

Relation between transmission type and miles per gallon

Martina Hoever

May 22nd - 2016

Data analysis on miles per gallon for automatic versus manual transmission

Executive summary

This report focuses on answering the following two questions:

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

The data analysis shows that automatic transmission is better for MPG than manual transmission. Taking weight and number of cylinders into account, having manual transmission will cause an increase in miles per gallon of 0.17 over having automatic transmission.

1. Overview and exploratory data analysis

*Note: To keep this report concise all code used in this report is hidden but can be found **here** in the .rmd file on GitHub*

This report focuses on a data analysis of difference in the miles per gallon for cars with automatic versus manual transmission. For this analysis the Motor Trend Car Road Tests dataset is used. This 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). A data frame with 32 observations on 11 variables. We are mostly interested in the variables ‘mpg’ and ‘am’, describing respectively the miles per (US) gallon and the transmission, with 0 being automatic and 1 being manual.

1.1 Load and summarize the data

The dataset mtcars will be loaded and summarized. Since we will be using the variable transmission as a factor, we will adjust this in the dataframe.

1.2 View the data

To get some insight in the distribution of the two variables that we are mostly interested in, we will plot these two variables in a violin plot.

Note: The violin plot can be found in the appendix

This plot shows us that manual transmission seems to have a higher MPG than automatic transmission. We will check this by calculating the mean for both transmission types. This results in a mean of 17.14 MPG for automatic transmission and 24.39 MPG for manual transmission.

To be able to answer the question “Is an automatic or manual transmission better for MPG” a t-test can be done to show if the difference shown in the previous plot is significant.

This shows that the true difference in means is not equal to 0, with a p value of 0.001374. Since we know that manual transmission has a higher mean than automatic, we now can conclude that manual is better for MPG than automatic.

2 Regression Analysis

We would like to see if we can create a better prediction model by considering more variables than only transmission and mpg. Therefor we will create a multivariable regression model.

2.1 Multivariable regression model

We will start by creating a multivariable regression model with all the variables.

By having all variables in the model, we see that not one of the variables has a p-value lower than 0.05. This implies that none of the variables show a significant result. To be able to create a better prediction model, we will have to remove some of the variables.

2.2 Evaluate correlation

To determine which variables should go into our model, we create a correlation matrix.

Note: The correlation graph can be found in the appendix

Looking at the correlation with mpg, we see that cyl, disp, hp, drat, wt, vs and am have the largest correlation.

2.1 Analysis of the variance

We will use analysis of the variance (ANOVA) to quantify the significance of adding additional regressors. After am, we will add the variables in order of their correlation with mpg. This results in 7 models each including one more attribute in the following order: am, wt, cyl, disp, hp, drat, vs.

From the result of this analysis we can see that from model 4 on the p-value becomes larger than 0.05. We will therefor choose the model with variables am, wt and cyl.

2.2.2 Evaluate coefficients

Firstly we will asses at this model by looking at the coefficients.

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	39.42	2.64	14.92	0.00
am1	0.18	1.30	0.14	0.89
wt	-3.13	0.91	-3.43	0.00
cyl	-1.51	0.42	-3.58	0.00

We now see that all the variables are significant, since their p-values are below 0.05. For the variable transmission this model implies that a manual transmission will cause an increase in miles per gallon of 0.17.

2.2.1 Evaluate residuals

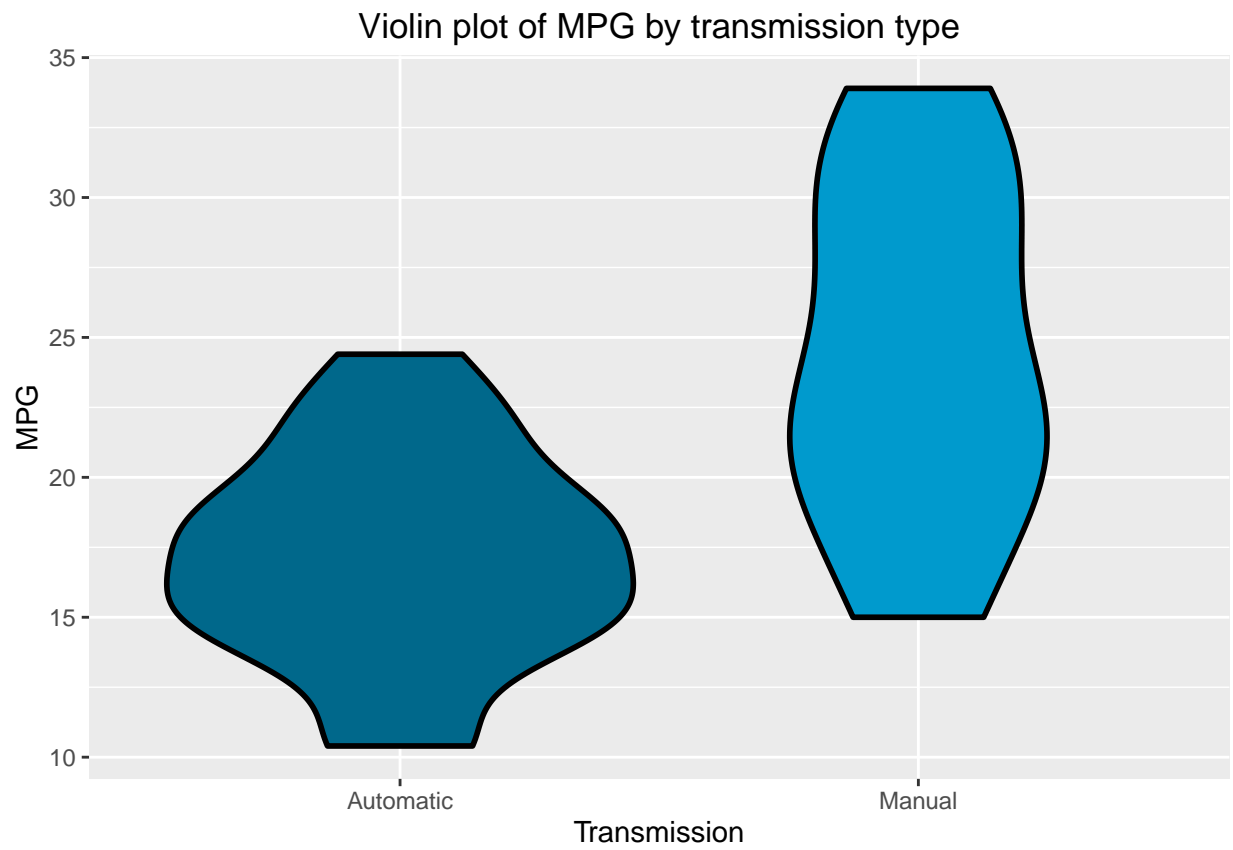
Analysis of the variance is sensitive to its assumption that model residuals are approximately normal. We will first take a look at these residuals in a plot, and then test if the residuals from are model are approximately normal by performing a Shapiro-Wilk test.

Note: The residuals plot can be found in the appendix

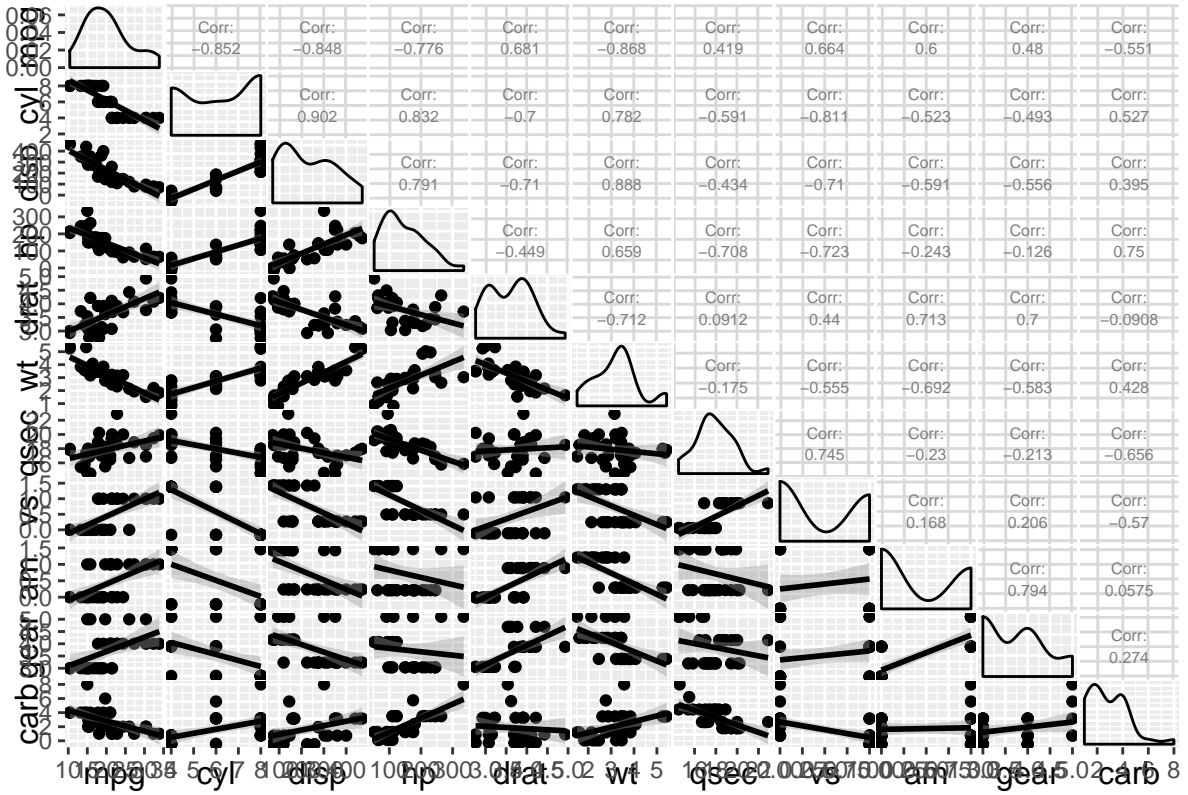
The plot shows that the residuals do not seem to follow a specific pattern. The Shapiro test P-value of 0.060108 is less than 0.1, which confirms the normality of the residuals.

3. Appendix

3.1 Violin plot of the variables transmission and mpg



3.2 Correlation graph



3.3 Plot of residuals

