Exploring the effect of various car variables on MPG

Marian Dragt

June 19, 2015

Summary In this report, we will analyze the Motor Trend Car Road Tests (mtcars) data set and explore the relationship between various variables and miles per gallon (MPG). This dataset 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). We assumed that there was a difference regarding performance between automatic and manual transmission, and found evidence for this assumption. Results reveal that cars with manual transmission have 7.2 MPG more than cars with automatic transmission. However, manual transmission alone was not the best predictor for MPG. Comparing several model, we found that weight, 1/4 mile time, and the interaction of weight and manual transmission together formed a stronger model regarding MPG.

We will address 2 **research questions**: "Is an automatic or manual transmission better for MPG" and "Quantify the MPG difference between automatic and manual transmissions".

Our **method** will start with loading and inspecting the data, and preprocess it when necessary. To address the first question, we will test the hypothesis "There is no difference between manual and automatic transmission". Hence, we will fit several linear models, using the backwards and forward modelling.

In our **analysis and results** we found confirmation that the data was normal distributed for both automatic and manual transmission, so we could use a t-test to test the hypothesis H0 "There is no difference between automatic and manual transmission". Given the results (p-value = 0.0014, df = 18,33, t = -3,77), the H0 can be rejected and accept the H1 hypothesis:"There is a difference between automatic and manual transmission"

In order to quantify these differences, we started with looking of at the effects of all the availabe variables on MPG. With this model we could explain nearly 91% of the MPG variance (RSE = 1.85, DF = 14), however, not all the used variables were significant (p-value smaller than 0.05). After applying stepwise techniques, we found our optimal model, resulting that transmission type, weight, 1/4 mile time and the interaction between weight and transmission type have a significant effect on MPG, explaining that a care with manual tranmission has 14.08 more MPG + 1.02 1/4 mile time, but the weight effects the MPG with -2.94 MPG and -4.14 MPG (interaction with transmission type) for every unit of car weight increase.

Hence, our **conclusion** that manual transmission is better for MPG, but it also depends on the weight of the car, and the 1/4 mile time, besides, the interaction between weight and the manual transmission type has a negative effect on the MPG. So, for light cars, manual tranmission type is better for MPG, but for heavy cars, this assumption is not supported.

Appendix

Data source and description The Motor Trend Car Road Tests **data** (mtcars) 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).

- 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

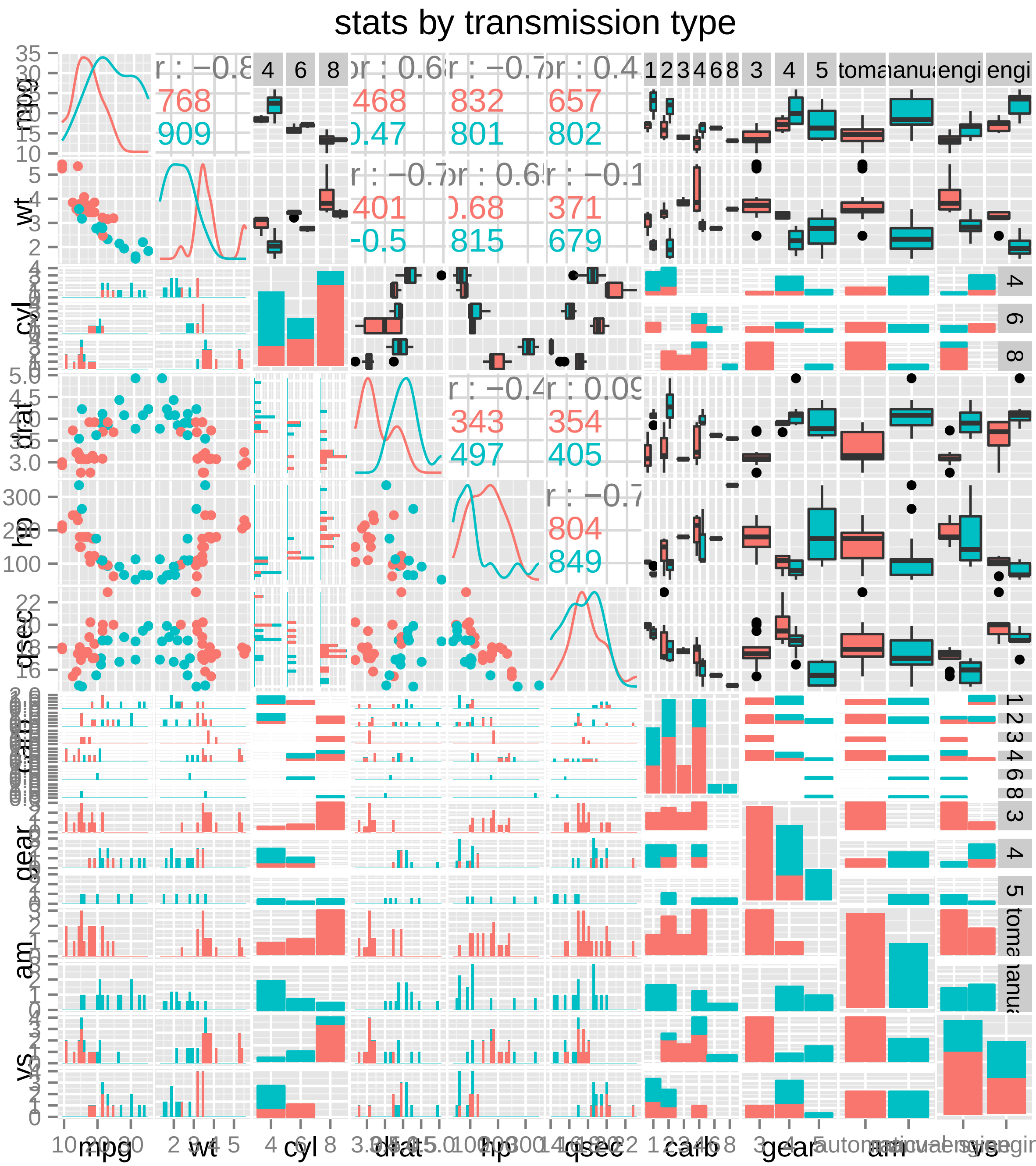
- vs V/S V-Engine / Straight engine
- am Transmission (0 = automatic, 1 = manual)
- gear Number of forward gears
- carb Number of carburetors

Source: Henderson and Velleman (1981), Building multiple regression models interactively. Biometrics, 37, 391-411.

Exploring the data Regarding the variable types, we have to transform cyl (Number of cylinders), vs (V/S V-Engine / Straight engine), am (Transmission (0 = automatic, 1 = manual)), gear (Number of forward gears), and carb (Number of carburetors) into factors.

We will also change the levels of transmission (0 = automatic, 1 = manual), and engine type (v-engine, s-engine) for clearer interpretation.

Hypothesis testing After checking for normality (0.8987358 and 0.5362729 are both not significant), we conducted a t-test, resulting in a p-value of 0.0013736, so we can conclude that there is a difference between manual and automatic transmission, rejecting the H0 hypothesis. There is a difference between these groups, showing that manual cars have 7.2449393 more MPG than automatic cars.



Exloratory Plots

Here we see the clear difference in transmission type, as manual cars have on average a higher MPG than automatic cars. Besides, we observe a relation between the transmission type and weight, as automatic cars tend to be heavier and have less MPG than manual cars. Actually, there are patterns in the data for nearly every variablea, accepting some noise.

But if it comes to correlation with MPG, than transmission type is not the best predictor, and probably models with the number of cylinders, weight, horsepower and displacement will do better.

Regression Models First we will test the **complete model**, that is, use all variables to predict MPG. With this initial step, we take the prior noticed correlations into account.

Our full model with additional possible confounding variables (RSE=1.42, DF=9) already explains nearly

0.944451 of the MPG variance, however, not all variables are significant. This can be also seen when looking at the confidence intervals, where only transmission typ, carb8, and the interaction term qsec:manual transmission are within the 2,5%-97,5% range. So, although the R-squared value is high, this is a messy model.

So we will try op to optimize this model using the stepwise approach, going backwards, eliminating variables step by step until reaching the best model. This also doesn't give a satisfying result, so will we use another strategy, starting with a **small model** that tests the regression of transmission type on mpg.

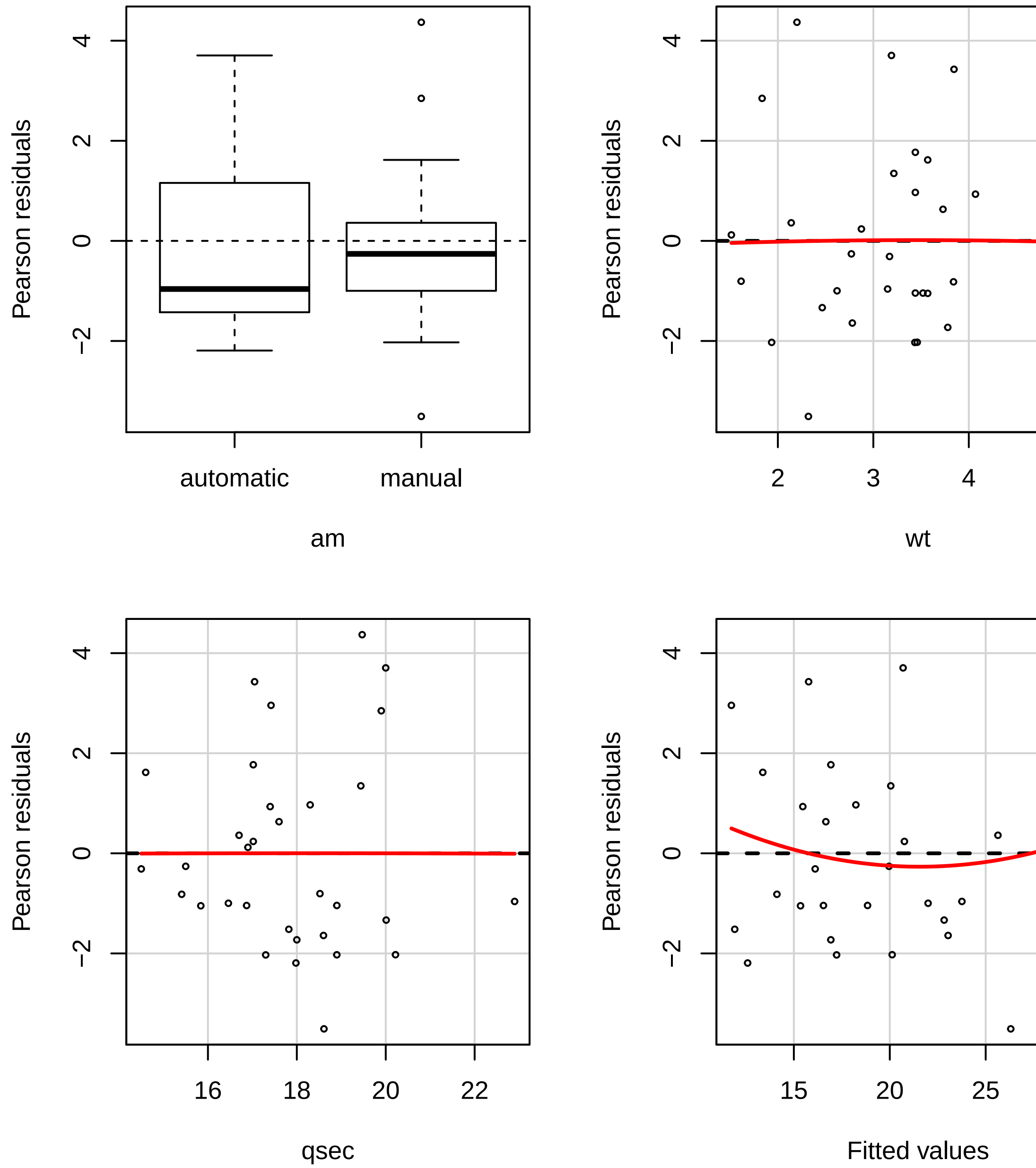
This model only explains 0.3384589 of the MPG variance (RSE=4.90, DF=30), so we use the same stepwise approach, but now going forward, including new variables by every step until reaching the best model. This gave a good result, and after deleting the non-significant horsepower, we obtained our **final model** (RSE=2.08, DF=27), with which we explain nearly 0.8804219 of the MPG variance. % latex table generated in R 3.2.0 by xtable 1.7-4 package % Sun Jun 21 14:03:42 2015

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	9.7231	5.8990	1.65	0.1109
ammanual	14.0794	3.4353	4.10	0.0003
wt	-2.9365	0.6660	-4.41	0.0001
qsec	1.0170	0.2520	4.04	0.0004
ammanual:wt	-4.1414	1.1968	-3.46	0.0018

% latex table generated in R 3.2.0 by xtable 1.7-4 package % Sun Jun 21 14:03:42 2015

	2.5 %	97.5 %
(Intercept)	-2.38	21.83
ammanual	7.03	21.13
wt	-4.30	-1.57
qsec	0.50	1.53
ammanual:wt	-6.60	-1.69

Inference and coefficients Regarding the final model, transmission type, weight, 1/4 mile time and the interaction between weight and transmission type have a significant effect. Looking at the confidence intervals, all confidence intervals are not including zero, confirming significance.



Residuals and Diagnostics
The plots of the residuals seem linear, which is ok. Also the diagnostic plots show normal results with some values diverse from the Normal QQ plot, however not drastically. Regarding the leverage, we see that Maserati Bora has more leverage, but this value is not surpassing the 0.5.

Comparing our final model with the model with only transmission type, we found evidence, using ANOVA, that our final model is better. We also found that all the predictors of the final model have significance. We didn't find outliers in the model, checking the leverage, nor did we found substantial influencers. [1] "Comparation of models" % latex table generated in R 3.2.0 by xtable 1.7-4 package % Sun Jun 21 14:03:42 2015

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	30	720.90				
2	27	117.28	3	603.62	46.32	0.0000

[1] “Significance testing of predictors” % latex table generated in R 3.2.0 by xtable 1.7-4 package % Sun Jun 21 14:03:42 2015

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
am	1	405.15	405.15	93.28	0.0000
wt	1	442.58	442.58	101.89	0.0000
qsec	1	109.03	109.03	25.10	0.0000
am:wt	1	52.01	52.01	11.97	0.0018
Residuals	27	117.28	4.34