off by noticing that the controls *hp* and *wt* have the expected negative sign and is highly significant. Furthermore we see that *am* is positive, but insigficant. We interpret them as follows: An increase of 1000 lbs. will reduce the milage of a car by 2.88

mpg, holding transmission type and number of horsepower constant. Increasing the horsepower by 1 will reduce the milage by 0.037, holding everything constant. And lastly, on average, automatic transmission will increase the milage by 2 mpg, holding everything else constant. In plot 1 I have plotted *hp* and *wt* with *mpg*. It does seem like automatic transmissions get slightly higher milage, when visually inspecting it by *hp* and by *wt*, albeit we could not say anything from the regression.
We plot the squared residuals against the wt and see that we get a negative curve. It does not seem that the relationship is particularly strong however, and a Breusch-Pegan test reveals that there is no heteroskedasticity. Hence, our assumptions are probably reasonable.

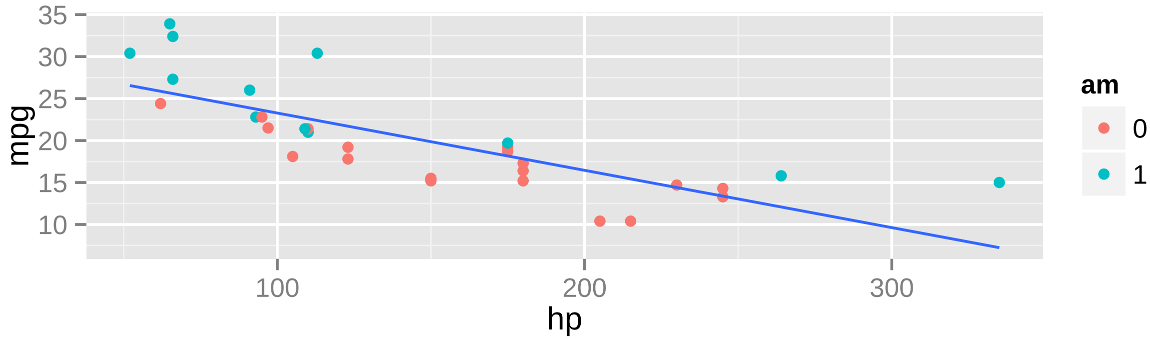
For a final test, we will examine whether there is *no* difference between the two groups (automatic vs. manual). We do this by creating interaction terms on all the variables, and test if the variables are jointly significant. The F-test reaveals a test-statistic of 3.772 and and p-value of 0.0226, and and such, we can rule out that they are indeed the same.
Looking at plot 2, it becomes quite apparent that the partial effect of *hp* is more or less the same, but the partial effect of weight is very different indeed. For light automatic cars, the milage is better, but this drops rather steeply, and much steeper than manuals. It should be noted that manuels in general are heavier, which makes for a hard comparison, but there do seem to be a difference, nevertheless.

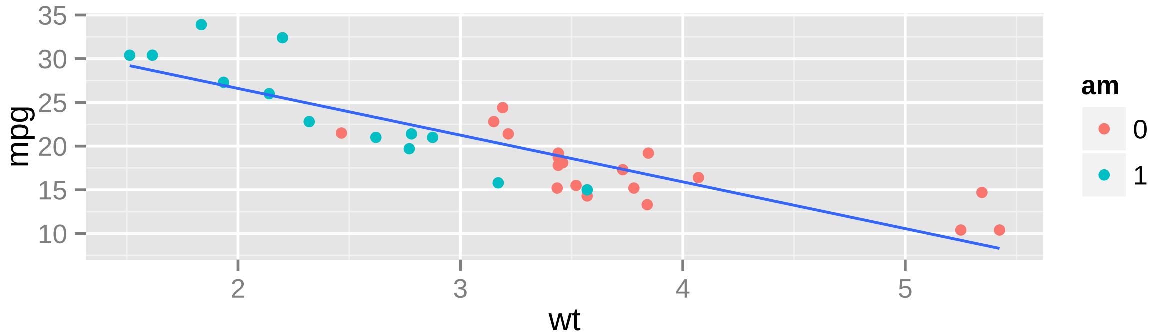
Table 2: Regression table

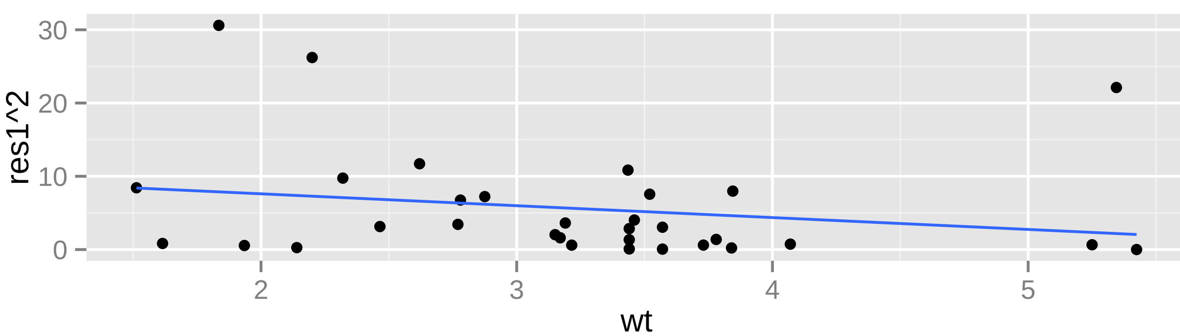
	Dependent variable:	
	mpg	
	(1)	(2)
am1	2.084	13.740***
	(1.348)	(4.223)
wt	-2.879^{***}	-1.856^{*}
	(0.892)	(0.945)
hp	-0.037^{***}	-0.041^{***}
	(0.008)	(0.014)
am1:wt		-5.769^{***}
		(2.072)
am1:hp		0.028
		(0.019)
Constant	34.000***	30.700***
	(2.597)	(2.675)
Observations	32	32
R^2	0.840	0.879
Adjusted R^2	0.823	0.856
Residual Std. Error	2.538 (df = 28)	2.286 (df = 26)
F Statistic	48.960*** (df = 3; 28)	37.890*** (df = 5; 26)
Note:		*p<0.1; **p<0.05; ***p<0.01

Appendix

Plot 1







Plot 2