Exploring Relationship Between Variables and MPG

Borye

Thursday, November 20, 2014

Introduction

This report is for *Motor Trend*, the magazine about the automobile industry, to exploring the relationship between a set of variables and miles per gallon (MPG). Particularly, this report is interested in the following two quetions:

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

Summary of Data

This report is using the data from **Motor Trend Car Road Tests**. 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 data frame with 32 observations on 11 variables. All variables are list as follow:

- Miles/(US) gallon `(mpg)`
- Number of cylinders `(cyl)`
- Displacement (cu.in.) `(disp)`
- Gross horsepower `(hp)`
- Rear axle ratio `(drat)`
- Weight (lb/1000) `(wt)`
- 1/4 mile time `(qsec)`
- V/S `(V/S)`
- Transmission (0 = automatic, 1 = manual) `(am)`
- Number of forward gears `(gear)`
- Number of carburetors `(carb)`

Exploratory Analysis

At first, a boxplot of mtcars shows the diffrence between automatic and manual transmission to the MPG [*Figure 1*]. It is clear that from the plot, manual transmission contribute more MPG compare to the automatic transmission. Secondly, a pairplot is presented to explore the relationship between all the different variables [*Figrue 2*]. From this figure, we can observed that lots of variables have the linear relationship between MPG. To clearly examine this, a correlation table is presented to see the exact correlation between variables and MPG [*Table 1*]. According to the table, (**drat = 0.681**, **qsec = 0.419**, **vs = 0.664**, **gear = 0.480**, and **am = 0.600**) are the variables that have a positive linear relationship between MPG. So these variables are chosen to build the model.

Model

After the basic variable selection, we have chosen (wt, qsec, and am) to form the model. To find the model, we use the method of anova table to have a further selection of the variables[*Table 2*]. The basic model is

just lm(mpg ~ am, mtcars). From the analysis above, we could obtain a new model which include the other variables which have great linear effect on the MPG. So the proposed new model is lm(mpg ~ am + wt + qsec, mtcars). From table 2, it can be seen that a p-value of 1.55e-09 was obtained. This value is very small, which means that the addition of the new variables are significant to improving the model. Summary of the model shows that the slop of am is 2.93, Which means that the average increase of MPG from Automatic to Manual transmission is 2.93 [*Table 3*].

Statistical Inference

A Two Sample t-test was conducted between the different transmission types. The null hypothesis that transmission types don't have an effect on the MPG is discarded for a p-value greater than 0.05. The results are shown in [*Table 4*]. The p-value of 0.001374 and difference in means show that manual transmission has significantly more MPG than automatic.

Conclusion

The transmission type of a car has a significant effect on its fuel efficiency. According to the model, manual transmission, on average, has **2.93** MPG more than automatics. According to the boxplot, manual transmission has **7.2** MPG more than automatics.

Appendix

Figure 1

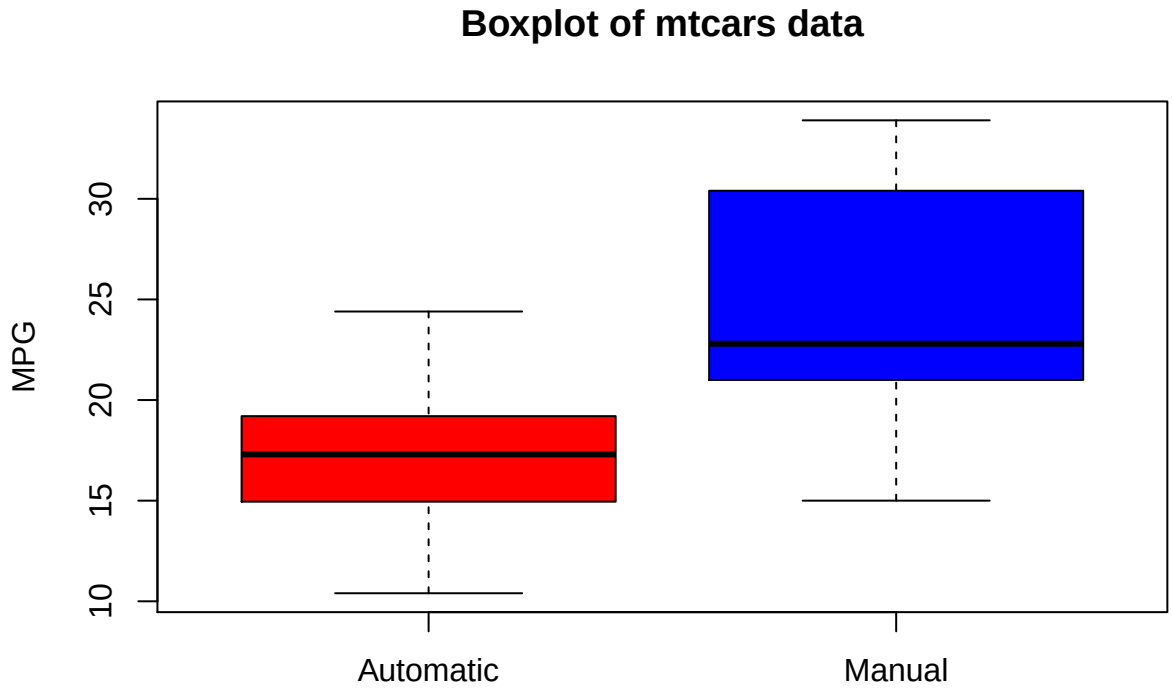

Figure 2

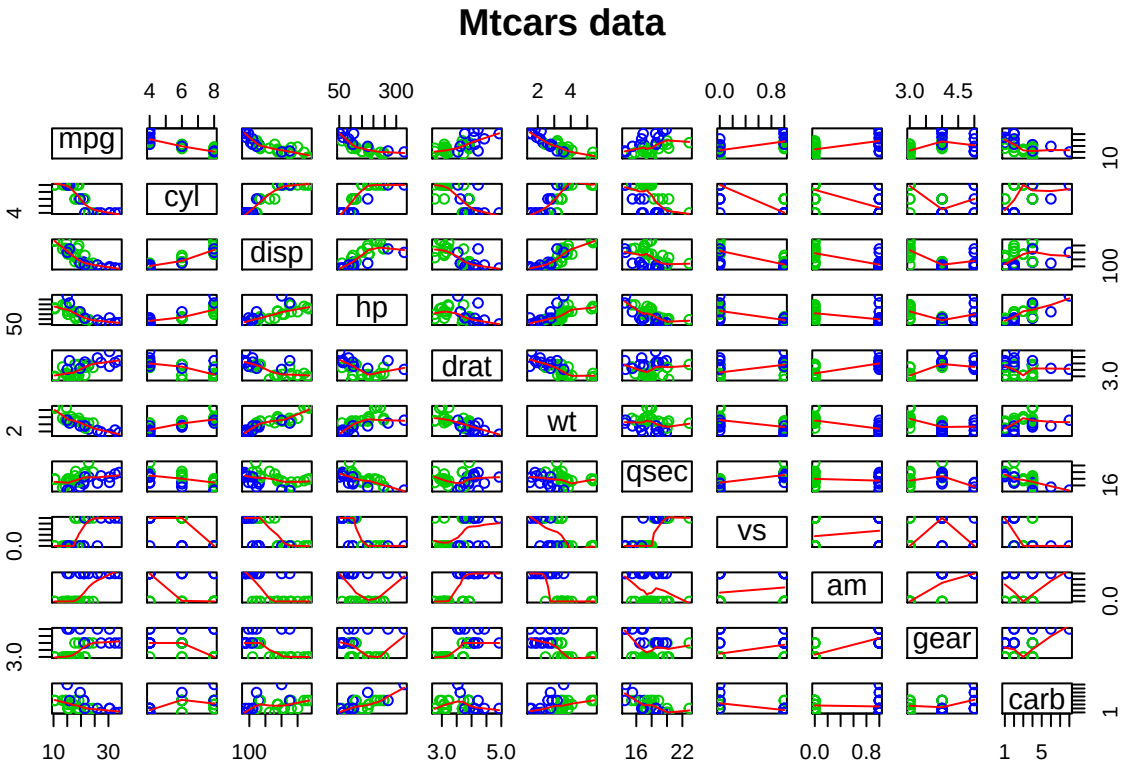

Table 1

```
##      mpg       cyl       disp         hp      drat         wt      qsec
## mpg    1 -0.852162 -0.8475514 -0.7761684 0.6811719 -0.8676594 0.418684
##            vs        am      gear       carb
## mpg 0.6640389 0.5998324 0.4802848 -0.5509251
```

Table 2

```
fit1 <- lm(mpg ~ am, data = mtcars)
fit2 <- lm(mpg ~ am + wt + qsec, data = mtcars)
anova(fit1, fit2)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + wt + qsec
##   Res.Df    RSS Df Sum of Sq      F   Pr(>F)
## 1     30 720.90
## 2     28 169.29  2    551.61 45.618 1.55e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Table 3

```r
summary(fit2)$coef
```

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept)  9.617781  6.9595930  1.381946 1.779152e-01
## am           2.935837  1.4109045  2.080819 4.671551e-02
## wt          -3.916504  0.7112016 -5.506882 6.952711e-06
## qsec         1.225886  0.2886696  4.246676 2.161737e-04
```

Table 4

```
##
##  Welch Two Sample t-test
##
## data:  mpg by am
## t = -3.7671, df = 18.332, p-value = 0.001374
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.280194  -3.209684
## sample estimates:
## mean in group 0 mean in group 1
##        17.14737        24.39231
```