Motor Trend Car Road Test - Effects of Automatic and Manual Transmission on MPG

Amy Su Jiang

December 19, 2014

Executive Summary

Motor Trend, an automobile trend magazine is interested in exploring the relationship between a set of variables and miles per gallon (MPG) outcome. In this project, we analyzed the mtcars dataset from the 1974 Motor Trend US magazine to answer the following questions:

- Is an automatic or manual transmission better for miles per gallon (MPG)?
- How different is the MPG between automatic and manual transmissions?

Through linear regression analysis, we have found a significant different between the mean MPG for automatic and manual transmission cars, that manual transmission cars have a higher mean MPG comparing to automatic transmission cars. Holding other variaables constant, we could estimate that manual transmission cars' MPG is about 1.8 times of automatic transmission cars' above 33.7 miles.

Data Processing

The following R code loaded mtcars dataset and transformed several variables into factors.

```
data(mtcars)
mtcars$cyl  <- factor(mtcars$cyl)
mtcars$vs   <- factor(mtcars$vs)
mtcars$gear <- factor(mtcars$gear)
mtcars$carb <- factor(mtcars$carb)
mtcars$am   <- factor(mtcars$am,labels=c("Automatic","Manual"))
```

Exploratory Analysis

To explore various relationships between variables, we plotted all the variables of the mtcars dataset, as shown in **Appendix Figure 1**. The plot shows that several variables - cyl, disp, hp, drat, wt, vs and am - have a strong correlation with mpg.

Since we are specifically interested in the effects of car transmission type on mpg in this analysis, we plotted a box polt to see the distribution of mpg for each level of am (Automatic or Manual), as shown in **Appendix Figure 2**. The plot clearly shows that manual transmissions tend to have higher MPG. We will further explore and quantify this relationship using linear regression analysis by fitting all variables into a linear model.

Regression Analysis

In this analysis, we first built linear regression models using different variables in order to find the best fit, then we compared it with the base model which we have used anova. After model selection, we also performed analysis on residuals.

Model building and selection

We started by building a model which includes all variables as predictors of mpg. The we used stepwise model selection to selected the most significat predictors for the best fit model. The step function we used here selects the best mdoel using AIC algorithm, which allow us to include the most significant variables for prediction while elminating the non-significant ones.

```
test.model <- lm(mpg ~ ., data = mtcars)
best.model <- step(test.model, direction = "both")
```

The best model obtained from the above computations shows that variables, cyl, wt and hp as confounders and am as the independent variable. Details of the model are shown as below.

```
summary(best.model)
```

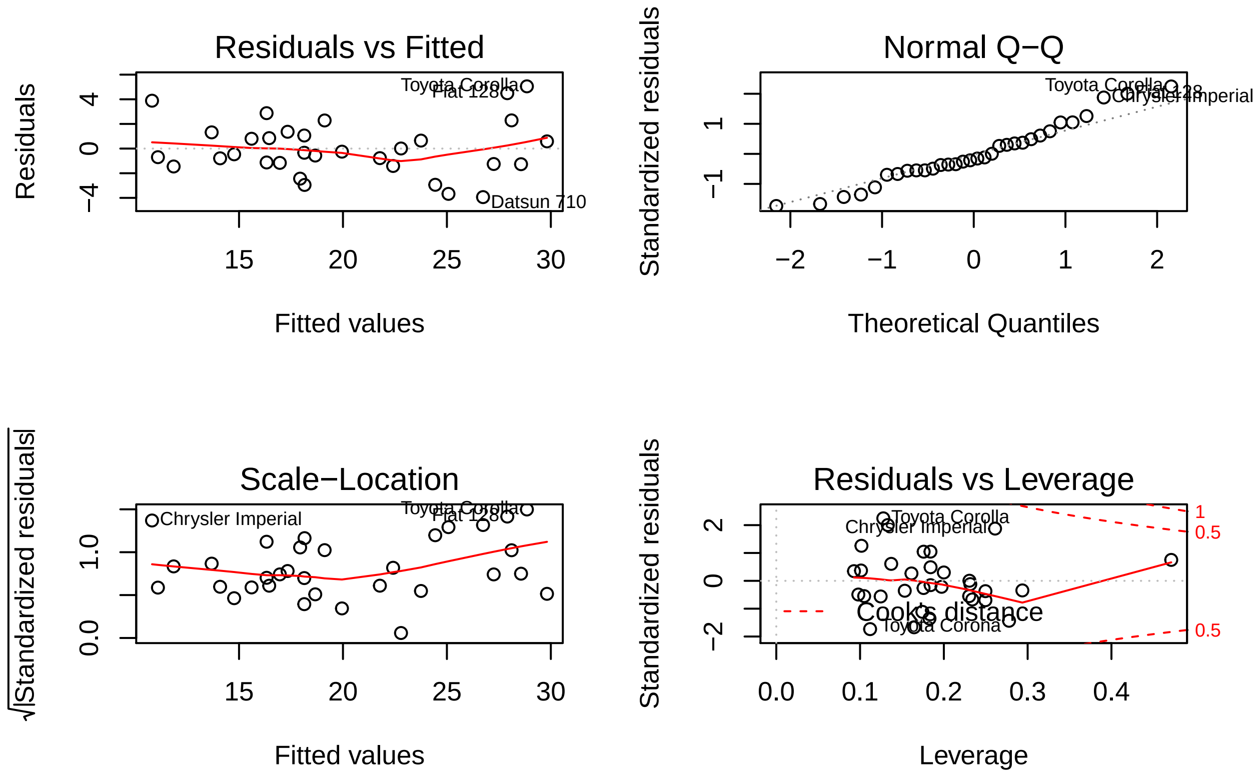
```
##
## Call:
## lm(formula = mpg ~ cyl + hp + wt + am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.9387 -1.2560 -0.4013  1.1253  5.0513
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 33.70832    2.60489  12.940 7.73e-13 ***
## cyl6        -3.03134    1.40728  -2.154  0.04068 *
## cyl8        -2.16368    2.28425  -0.947  0.35225
## hp          -0.03211    0.01369  -2.345  0.02693 *
## wt          -2.49683    0.88559  -2.819  0.00908 **
## amManual     1.80921    1.39630   1.296  0.20646
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.41 on 26 degrees of freedom
## Multiple R-squared:  0.8659, Adjusted R-squared:  0.8401
## F-statistic: 33.57 on 5 and 26 DF,  p-value: 1.506e-10
```

The R square value, 0.84, is the maximum obtained considering all combinations of variables. Hence we conclude that more than 84% of the data is explained by the above model.

Model Residuals and Diagnostics

The purpose of analysing residuals of the regression model is to find any leverage points which maybe a potential problem of the model.

```r
par(mfrow=c(2, 2))
plot(best.model)
```



From above plots we made following observations:

- The points in the Residuals vs. Fitted plot are randomly scattered that verifies the independence condition of all predictors.
- The Normal Q-Q plot consists of the points which mostly fall on the line indicating that the residuals are normally distributed.
- The Scale-Location plot consists of points scattered in a constant band pattern, indicating constant variance.
- There are some distinct points of interest (outliers or leverage points) in the top right of the plots that may indicate values of increased leverage of outliers.

Inferences

We performed a t-test on the two subsets of mpg data: manual and automatic transmission assuming that the transmission data has a normal distribution and tests the null hypothesis that they come from the same distribution.

```r
t.test(mpg ~ am, data = mtcars)
```

```
##
##  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 Automatic    mean in group Manual
##                17.14737                24.39231
```

Based on the t-test results, we reject the null hypothesis that the mpg distributions for manual and automatic transmissions are the same.

Conclusions

Based on our analysis, we can conclude that: * Manual transmission cars' MPG is about 1.8 times of automatic transmission cars' above 33.7 miles. * mpg will decrease by 2.5 for every 1000 lb increase in wt above 33.7 miles. * mpg will decrease 0.032 with every increase in hp above 33.7 miles. * If number of cylinders, cyl increases from 4 to 6 and 8, mpg will decrease by a factor of 3 and 2.2 respectively above 33.7 miles.

Appendix

Figure 1 - Pairs plot for the "mtcars" dataset

```r
pairs(mpg ~ ., data = mtcars)
```

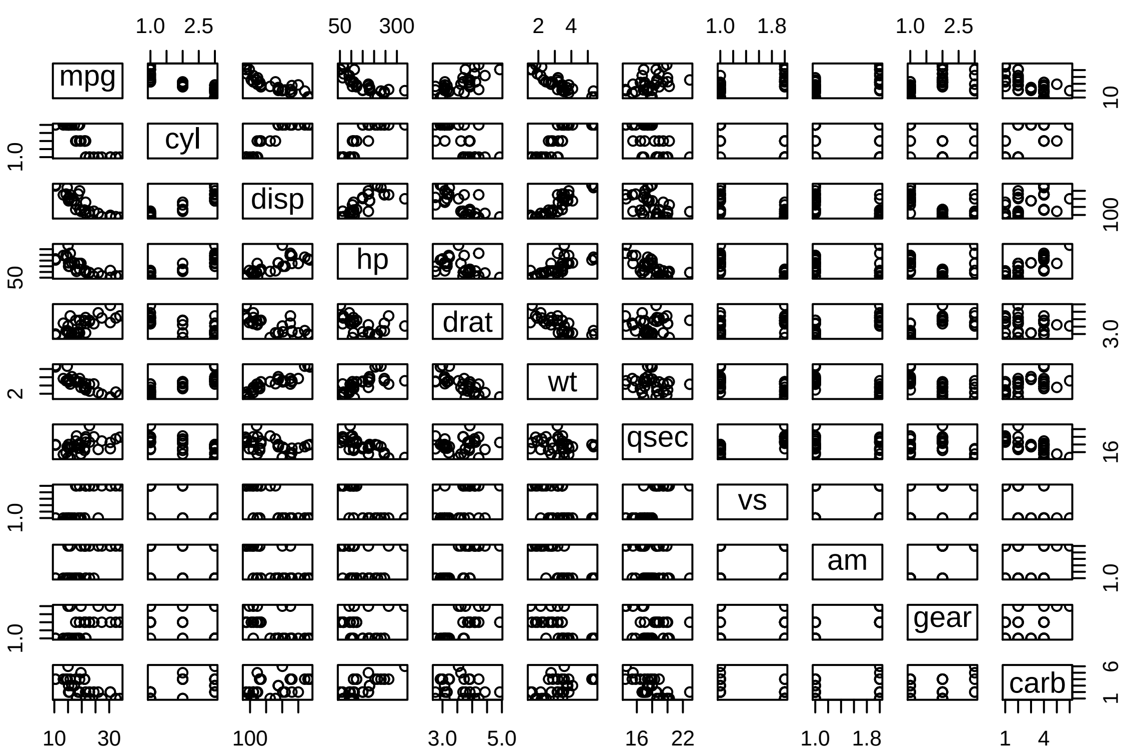


Figure 2 - Boxplot of miles per gallon by transmission type

```
boxplot(mpg ~ am, data = mtcars, col = (c("red","blue")), ylab = "Miles Per Gallon", xlab = "Transmissi
```