

# Analysis of feuel consumption using multivariable regression

*Reza*

*20 May 2016*

## Summary

The goal of this study is to check whether automatic transmission has an impact on MPG in cars. We have a dataset of 32 cars with 11 measurement such as MPG, weight, displacement etc from each. We use regression analysis to check for the impact of transmission type on fuel consumption.

At the end we will see that, although it seems at the first look that automatic transmission decreases MPG but this is not an statistically meaningful conclusion.

## Data Explanation

The data was extracted from the 1974 Motor Trend US magazine, and includes fuel consumption (mpg), Number of cylinders(cyl), Displacement (disp), Gross horsepower(hp), Rear axle ratio (drat), Weight (wt), 1/4 mile time (qsec), V/S (vs), Transmission (am, 0 = automatic, 1 = manual), Number of forward gears (gear), Number of carburetors of (carb), for 32 automobiles (1973-74 models).

Comparing the MPG of automatic transmissible cars vs manually transmitted ones (17.1473684, 24.3923077) make us think that automatic transmission has a negative impact on fuel economy. But having a closer look at the data set (see Fig. 1), reveals that in particular in the case of weight(wt) and rear axle ratio (drat) variables, in one hand they both have impact on MPG and on the other hand, they can roughly separate automatic and manually transmitted cars. In other words more weight means less mpg also cars with automatic transmission are heavier in this data set. So we use multivariate regression analysis to face this lack of data in the dataset.

## Regression analysis

For start we fit a linear model for mpg vs all other variables and look at the coefficients of the model.

Table 1: Coefficients of linear model

(Intercept)	cyl	disp	hp	drat	wt	qsec	vs1	am1	gear	carb
12.3	-0.11	0.01	-0.02	0.79	-3.72	0.82	0.32	2.52	0.66	-0.2

It says that keeping other factors constant, the change from automatic transmission to manual increases mpg by 2.52 miles per gallon. But the confidence interval for this coefficient is (-1.76, 6.8) containing 0 which means that this improvement is not statistically significant. But there might be extra variables in the model which introduce extra variance. To seek for that we use the variable inflation factor (VIF). These are the VIF for the model.

Table 2: Variable inflation factors

cyl	disp	hp	drat	wt	qsec	vs1	am1	gear	carb
15.37	21.62	9.83	3.37	15.16	7.53	4.97	4.65	5.36	7.91

On the other hand one can look at the importance index of the variables included in the model.

Table 3: Importance index

	cyl	disp	hp	drat	wt	qsec	vs1	am1	gear	carb
Overall Importance	0	34.52	47.46	20.2	100	54.83	2.39	60.33	17.92	7.22

“cyl”, “vs” and “carb” variables have the least importance index and the quite high inflation factors. So there is a big chance that they are extra variables which might make bias in the perdition. Therefore we omit them and fit a new model

Table 4: Coefficients of the new linear model

(Intercept)	disp	hp	drat	wt	qsec	am1	gear
9.2	0.02	-0.02	0.81	-4.13	1.01	2.59	0.61

The “amt” coefficient is a bit bigger but not that much changed. It is still insignificant with 95% confidence interval of  $(-1.2, 6.38)$ . Let us take a look at the VIF’s of the new model.

Table 5: Variable inflation factor in the new linear model

disp	hp	drat	wt	qsec	am	gear
11.34	6	3.01	7.33	3.82	4.2	3.97

As the inflation of the variable “dis” is too high, we omit it as well for the last model.

Table 6: Coefficients of the third linear model

(Intercept)	hp	drat	wt	qsec	am1	gear
15.45	-0.02	0.65	-3.12	0.78	2.56	0

Does not show that much change in coefficient also the 95% confidence interval is  $(-2.31, 2.31)$ , which means non significant. But to get sure that omitting these variables was harmless, I ran a analysis of variance between two models. Here are the results

Table 7: Analysis of variance between three models

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
25	158.64	NA	NA	NA	NA
24	148.53	1	10.11	1.44	0.24
21	147.49	3	1.03	0.05	0.99

The P-value (columns  $\Pr(>F)$ ) between the third and second model is smaller than 0.95 which means that we shouldn't have omitted "disp". But the P-Value is high in comparing second and first models which means that omitting the variables "cyl", "vs" and "carb" was a good idea. So we stick on our second model with shows an *nonsignificant* 2.59 miles per gallon increase in MPG in manual transmitted cars.

To get sure about the validity of the last model I plot the residual plot (see Fig.2). As you can see there is no special trend in the residual plot. So the model appears to be valid.

## Conclution

Although a manually transmitted car seems to show better fuel economy, this improvement does not appear to be significant. It should be noted that this analysis is done on an old and small dataset, so is not that reliable. For assurance more investigation needs to be done.

## Appendix: Plots

Fig.1: Impact of different variables on MPG based on transmission type

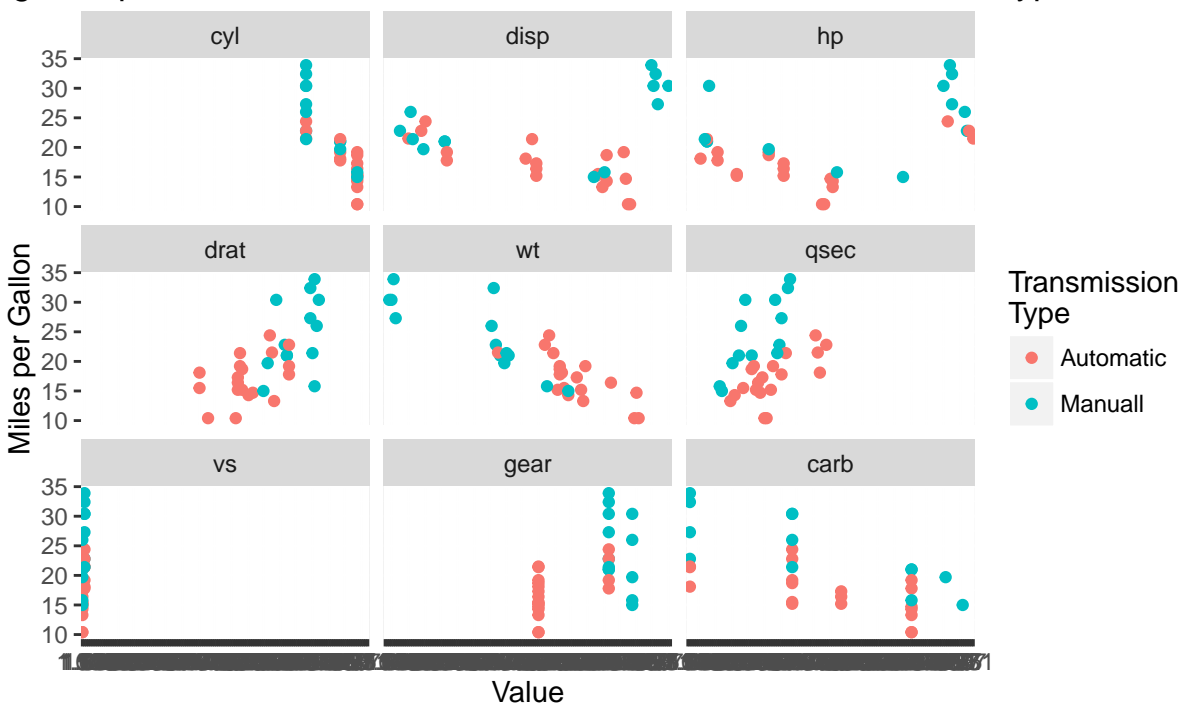
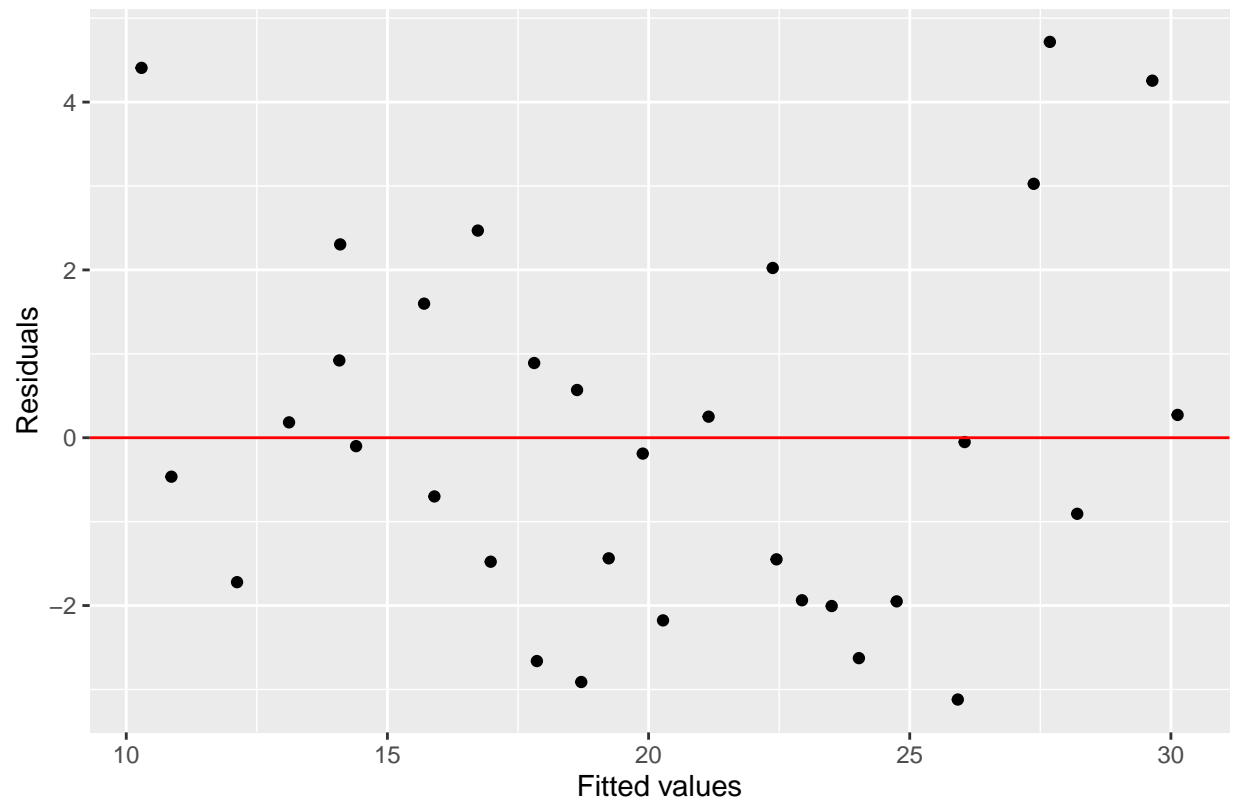


Fig.2: Residual plot



## Reference

- mtcars dataset in R.