Regression models

Casper Christiansen

Sunday, July 19, 2015

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

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 see from table 1 (in the appendix) 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 conditional mean is 0, such that $E[u|X] = E[u] = 0$.
5. There is no heteroskedasciticy
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, when controlling for *hp*. 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, using the standard $\alpha = 0.05$, (which is used throughout the report). Also, our R^2 is quite high, which supports the model by explaining a very large part of the observed variation. 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. 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, it is reasoable to assume that our model is not misspecified as such. This holds for all the models.

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, e.i. we test the model $mpg_i = \beta_0 + \beta_1 am_i + \beta_2 wt_i + \beta_3 hp_i + \delta_1 am_i wt_i + \delta_2 am_i hp_i$, if $\delta_1 = \delta_2 = 0$. The F-test reaveals a test-statistic of 4.2464 and and p-value of 0.0254, and and such, we can rule out that they are indeed the same. This could be done using Anova as well, but I prefer the F-test. Looking at plot 2, it becomes quite apparent that the partial effect of *hp* is more or less the same, but there is a distinct level difference, meaning that automatic gives higher milage for all levels of *hp*.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.
The third model regress the model with the intercepts at the average values, so for an averaged size car with an average amount of hp (146 and roughly 3200 lbs), we see that am has become virtually 0 and insiginificant. This basically means that for the average car, it does not matter for milage whether you choose manual or auto. The car will do around 18.7 mpg regardless. We should be fairly confident of this result, as it is, by definition, right in the center of our observations.

Table 1: Regression table

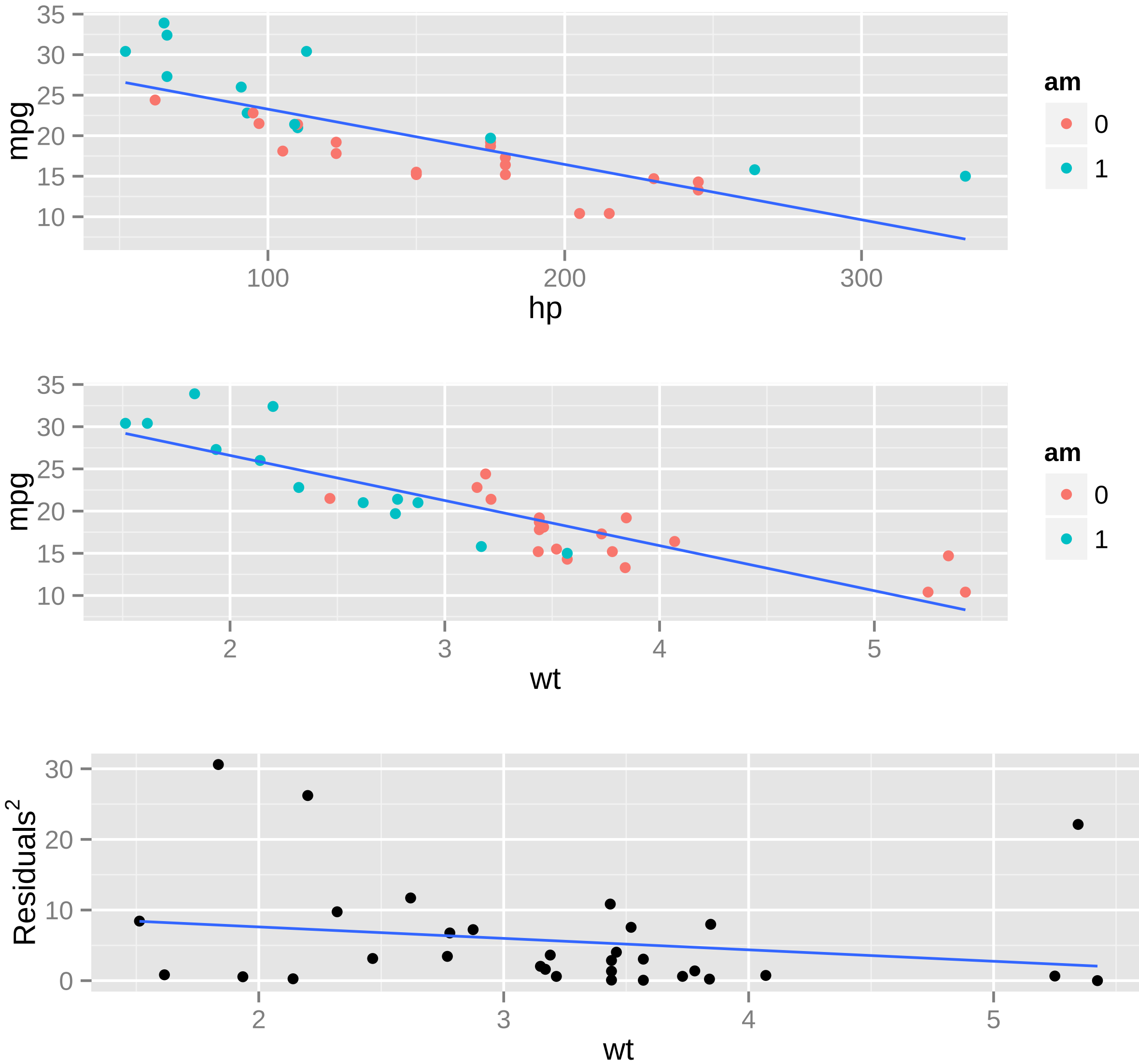
	Dependent variable:		
	mpg		
	(1)	(2)	(3)
am1	2.084	13.740***	−0.743
	(1.348)	(4.223)	(1.576)
wt	−2.879***	−1.856*	
	(0.892)	(0.945)	
hp	−0.037***	−0.041***	
	(0.008)	(0.014)	
am1:wt		−5.769***	
		(2.072)	
am1:hp		0.028	
		(0.019)	
I(wt - mean(wt))			−1.856*
			(0.945)
I(hp - mean(hp))			−0.041***
			(0.014)
am1:I(wt - mean(wt))			−5.769***
			(2.072)
am1:I(hp - mean(hp))			0.028
			(0.019)
Constant	34.000***	30.700***	18.730***
	(2.597)	(2.675)	(0.671)
Observations	32	32	32
R^2	0.840	0.879	0.879
Note:			*p<0.1; **p<0.05; ***p<0.01

Summary

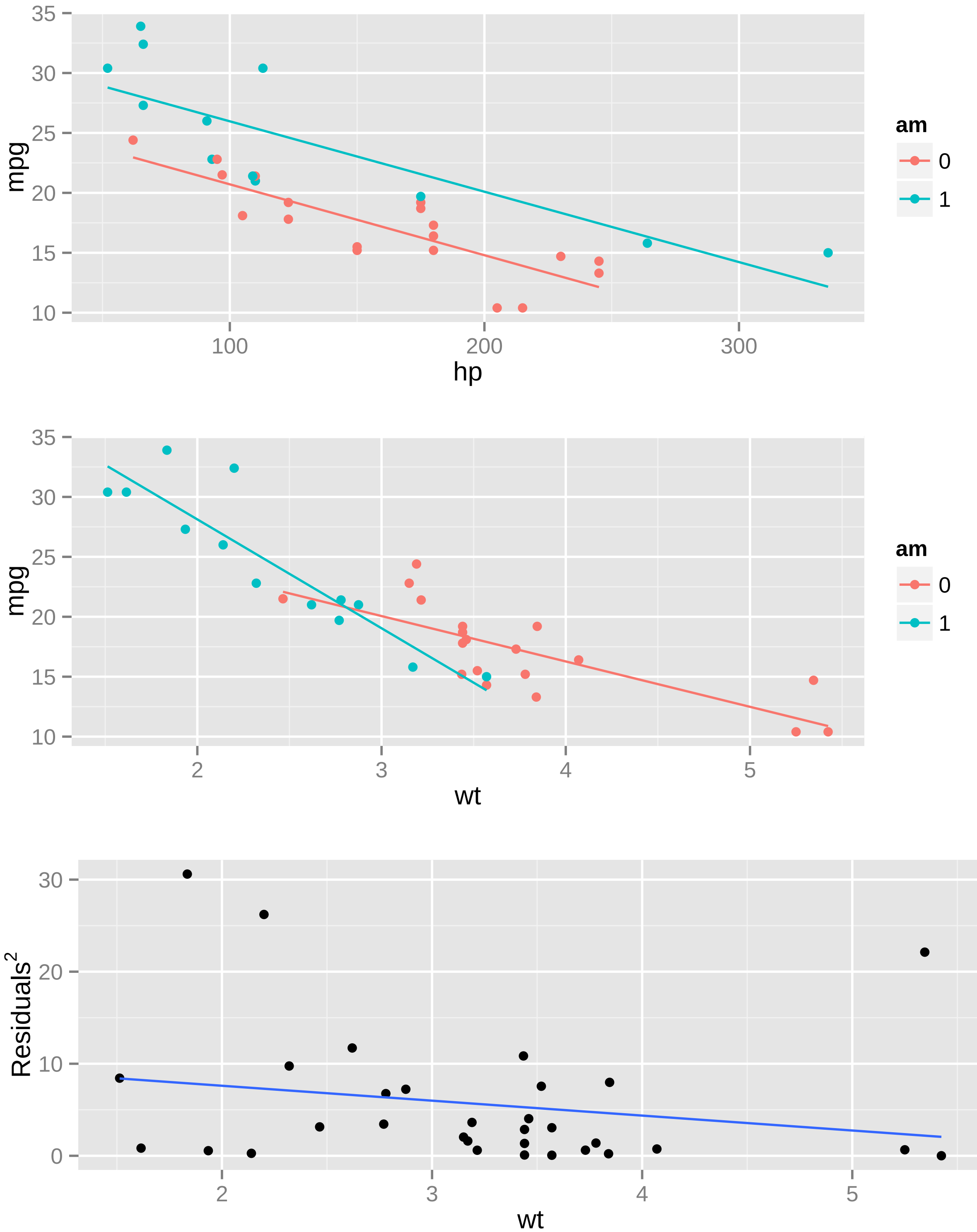
The takeaway from this seems to be that AM does seem to outperform manual, atleast for lighter cars. Up until around 2380 pounds, the automatic is most efficient (given amount of *hp*), after which it the manual becomes more efficient. We should however note that if if we evaluate at the averages, we have no clear winner. As such, if you buy an averaged size car, there is no difference between automatic and manual, your car will do around 18.7 mpg regardless. The weight of the car seems to have a significant impact when measuring the difference between manuals and auto. On average, having an automatic decrease your milage by 5.8 mpg *more* than the corresponding manual will. For better milage, one should avoid heavy automatic cars, and go for lighter ones instead. If you wish to purchase a large car, you should choose a manual.

Appendix

Plot 1



Plot 2



mpg
35
30
25
20
15
10
100
200
300
hp
am
0
1
mpg
35
30
25
20
15
10
2
3
4
5
wt
am
0
1
Residuals^2
30
20
10
0
2
3
4
5
wt

Table 2: 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
res1	32	0.000	2.412	−3.422	5.532
yhat	32	20.090	5.652	11.370	31.450

Table 3: Overview of the variables, manuals

Statistic	N	Mean	St. Dev.	Min	Max
mpg	19	17.150	3.834	10.400	24.400
cyl	19	6.947	1.545	4	8
disp	19	290.400	110.200	120.100	472.000
hp	19	160.300	53.910	62	245
drat	19	3.286	0.392	2.760	3.920
wt	19	3.769	0.777	2.465	5.424
qsec	19	18.180	1.751	15.410	22.900
vs	19	0.368	0.496	0	1
gear	19	3.211	0.419	3	4
carb	19	2.737	1.147	1	4
res1	19	0.000	1.970	−3.293	4.703
yhat	19	17.150	3.359	11.370	22.250

Table 4: Overview of the variables, auto

Statistic	N	Mean	St. Dev.	Min	Max
mpg	13	24.390	6.167	15.000	33.900
cyl	13	5.077	1.553	4	8
disp	13	143.500	87.200	71.100	351.000
hp	13	126.800	84.060	52	335
drat	13	4.050	0.364	3.540	4.930
wt	13	2.411	0.617	1.513	3.570
qsec	13	17.360	1.792	14.500	19.900
vs	13	0.538	0.519	0	1
gear	13	4.385	0.506	4	5
carb	13	2.923	2.178	1	8
res1	13	0.000	3.033	−3.422	5.532
yhat	13	24.390	5.641	12.820	31.450