Motor Trend: Estimating MPG

Hugo Janssen

July 23, 2015

This edition of Motor Trend magazine explores the relationship between a set of variables and miles per gallon (MPG) for a collection of cars. In particular the following two questions are addressed:

- Is an automatic or manual transmission better for MPG?
- How to quantify the MPG difference between automatic and manual transmissions?

Source: `mtcars {datasets}` 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).

Executive Summary

The article starts with an exploratory analysis in which a baseline model is established based on transmission type alone. It is found that this model explains only 36% of the variance. In the data analysis that follows, more variables are introduced in the model. The proposed formula for the best fit is `mpg ~ 9.62 + 2.94 * am - 3.92 * wt + 1.23 * qsec` which explains 83% of the variance with a 95% confidence.

Data Overview

The data set consists of a data frame with 32 observations on 11 variables. The first 4 entries of the data set are shown in table 1.

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160	110	3.90	2.62	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160	110	3.90	2.88	17.02	0	1	4	4
Datsun 710	22.8	4	108	93	3.85	2.32	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258	110	3.08	3.21	19.44	1	0	3	1

Table 1: The first entries of the mtcars table

Exploratory Analysis

The first model to fit is an univariate model where MPG is predicted by the transmission type alone. This provides a benchmark against which other models can be evaluated. The resulting coefficients are listed in table 2.

```
fit0 <- lm(mpg ~ am, data = mtcars)
```

This model estimates an expected 7.24 increase in MPG for cars with manual transmission compared to cars with automatic transmission.

This claim is supported by the boxplot in figure 1 that shows the distribution of MPG per transmission type: cars with automatic transmission have an average of 17.15 MPG (which is equal to the intercept), compared to 24.39 MPG for cars with manual transmission.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	17.15	1.12	15.25	0.000000
am	7.24	1.76	4.11	0.000285

Table 2: Coefficients of a univariate model considering transmission type.

The predictor is significant, due to its small `p-value` and a [3.64, 10.85] confidence interval. However, with an `R-squared` of 0.36 it explains only 35.98% of the variance.

Data Analysis

According to Newton's law of physics, more force is needed to move objects with a higher mass. This suggests that the weight of a car is of influence on its MPG. The next model incorporates both the transmission type and the cars' weight:

```r
fit1 <- lm(mpg ~ am + wt, data = mtcars)
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	37.32	3.05	12.22	0.000000
am	-0.02	1.55	-0.02	0.987915
wt	-5.35	0.79	-6.79	0.000000

Table 3: Coefficients of a multivariate model considering transmission type and weight.

Indeed, the weight appears to be a good predictor as indicated by the very small `p-value`. Furthermore, the model explains 75.28% of the variance.

Another interesting observation is that one can see from the `p-values` in both models that holding the weight constant, transmission type appears to much have less of an impact on MPG than if weight is disregarded. This suggests that the weight of a car is far more important with respect to MPG than its transmission type.

After trying several other variables, `qsec` was selected as third argument in the model under the reasoning is that the faster the car, the higher its fuel consumption and thus the less MPG. Table 4 lists the variance table for the three models.

```r
fit2 <- lm(mpg ~ am + wt + qsec, data = mtcars)
```

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	30	720.90				
2	29	278.32	1.0	442.58	73.20	0.000000
3	28	169.29	1.0	109.03	18.03	0.000216

Table 4: Analysis of variance tables for the three models.

Figure 3 shows some minor residual heteroskedacity and the distribution plot of the residuals of manual cars (Figure 4) shows a slight shift just to the left of the mean, indicating that the residuals of the regression do not appear biased with mean close to 0.

Conclusion

The third model, including `am`, `wt` and `qsec` as variables explains 84.97% of the variance. Additional variables did not pass the statistical significance test and/or did not contribute to variance of the model.

Appendix 1: Figures

Figure 1. Boxplot of MPG per transmission type

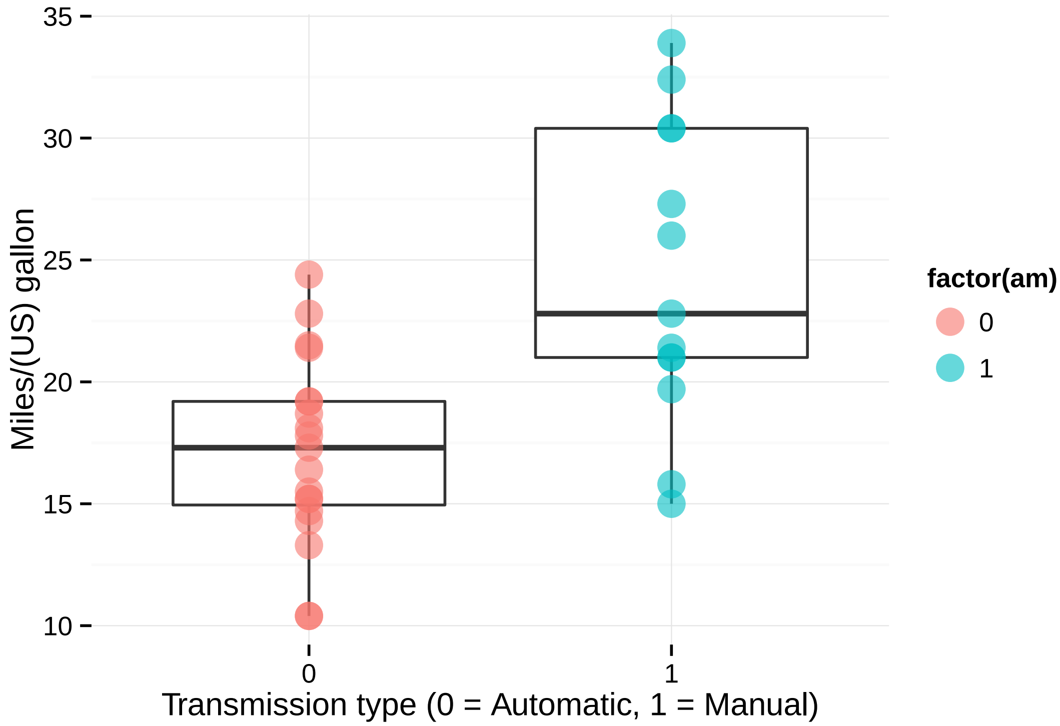

Figure 2. Regression line of MPG vs. weight per transmission type

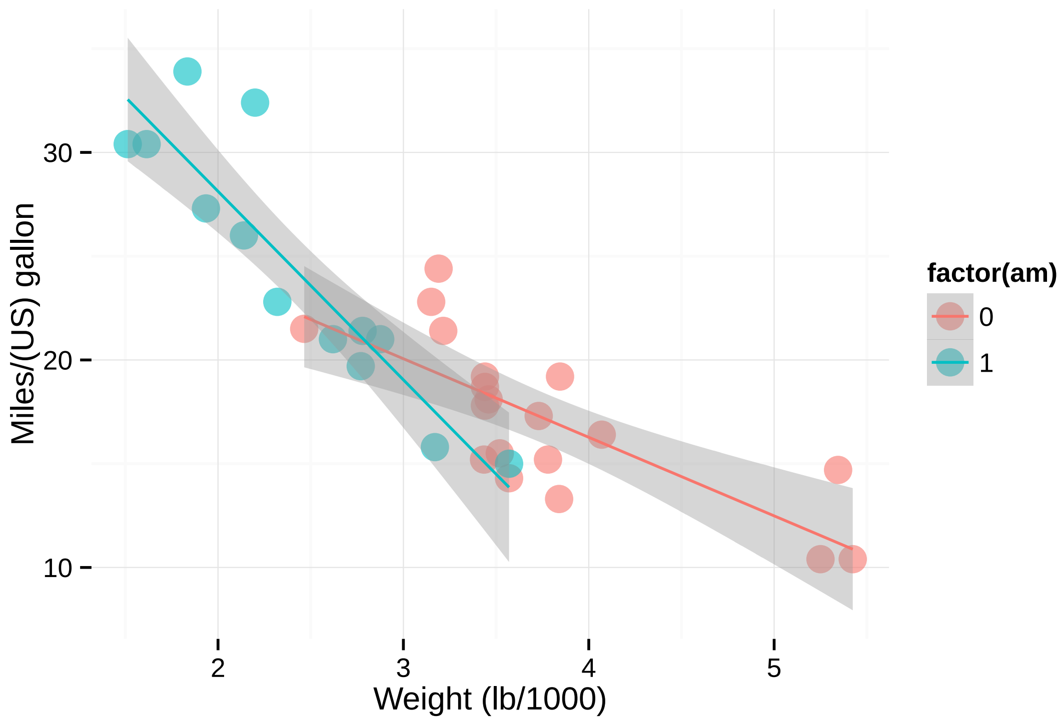

Figure 3. Residual plot

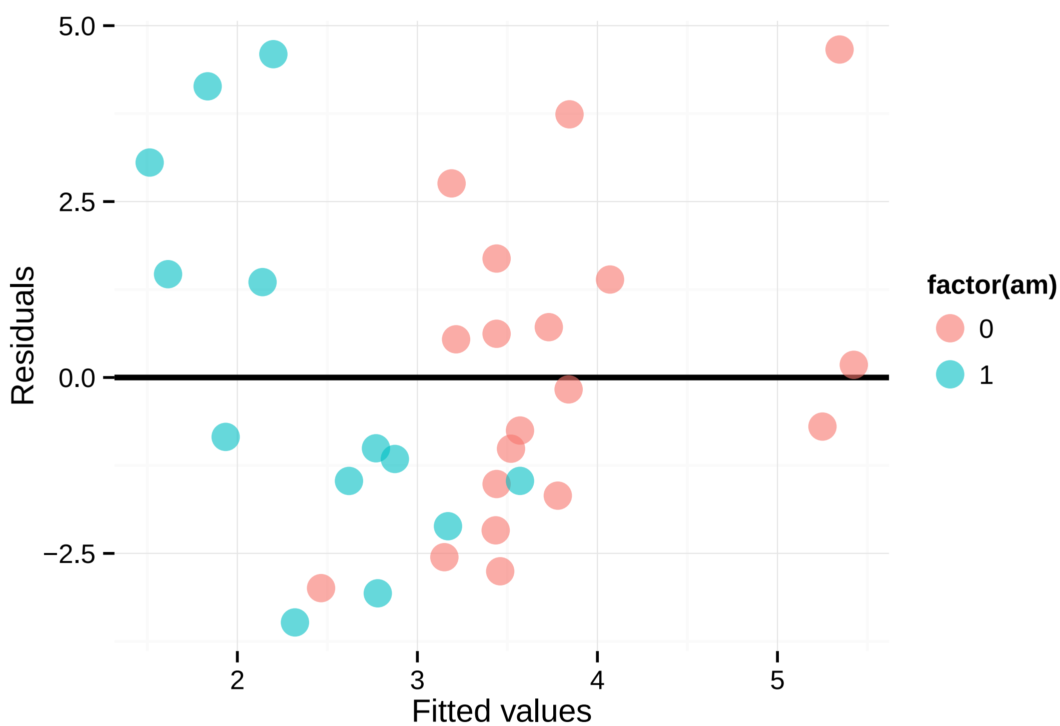

Figure 4. Residual histogram