Weight and horsepower are the most important factors for MPG

Stepan Kuntco steplg@gmail.com

Thursday, August 21, 2014

Summary

This paper describes analysis of the set of 32 automobiles (1973-74 models) to determine the relationship of the MPG and different aspects of automobile design and performance.

It appears that 88.47% of the variance in MPG can be described by automobile weight and horse power. For average horse power 146.68 increase in weight per 1000 lbs will reduce MPG by -4.1316, but this number declines with increase in horsepower and for 264 hp increase in weight per 1000 lbs will reduce MPG only by -0.8647.

While there is significant difference in MPG between automobiles with automatic and manual transmission (95% confidence interval is [3.2, 11.3]) but this factor by itself doesn't seems to be important for MPG prediction. It appears that automatic transmission is using mostly on automobiles with lower weight (95% confidence interval for difference in weight is [-1.86, -0.85]) and according to our model this leads to increase in MPG.

Exploritory analyses

The data was extracted from the 1974 Motor Trend US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973-74 models). Dataset contains following columns:

- **mpg** - Miles/(US) gallon
- **cyl** - Number of cylinders
- **disp** - Displacement (cu.in.)
- **hp** - Gross horsepower
- **drat** - Rear axle ratio
- **wt** - Weight (lb/1000)
- **qsec** - 1/4 mile time
- **vs** - V-engine or Straight engine
- **am** - Automatic/manual transmission
- **gear** - Number of forward gears
- **carb** - Number of carburetors

Summary for factor variables:

```
##  cyl                 vs             am     gear   carb
##  4:11   Straight engine:18   Automatic:13   3:15   1: 7
##  6: 7   V-engine       :14   Manual   :19   4:12   2:10
##  8:14                                      5: 5   3: 3
##                                                   4:10
##                                                   6: 1
##                                                   8: 1
```

Table 1: Correlation tables

Table 2: Most correlated variables pairs

Var1	Var2	value
wt	disp	0.8880
wt	mpg	-0.8677
mpg	disp	-0.8476
hp	disp	0.7909
mpg	hp	-0.7762

Table 3: Correlation with mpg

Var2	value
wt	-0.8677
disp	-0.8476
hp	-0.7762
drat	0.6812
qsec	0.4187

Model fitting

Variables with highest correlation with mpg are wt (weight), disp (displacement) and hp (horse power). But displacement is highly correlated with both weight and horse power, so let's start with horse power and weight as predictors:

```
lm1 <- lm(mpg ~ hp + wt, data=d)
```

Adding any one of the remaining variables gives a model which can't be distinguished from this one with the confidence level < 0.05, but residuals seems to be dependent of the fitted value. Adding interaction term corrects this issue.

```
lm2 <- update(lm1, '. ~ . + hp*wt')
```

This model is better than `lm1` with p-value 8.1083 $\times$ 10-4 and explains more MPG variation than the `lm1` model (88.4764% vs. 82.6785%).

Model analysis

`lm2` model coefficients:

```
## (Intercept)          hp          wt       hp:wt
##    49.80842    -0.12010    -8.21662     0.02785
```

In average MPG will decline with an increase in weight. For example for average horse power (146.6875) increase in weight per 1000 lbs will reduce MPG by -4.1316. But with increase in horsepower influence of weight becomes lower. Without considering outliner `Maserati Bora` with 335 horse power maximum horse power among other cars is 264. For automobiles with such horsepower increase in weight per 1000 lbs will reduce MPG by -0.8647.

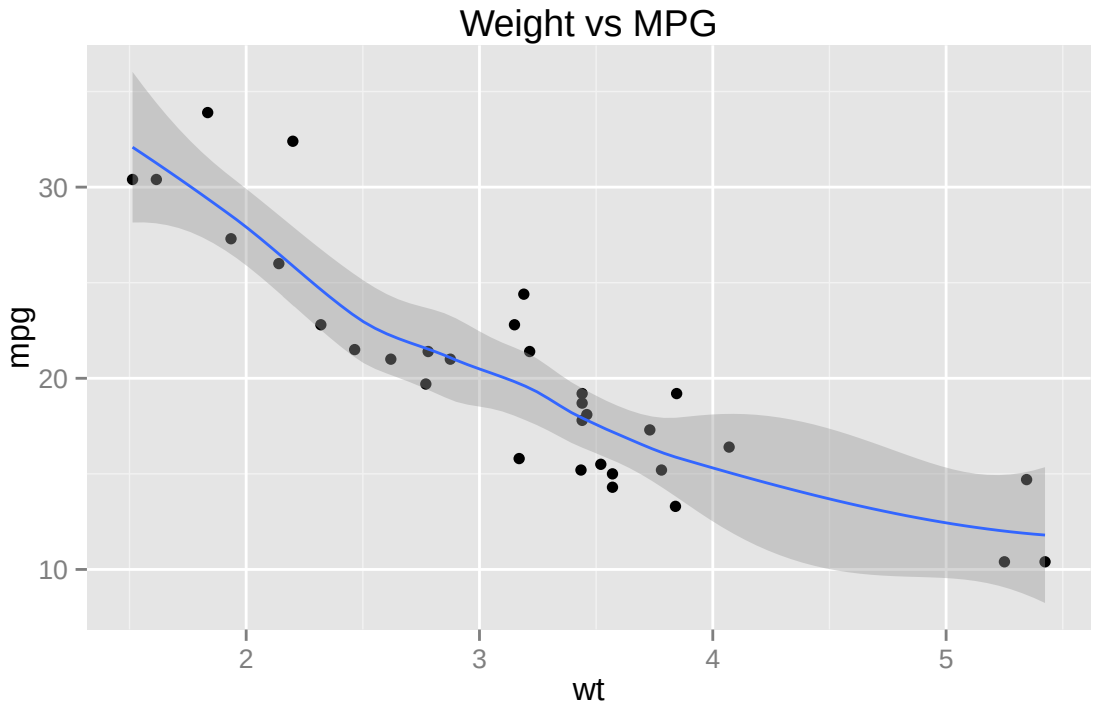
Transmission influence

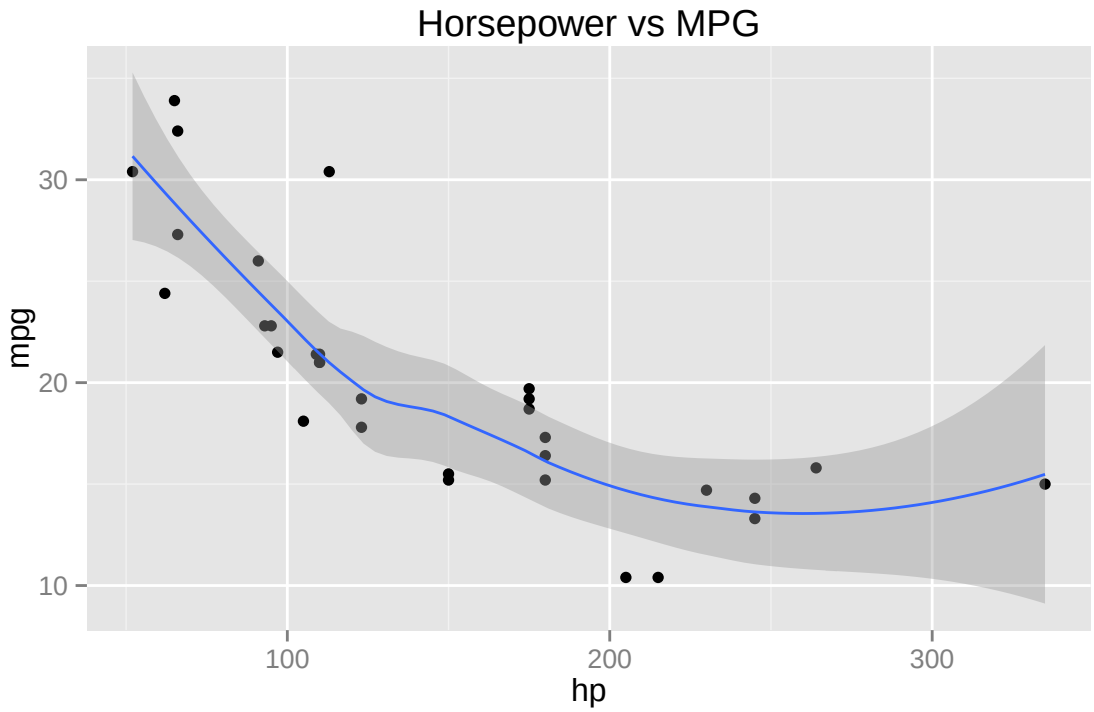
Overall confidence interval for MPG difference for vehicles with automatic/manual transmission is [3.2097, 11.2802] (with confidence score is 95%). That means that in average models with automatic transmission have higher MPG that the models with manual.

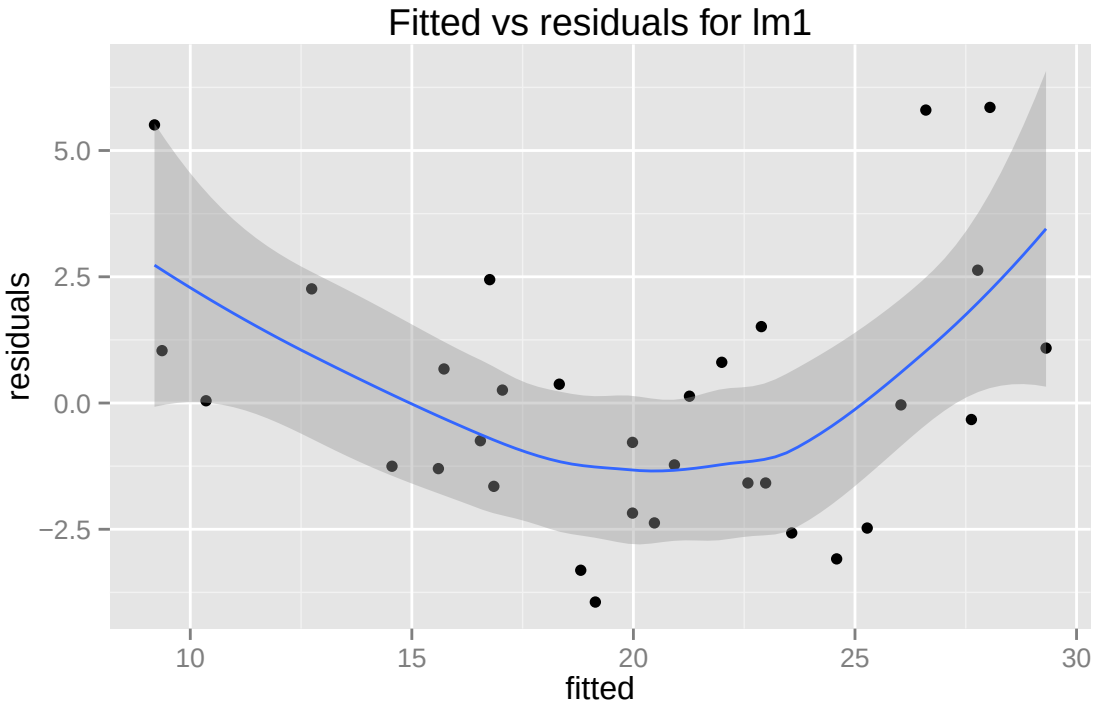
But this difference is explained by other factors (weight and horse power). Anova test for `lm2` model with added `am` factor has p-value 0.9259, what means that we can't distinguish `lm2` with and without `am` factor.

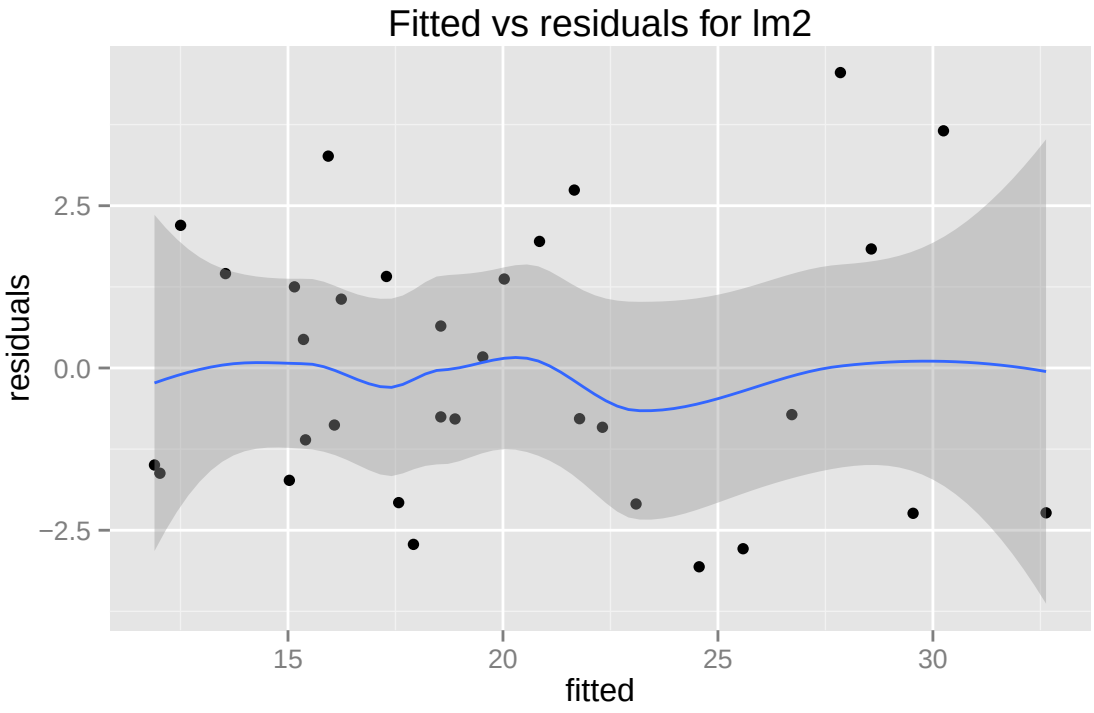
It appears that automatic transmission is using mostly on automobiles with lower weight (95% confidence interval for difference weight is [-1.8632, -0.8526]) and according to `lm2` model this leads to increase in MPG.

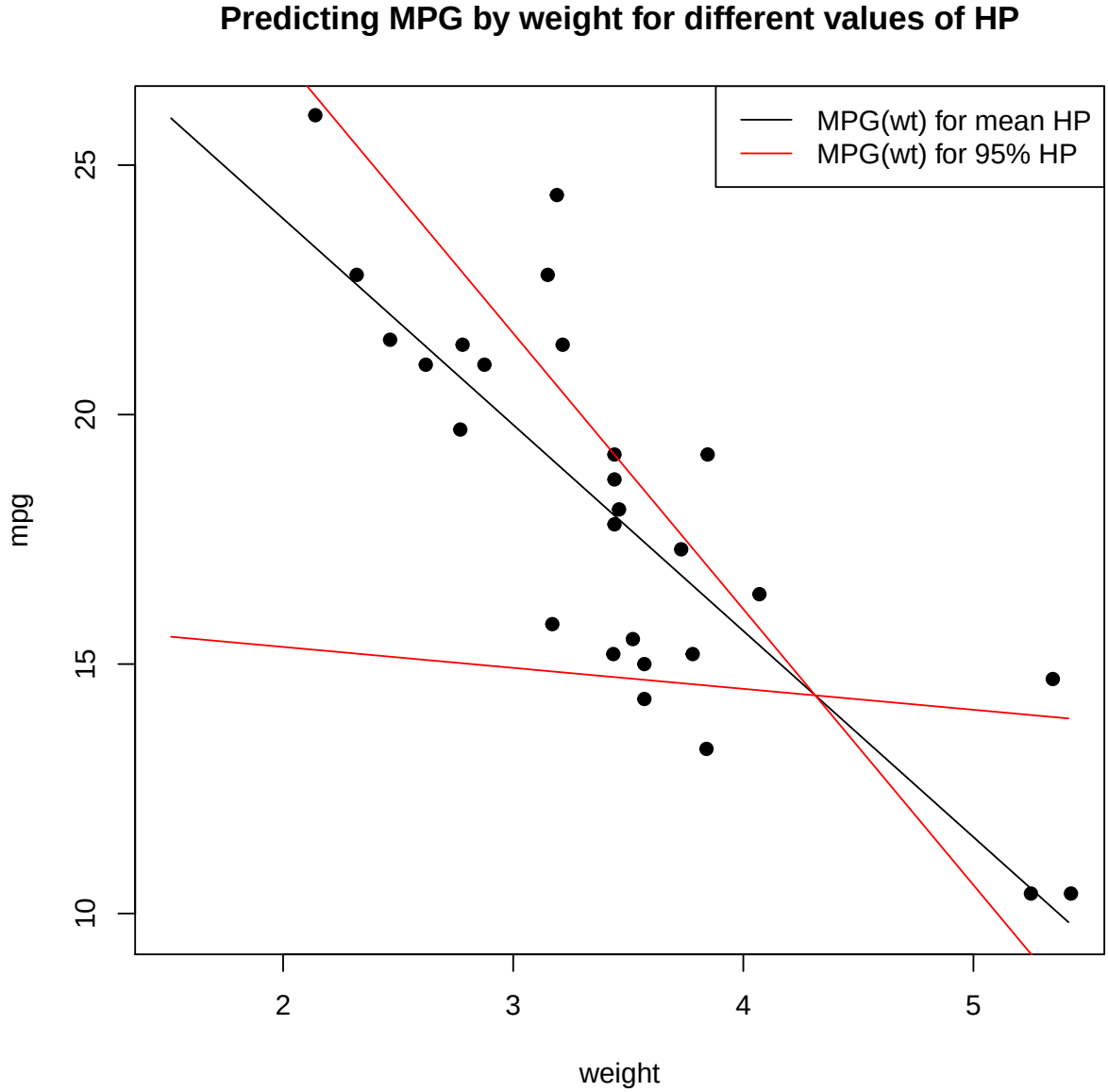
Appendix


Weight vs MPG
mpg
30
20
10
2
3
4
5
wt


Horsepower vs MPG
mpg
30
20
10
100
200
300
hp


Fitted vs residuals for lm1
residuals
5.0
2.5
0.0
−2.5
10
15
20
25
30
fitted


Fitted vs residuals for lm2
residuals
2.5
0.0
−2.5
15
20
25
30
fitted


Predicting MPG by weight for different values of HP
MPG(wt) for mean HP
MPG(wt) for 95% HP
mpg
10
15
20
25
2
3
4
5
weight