

Assessing Factors that Affect mpg

PilotGtec EFTL

21 January 2015

Analyze mtcars data. The dataset has 32 observations on 11 variables, namely `mpg` (miles per gallon), `cyl`, `disp`, `hp`, `drat`, `wt`, `qsec`, `vs`, `am`, `gear`, and `carb`. Details on the variable descriptions can be found [here](#). In particular, the automobile industry is interested in the following:

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

Executive Summary

We explore the `mtcars` data and determine the effects of various factors to the `mpg` values of automobiles, particularly focusing on the type of transmission (as an initial approach). Three regression models are set up and compared (using anova).

- Model 1: Transmission type only
- Model 2: Transmission and weight
- Model 3: Stepwise algorithm

Results show that **Model 3** (`lm(formula = mpg ~ cyl + disp + wt + am, data = mtcars)`) is the most optimal. **Model 3** captures 86.588% of the the total variance (84.0088%, adjusted). From the coefficient values, a 1000-lb increase in an automobile’s weight would result to a -2.4968 change in the `mpg`. With respect to the transmission, automatic transmission has 1.8092 less `mpg` than manual. For `hp`, on the other hand, an increase would result to a -0.0321 (decrease) in `mpg`. Finally, in general, more cylinders (`cyl`) would decrease the automobile’s `mpg`. Analysis of residuals is also presented in this report.

Exploratory Data Analysis

First we look at the different relationships of each of the pair of variables in the dataset by plotting the pairs (Figure A1, *Appendix*). We also mark those x - y pairs that fit ($r^2 > 0.700$) into a linear model. In the plot, we have excluded `cyl`, `vs`, `am`, `gear`, and `carb` as they are categorical variables— we plot them independently (Figure A2, *Appendix*). The plots give us an idea about the correlations of the variables (*Note: categorical variables are considered numerical*). To quantify the correlations, we implement the `cor` function with `wt` (-0.8677), `cyl` (-0.8522), `disp` (-0.8476), and `hp` (-0.7762) showing the highest correlation values.

Question 1: “Is an automatic or manual transmission better for MPG”

From Figure A2, we see that there is a relationship between MPG and the type of transmission an automobile has: **Automatic** has lower `mpg` than **Manual** with mean values 17.1474 and 24.3923, respectively. To further **quantify the statistical significance of the difference**, we conduct a *Welch Two Sample t-test* on the categories with a 95% confidence interval. This results to a **p-value** of 0.0014 suggesting that indeed **Automatic** has lower `mpg` than **Manual**. We have to note, however, that there are other variables that need to be considered as well. This is where the regression models come in.

The Models

Note: Categorical variables are converted from numerical to factor.

In this section, we explore three models listed below. Here, we excluded the model (**Model 0**) involving all variables. This was decided after implementing the `anova` function comparing it against the model (**Model 3**) that resulted from a stepwise algorithm performed on **Model 0**. Even though **Model 3** is not significantly better ($\Pr(>F) \sim 0.9588$) than **Model 0**, it has fewer variables.

- Model 1: Transmission type only
- Model 2: Transmission and weight
- Model 3: Stepwise algorithm

Comparing the Models

Computing an anova for the three models shows that **Model 3** is the most optimal model compared to **Model 2** and **Model 1**.

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ wt + am
## Model 3: mpg ~ cyl + hp + wt + am
##   Res.Df RSS Df Sum of Sq    F Pr(>F)
## 1      30 721
## 2      29 278  1      443 76.2 3.3e-09 ***
## 3      26 151  3      127  7.3  0.001 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Model 3 captures 86.588% of the the total variance and 84.0088% (adjusted). From the coefficients, we can see that a 1000-lb increase in an automobile's weight would result to a -2.4968 change to the `mpg`. With respect to the transmission, automatic transmission has 1.8092 less mpg than manual. In terms of `hp`, an increase would result to a -0.0321 (decrease) in `mpg`. Finally, in general, more cylinders (`cyl`) would decrease the automobile's `mpg`.

```
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) 33.70832    2.60489  12.9404 7.733e-13
## cyl6        -3.03134    1.40728  -2.1540 4.068e-02
## cyl8        -2.16368    2.28425  -0.9472 3.523e-01
## hp          -0.03211    0.01369  -2.3450 2.693e-02
## wt          -2.49683    0.88559  -2.8194 9.081e-03
## am1         1.80921    1.39630   1.2957 2.065e-01
```

In the Appendix section (Figure A3), we analyze and plot the residuals against the fitted and leverage values for **Model 3**. We also provide the Normal Q-Q plot and Scale-Location plots.

Appendix

Figure A1

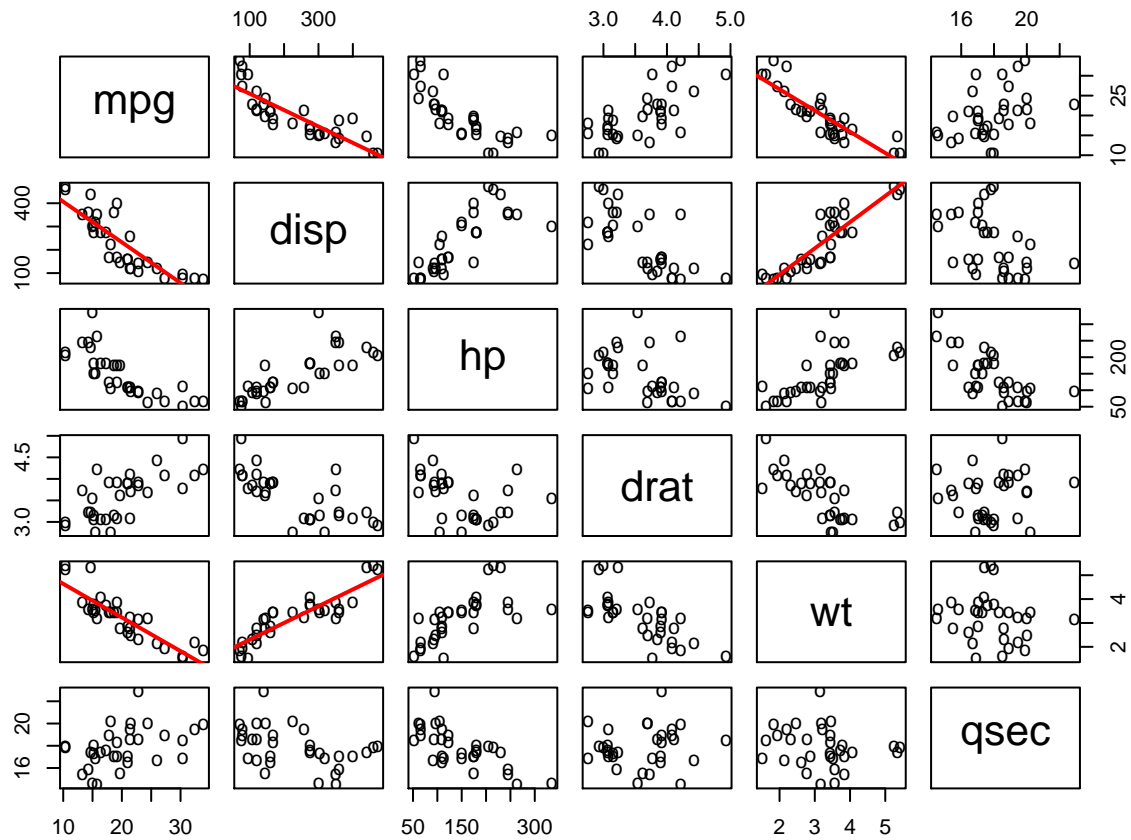


Figure A2

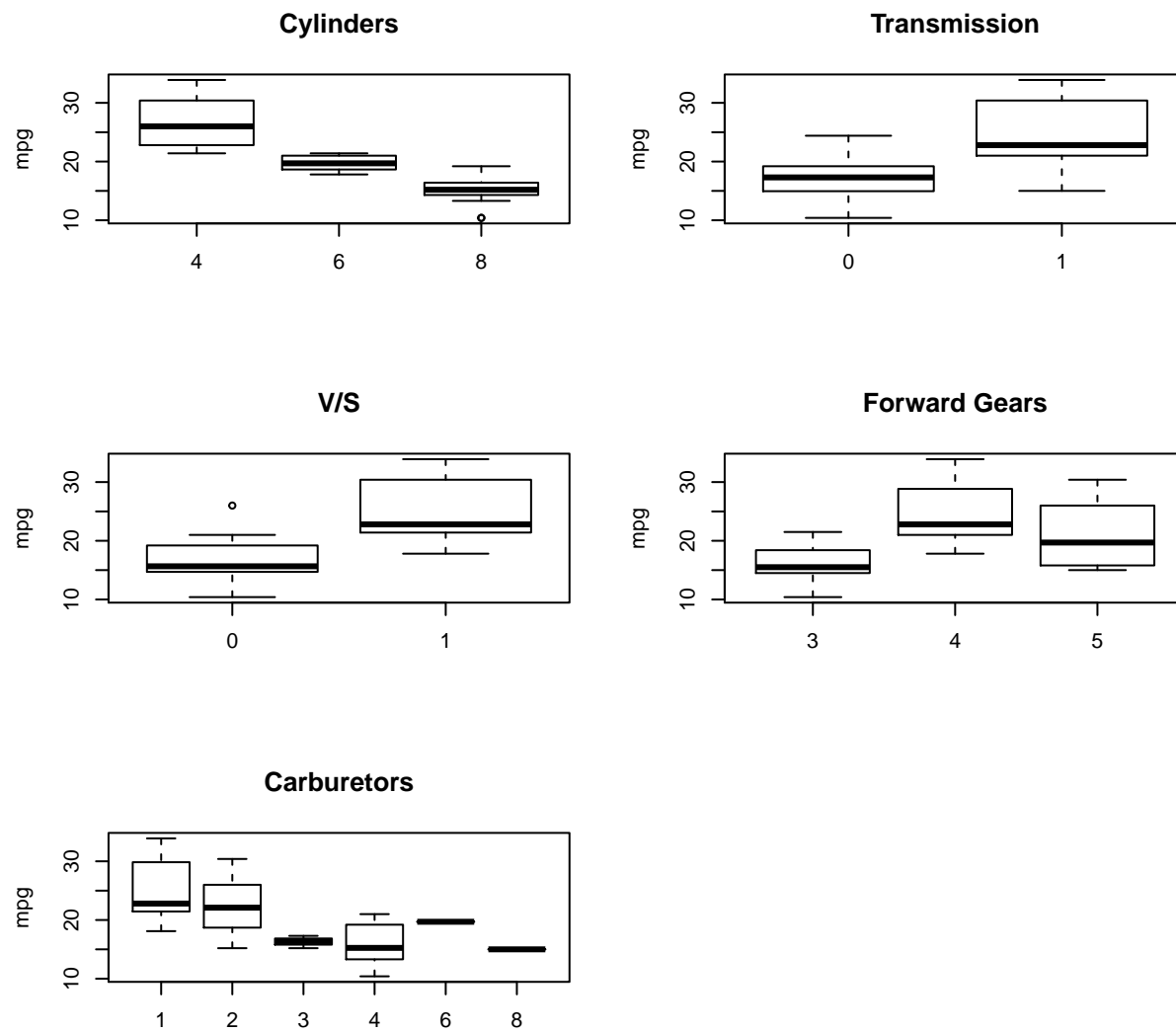


Figure A3

