Miles per gallon and its relationship with car variables: transmission.

Fábio Corrêa Didoné

Thursday, February 19, 2015

Executive Summary

This report will analyse the car variables, focusing in mile per gallon (MPG) and its relationship with others variables like weight, transmission, cylinders and others. The basic objective is to answer if the automatic cars have a higher MPG than manual cars. Linear models and correlation were used to analyse this question. Also, it was fitted a model than can predict the MPG of a car based on its weight, transmission, and second for a quarter mile. It was used analysis of variance, F-test, confidence intervals, variance inflation, normality test and other diagnostic to choose the best fitted model. The results show that manual car have a higher MPG than automatic cars. When the manual car was compared to automatic cars the difference in average was 7.2 increase in MPG for manual cars. When all variable analysed were fixed the manual cars had an increase in MPG of 2.52.

Objectives

Motor trend is interested in analysing the Miles per gallon relationship with other variables. Hence, we collected data about 32 cars, analysing the variables of design and performance. Special attention is paid to the variable transmissions (automatic or manual). The main objective of this article is to answer these questions:
Is an automatic or manual transmission better for MPG?
Quantify the MPG difference between automatic and manual transmissions?
It also will analysis the relationship of the transmission with another variables regarding the fuel consumption.

Exploratory Data analyses

The variables analysed were: Number of cylinders (cyl), displacement (disp), gross horsepower (hp), rear axle ratio (drat), weight (wt), seconds for one quarter of mile (qsec), v engine or straight engine (vs), transmission (am), forward gears (gear) and carburetors (carb)

The figure 1 shows the shape of the distribution of the MPG. The figure 2 shows the relationship of the MPG with the other variables regarding the transmissions. With these figures it is possible to observe that the mpg has a negative relationship with cylinders, displacement, horsepower, carburetors and weight. The mpg has a positive relationship with: Rear Axle Ratio, v engine or straight engine, seconds for one quarter of mile and it's seems there is no apparent relationship with gear.

The table 1 shows the actual magnitude of the correlation between the MPG and the other variables

Linear Models

Linear models can be used for a variety of functions.First, it was used to see the difference between the MPG in Automatic cars and Manual cars

The table 2 shows the difference in MPG between automatic and manual cars is 7.24 with t-value of 4.11,with the confidence interval for alpha = 5% between 3.64 and 10.85.TThe figure 3 shows the residuals have a pattern and are not normally distributed.

To know each variable relationship, it was made a linear models with all the variables as the regressors and MPG as the outcome. The table 3 informs the coefficients and the variance inflation for each regressor.

The table 3 shows first that when all the other variables are held constant, a manual car has 2.52 higher mpg than an automatic car. Other information that table 3 shows is the p value of the regressor, the variables with the lowest p value are wt, am, qsec. The variables with lowest variance inflation are drat, am and v/s, hence they are not correlated with other variable in the model

In order to do a regression line to explain the MPG variable. It was tested first the variables with the lowest p value as the regressor. After, it is was included the variables with low variation inflation. The regressors tested in the order were : am, wt, qsec, hp,drat,disp and cyl. The table 4 shows the ANOVA for the models tested

The model with wt, qsec and wt as regressor was chosen. The table 5 shows the coefficients of the model. All variables have a low p-value and the variance inflated is low. Also, when the variable weight and qsec are fixed the difference in mpg between a manual car and an automatic car is 2.94 with a confidence interval between 0.05 and 5.83

The table 6 shows some variables that diagnose the model chosen. The r^2 is high and the residuals are normal according the shapiro test. The figure 4 shows other diagnostic variables like the residuals plot against the fitted values and the hat values.

Assumptions

In order to realize this analysis some assumptions were made:

The variables are independent and identically distributed . The alpha used for all inference in the data was 0.05. The ANOVA assumes normality and constant variance of the variables.

Conclusions

The variables selected for the regression were transmission (am), the weight(wt) and second for one quarter of mile(qsec). Probably, because cusec is a performance variable that it is related with other variables like horse power, displacement and carburetors. While, weight is the variable most correlated with MPG. The transmission is a factor that can affect the MPG.
The comparison of automatic cars against manual cars showed that manual have 7.2 in average increase in the MPG. When all variables were fixed to account the manual cars had 2.52 increase in mpg. In conclusion, we can affirm that manual cars have higher MPG than automatic cars.

Notes

This file was produced with the knitr in the R studio, the code was omitted in order to reduce the length of the file.
The file with Rmd file is available in: https://github.com/fcdidone/reg_project .
.
.
.

Appendix

Tables

MPG	Cylinders	Displacement	hp	Rear Ratio	Weight	Qsec	V/S	Transmission	Gear	Carb
1	-0.85	-0.85	-0.78	0.68	-0.87	0.42	0.66	0.6	0.48	-0.55

Table 1: Correlation table

	Estimates	Std. Error	t value	Pvalue
Automatic	17.1474	1.1246	15.2475	0e+00
Manual	7.2449	1.7644	4.1061	3e-04

Table 2: Coefficient of the linear regression

	Estimate	Std. Error	t value	Pr(>|t|)	c(0, vif(f2))
(Intercept)	12.30	18.72	0.66	0.52	0.00
cyl	-0.11	1.05	-0.11	0.92	15.37
disp	0.01	0.02	0.75	0.46	21.62
hp	-0.02	0.02	-0.99	0.33	9.83
drat	0.79	1.64	0.48	0.64	3.37
wt	-3.72	1.89	-1.96	0.06	15.16
qsec	0.82	0.73	1.12	0.27	7.53
vs	0.32	2.10	0.15	0.88	4.97
am	2.52	2.06	1.23	0.23	4.65
gear	0.66	1.49	0.44	0.67	5.36
carb	-0.20	0.83	-0.24	0.81	7.91

Table 3: Coefficients of the linear regression and the variance inflation

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
am	30	720.897	NA	NA	NA	NA
wt	29	278.320	1	442.577	72.088	0.000
qsec	28	169.286	1	109.034	17.760	0.000
hp	27	160.066	1	9.219	1.502	0.233
drat	26	158.639	1	1.428	0.233	0.634
disp	25	150.093	1	8.545	1.392	0.250

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
cyl	23	141.206	2	8.887	0.724	0.496

Table 4: Anova with the model tested. Each row represents the variable included in the mdoel

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	9.62	6.96	1.38	0.18
amManual	2.94	1.41	2.08	0.05
wt	-3.92	0.71	-5.51	0.00
qsec	1.23	0.29	4.25	0.00

Table 5: coefficients and variance inflation for the model chosen

r^2	adj.r^2	shapiro.test
0.8497	0.8336	0.0804

Table 6: Diagnostic variables of the model chosen

Figures

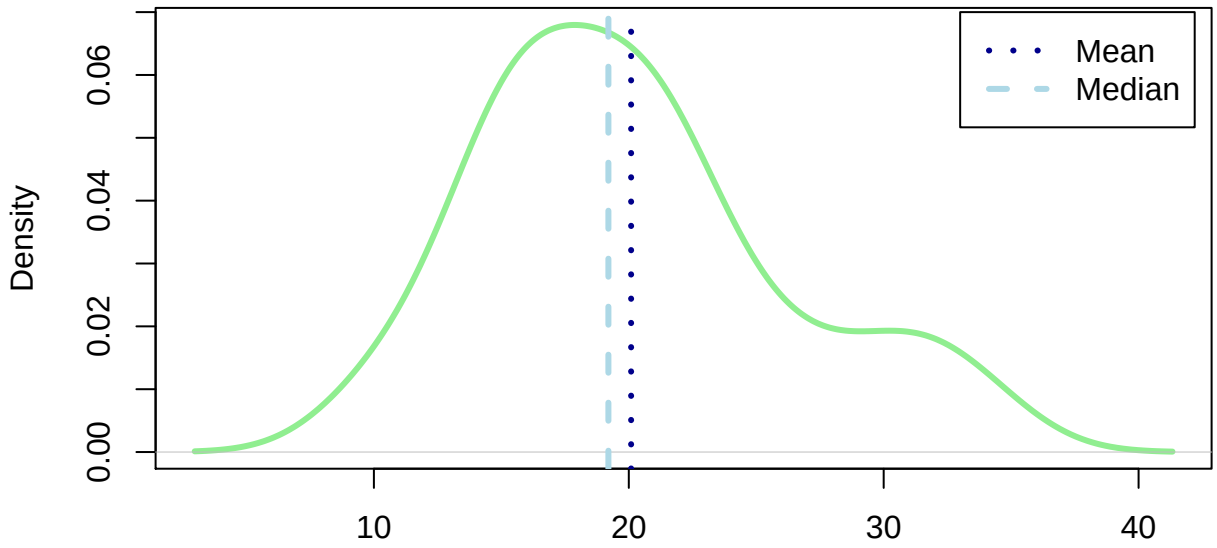

Figure 1: Density distribution of the MPG. The dotted line represents the mean and the dashed line represents the median

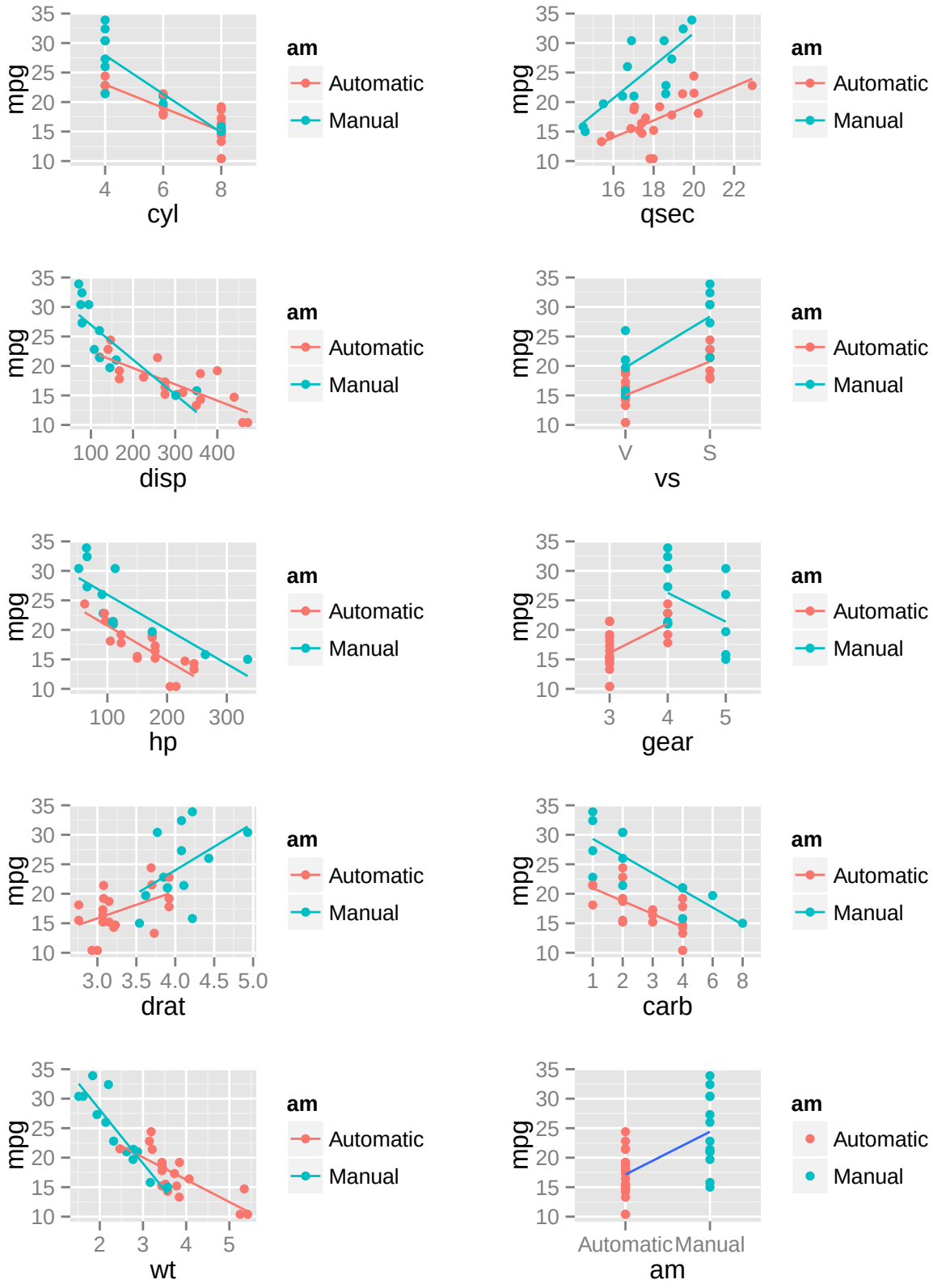

Figure 2: Panel with the relationship of MPG with all other variables. The regression line is plotted for each car transmission

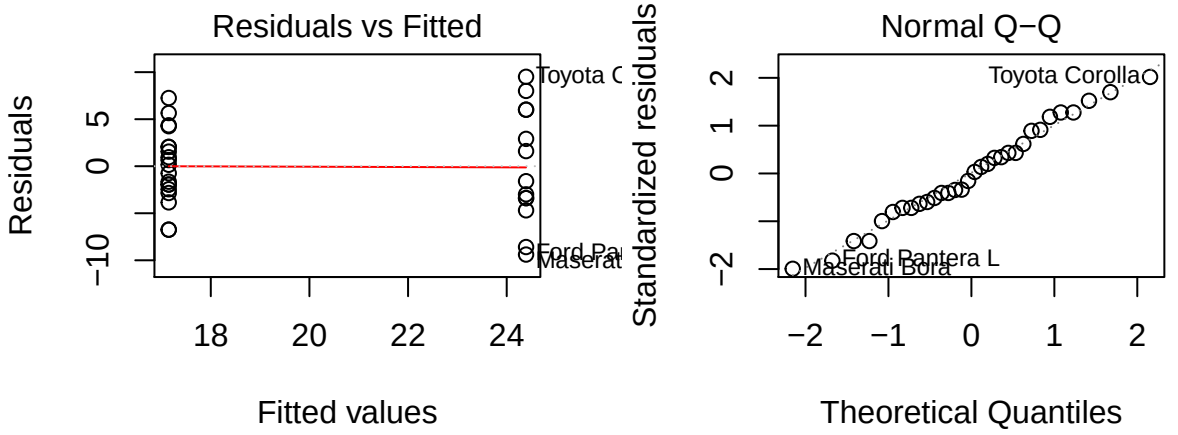

Figure 3: Diagonstic plots for the model

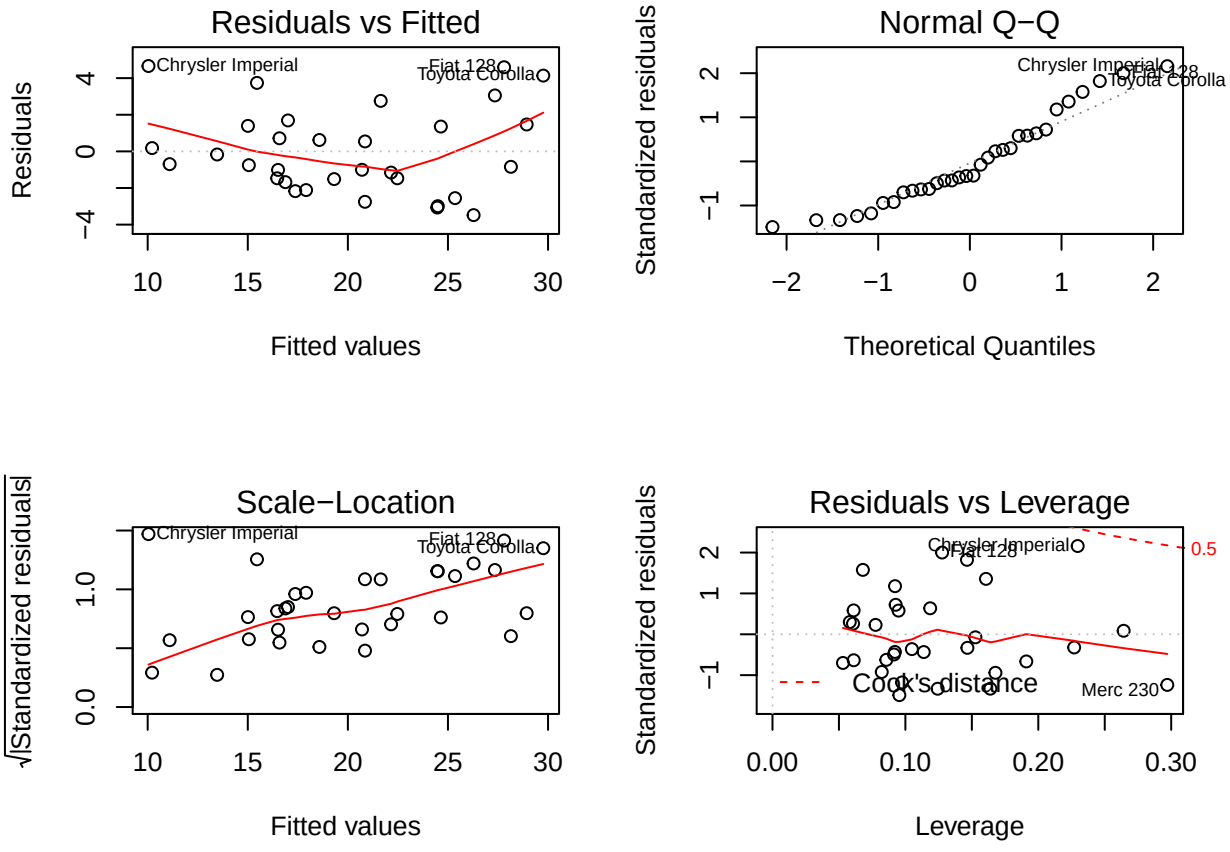

Figure 4: Diagonstic plots for the model chosen