Regression Models Project

Carlos Espino García

July 23, 2015

1 Overview

The purpose of this project is to analyze a data set of a collection of cars, and explore the relationship between a set of variables and miles per gallon. We are particularly interested in answering two questions: 1. Is an automatic or manual transmission better for MPG? 2. Quantify the MPG difference between automatic and manual transmissions

2 Dataset

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). The 11 variables are the following:

1. **mpg** Miles/(US) gallon
2. **cyl** Number of cylinders
3. **disp** Displacement (cu.in.)
4. **hp** Gross horsepower
5. **drat** Rear axle ratio
6. **wt** Weight (lb/1000)
7. **qsec** 1/4 mile time
8. **vs** V/S
9. **am** Transmission (A = automatic, M = manual)
10. **gear** Number of forward gears
11. **carb** Number of carburetors

3 Exploratory Analysis

To see the influence of the variables, especifically the transmission, in the miles per gallon, Figure 1 shows a matrix that illustrates the relation of the variables. We can appreciate that disp, hp, drat and wt have a strong linear correlation with mpg. We can appreciate as well, that the transmission seems to influence mpg, since the distribution of mpg is clearly divided by the type of transimission, we can see it more clearly in Figure 2.

4 Regression Analysis

To determine wich variables influence mpg and wether or not the transmission is important, we proceed to perform a regression analysis.

4.1 Model Selection

We first need to determine the method to use to select the "best" model. We are going choose the model by AIC in a backwards stepwise algorithm.

The Akaike Information Criteria (AIC) is a commonly used statistic to measures the relative goodness of fit. The AIC is a way of comparing different models against each other.

The backwards stepwise algorithm consists in the following:

1. Start with all candidate variables.
2. Test the deletion of each variable using a chosen model comparison criterion, in this case the AIC criterion.
3. Delete the variable (if any) that improves the model the most by being deleted.
4. Repeat this process until no further improvement is possible.

To perform the analysis, we take the transmission level "Automatic" (A) as the base level for the variable am. We define the variable $amM = I_{am=M}$.

After following the last steps, we get the following model.

$$\text{mpg} = 9.62 - 3.92\text{wt} + 1.23\text{qsec} + 2.94\text{amM} + \epsilon$$

The summary of the coefficients is shown in the next table

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	9.6178	6.9596	1.38	0.1779
wt	-3.9165	0.7112	-5.51	0.0000
qsec	1.2259	0.2887	4.25	0.0002
amM	2.9358	1.4109	2.08	0.0467

The steps of the model selection are not shown due to limitations of space

4.2 Coefficient Interpretation

We can interpret the coefficient as follows: 1. For each additional 1000 lbs. of weight, the miles per gallon decrease in 3.92. 2. If each quarter mile increase, the miles per gallon increase in 1.22. 3. The manual transmission increases the miles per gallon in 2.94 in average from the automatic transmission.

4.3 Residual Analysis

We need to perform a residual analysis to prove the assumptions of regression analysis. Figure **??** shows some plots that help for this analysis. We can see in the QQ-plot that the quantiles of th residuals against the theoretical residual approximately follow the reference line. The assumption of homoscedastic seems to be followed. We can appreciate as well that Fiat 128, Toyota Corolla and Chrysler Imperial are outlyers and need a deeper analysis.

5 Conclusion

According to the model selection criteria used, the variables wt, qsec and am are the most important to explain mpg. **We saw that the manual transmission is better for the performance of the car as it increases its miles per gallon in 2.94.**

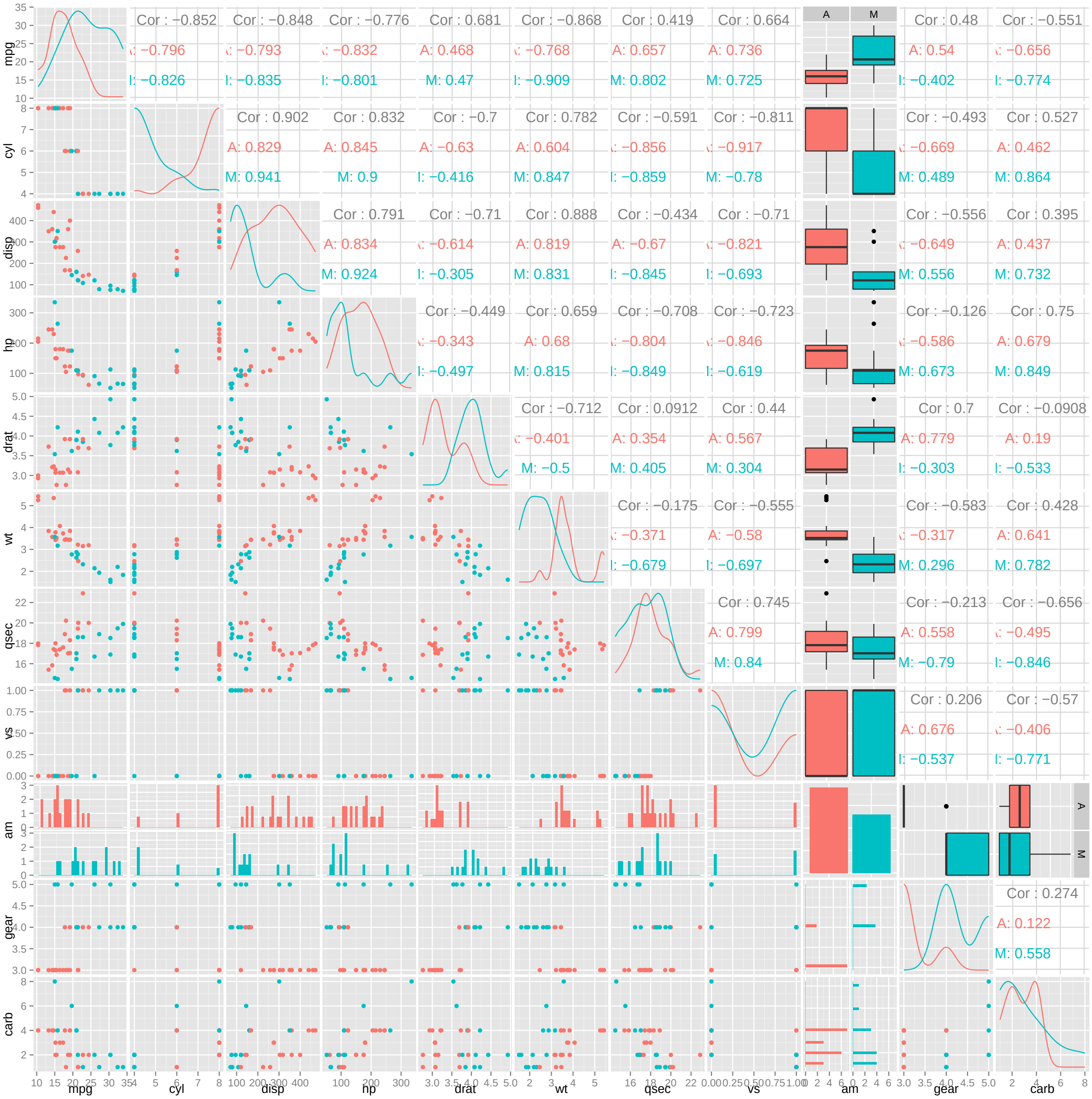

Figure 1: Correlations by transmission

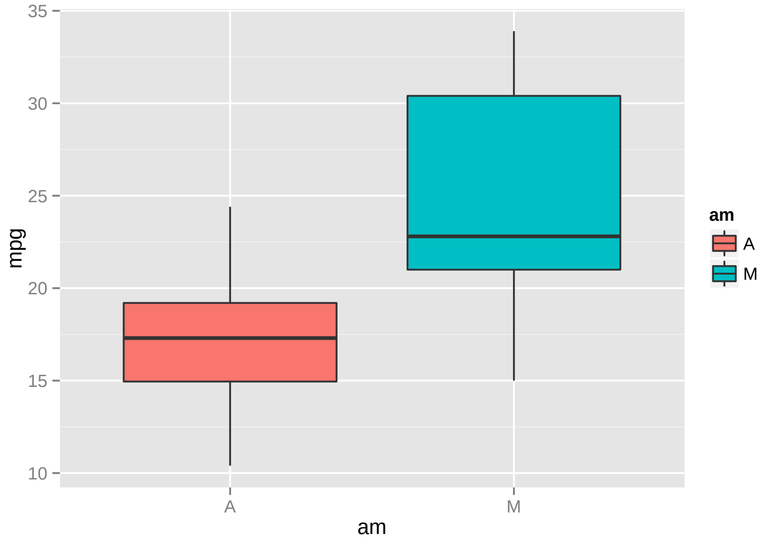

Figure 2: mpg by transmission

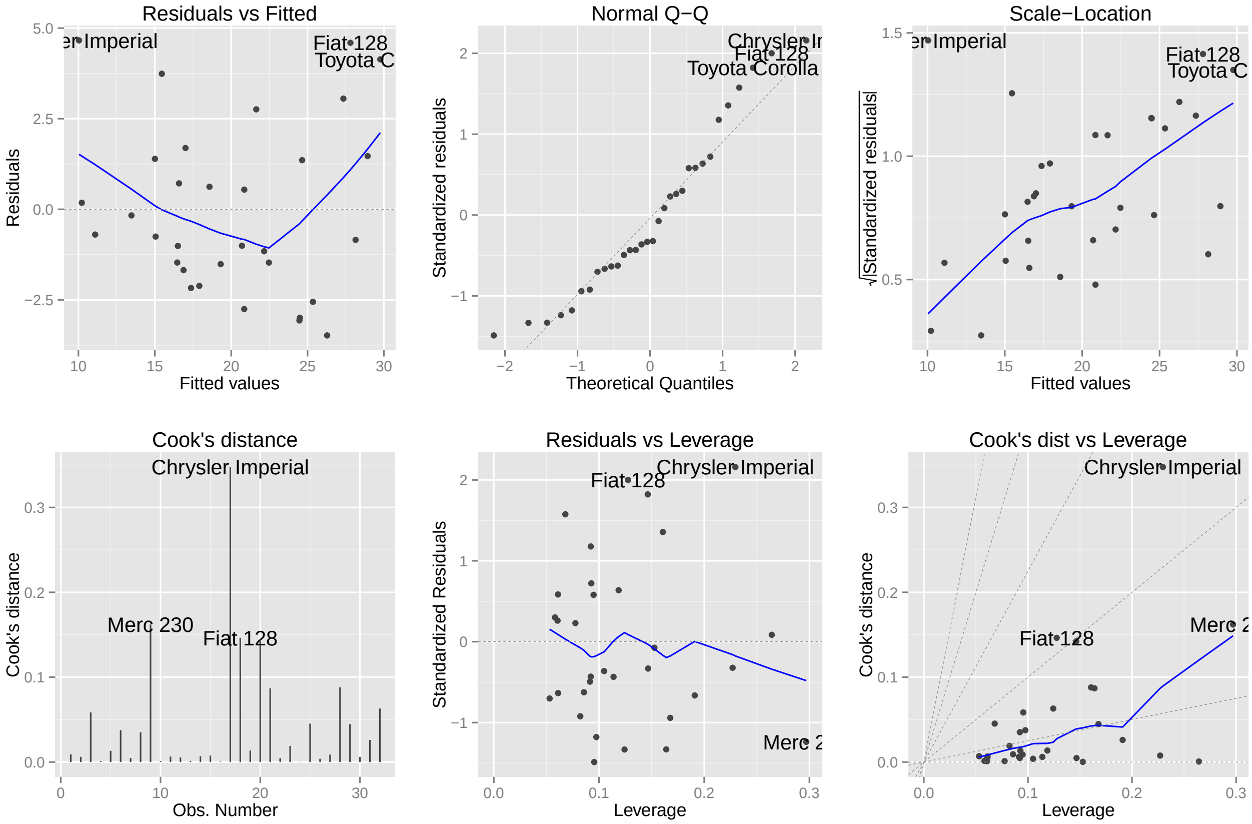

Figure 3: Residual analysis plots