Coursera Regression Model Course Project

Chuk Yong

31/12/2016

Executive Summary

Motor Trend is a magazine about the automobile industry. Looking at a data set of a collection of cars, they are interested in exploring the relationship between a set of variables and miles per gallon (MPG) (outcome). They are particularly interested in the following two questions:

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

Through our study, we can conclude that Manual Transmission is better for MPG than Automatic Transmission. The difference on first look was 7mpg or about 30% more fuel efficient in favour of Manual Transmission. However, on further anaylsis by including other relevent factors, the difference turn out to be less. The MPG difference between them, taking account for factors like cylinder, horse-power, displacement and weight, is only 1.8MPG or about 5%.

A study of MPG between Automatic and Manual Transmission

Load dataset

```
data(mtcars)
```

Transform some variables into factor form

```
mtcars$cyl  <- factor(mtcars$cyl)
mtcars$vs   <- factor(mtcars$vs)
mtcars$am   <- factor(mtcars$am,labels=c("Auto","Man"))
mtcars$gear <- factor(mtcars$gear)
mtcars$carb <- factor(mtcars$carb)
```

Exploratory Data Analysis

```
mpgAuto <- subset(mtcars, mtcars$am == "Auto", select = c(mpg))
mpgMan <- subset(mtcars, mtcars$am == "Man", select = c(mpg))
t.test(mpgAuto$mpg, mpgMan$mpg)
```

The T Test result gives us:

1. Average mpg for Automatic transmission is 17.15.

2. Average mpg for Manual transmission is 24,39. 3. p-value = 0.001374

Clearly, Manual transmission has a much better mpg and the test is significant (p-value < 0.01).

We can now proceed with regression modelling to see how well the is the fit.

```
fit <- lm(mpg ~ am, data = mtcars)
summary(fit)
```

As confimed by the coefficients - Manual transmission is 7.245mpg better than Automatic transmission. p-value is much less than 0.01 and so they are significant. However, looking at R-squared and Adjusted R-squared, the value of 0.36 and 0.33 tells us that there is room for improvement.

Quantifying the MPG difference between Automatic and Manual Transmission

So far, we know than there is a significant difference between Automatic and Manual Transmission. However, there might be other factors at play that may impact MPG. In order to make a better decision on what other variables could possible impact mpg, we created a pair-plot, Fig 1, in Appendix A.

From the pair-plot, we can see that there are possible correlations between cyl - cylinder, disp - displacement, hp - horse power and wt - weight.

Multivariable Regression Study

We will fit in the variables and investigate if introduction them improve on our original fit.

A step by step approach is applied. We add the variables one at a time and evaluate the R.Squared value to see if we are making an improvement. While R.Squared may not tell the complete story, it is a start. Other diagnostic methods will be applied on our final model.

Fig 2 shows the result of R.Sqaured after adding the variables.

Final regression with 4 added variables: cyl, disp, hp and wt.

```
fit5 <- lm(mpg ~ am+cyl+disp+hp+wt, data = mtcars)
summary(fit5)
```

Now we can see that although Manual transmission still shows a better MPG vs Automatic, the difference is greatly reduced. At 1.8MPG over Automatic transmission, the difference is now only 5%.

Fig 2 in Appendix A shows the R.Sqaured value as we add the variables into the regression. We can see from the graph that all the variables do improve the fit.

Finally, we use diagnostic plots as a verification of our findings. Fig 3 in Appendix A shows 4 diagnostic plots. All 4 plots do not show any unexpected errors.

Appendix A

Fig 1 - Pair plots betweem all variables in mtcars dataset

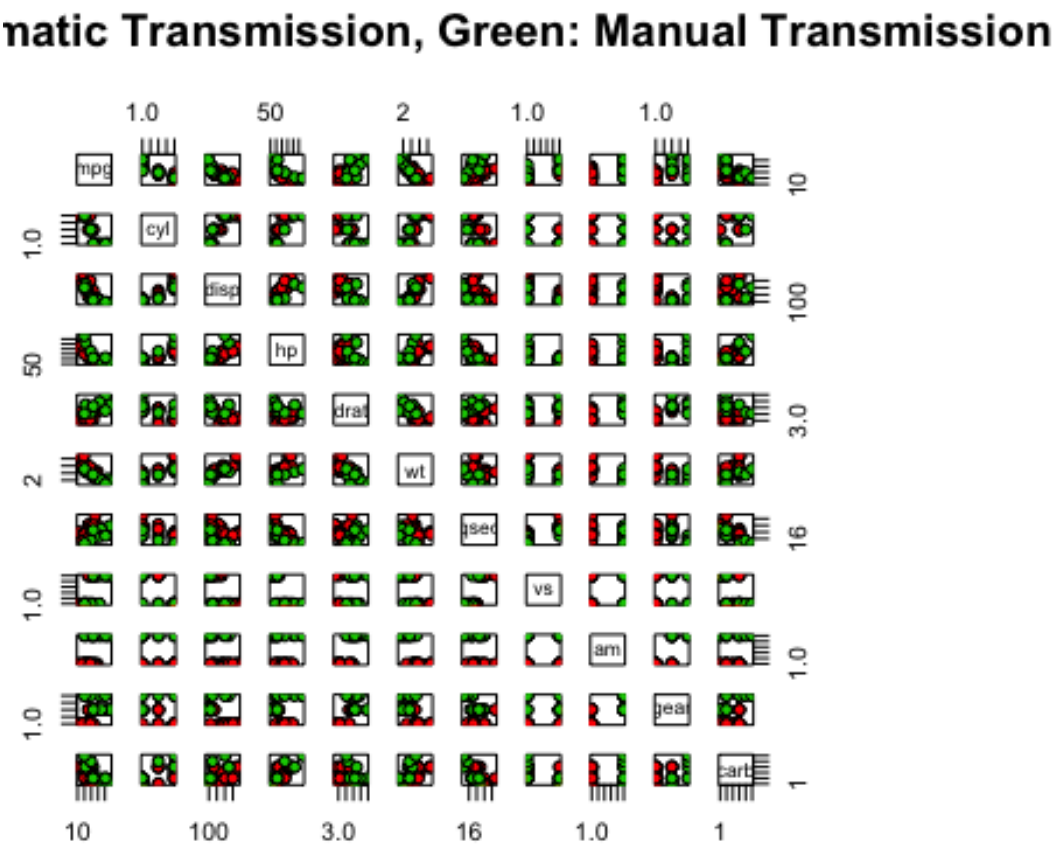
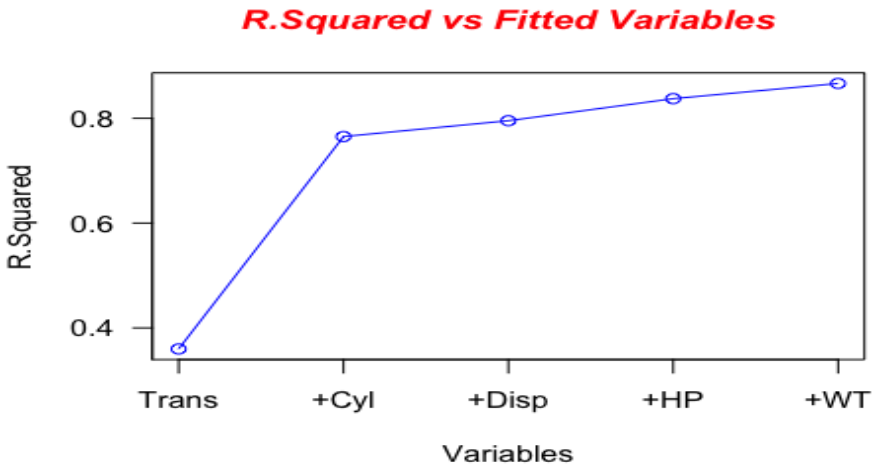


Fig 2 - Plot of how R.Sqaured perform when we add in our variables



X-axis indicates the variables we added (cumulative) to the regression.

Fig 3 - Diagnostic Plots