Regression Models_Project_Motor Trend Analysis

Maanas Mutreja

18 September 2016

Executive Summary

In this report, we will analyze mtcars data set and explore the relationship between a set of variables and miles per gallon (MPG). 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). We use regression models and exploratory data analyses to mainly explore how automatic (am = 0) and manual (am = 1) transmissions features affect the MPG feature. The t-test shows that the performance difference between cars with automatic and manual transmission. And it is about 7 MPG more for cars with manual transmission than those with automatic transmission. Then, we fit several linear regression models and select the one with highest Adjusted R-squared value. So, given that weight and 1/4 mile time are held constant, manual transmitted cars are 14.079 + (-4.141)*weight more MPG (miles per gallon) on average better than automatic transmitted cars. Thus we conclude that cars that are lighter in weight with a manual transmission and cars that are heavier in weight with an automatic transmission will have higher MPG values.

Method - Data Set

The data consists of 32 observations of 11 variables. mpg: Miles per US gallon cyl: Number of cylinders disp: Displacement (cubic inches) hp: Gross horsepower drat: Rear axle ratio wt: Weight (lb / 1000) qsec: 1/4 mile time vs: V/S am: Transmission (0 = automatic, 1 = manual) gear: Number of forward gears carb: Number of carburetors

Results

We first did some basic exploratory data analyses. Please refer to the **Appendix** section for the plots and exploratory analysis. According to the box plot, we see that manual transmission yields higher values of MPG in general. And as for the pair graph, we can see some higher correlations between variables like “wt”, “disp”, “cyl” and “hp”.

Inference At this step, we make the null hypothesis as the MPG of the automatic and manual transmissions are from the same population (assuming the MPG has a normal distribution). We use the two sample T-test to show it.

```
## The following object is masked from package:ggplot2:
##
##     mpg
```

```
result <- t.test(mpg ~ am, data = mtcars)
result$p.value
result$estimate
```

Since the p-value is 0.00137, we reject our null hypothesis. So, the automatic and manual transmissions are from different populations. And the mean for MPG of manual transmitted cars is about 7 more than that of automatic transmitted cars.

Regression Analysis First, we fit the full model as the following.

```r
fullModel <- lm(mpg ~ ., data=mtcars)
summary(fullModel) # results hidden
```

This model has the Residual standard error (RSE) as 2.833 on 15 degrees of freedom. And the Adjusted R-squared value is 0.779, which means that the model can explain about 78% of the variance of the MPG variable. However, none of the coefficients are significant at 0.05 significant level. Then, we use backward selection to select some statistically significant variables.

```r
stepModel <- step(fullModel, k=log(nrow(mtcars)))
summary(stepModel) # results hidden
```

This model is "mpg ~ wt + qsec + am". It has the RSE as 2.459 on 28 degrees of freedom. And the Adjusted R-squared value is 0.8336, which means that the model can explain about 83% of the variance of the MPG variable. All of the coefficients are significant at 0.05 significant level. According to the scatter plot (**Appendix**), it indicates that there appear to be an interaction term between "wt" variable and "am" variable, since automatic cars tend to weigh heavier than manual cars. Thus, we have the following model including the interaction term:

```r
amIntWtModel<-lm(mpg ~ wt + qsec + am + wt:am, data=mtcars)
summary(amIntWtModel) # results hidden
```

This model has the RSE as 2.084 on 27 degrees of freedom. And the Adjusted R-squared value is 0.8804, which means that the model can explain about 88% of the variance of the MPG variable. All of the coefficients are significant at 0.05 significant level. This is a pretty good one. Next, we fit the simple model with MPG as the outcome variable and Transmission as the predictor variable.

```r
amModel<-lm(mpg ~ am, data=mtcars)
summary(amModel) # results hidden
```

It shows that on average, a car has 17.147 mpg with automatic transmission, and if it is manual transmission, 7.245 mpg is increased. This model has the Residual standard error as 4.902 on 30 degrees of freedom. And the Adjusted R-squared value is 0.3385, which means that the model can explain about 34% of the variance of the MPG variable. The low Adjusted R-squared value also indicates that we need to add other variables to the model. Finally, we select the final model.

```r
anova(amModel, stepModel, fullModel, amIntWtModel)
confint(amIntWtModel) # results hidden
```

We end up selecting the model with the highest Adjusted R-squared value, "mpg ~ wt + qsec + am + wt:am"

```r
summary(amIntWtModel)$coef
```

Conclusion

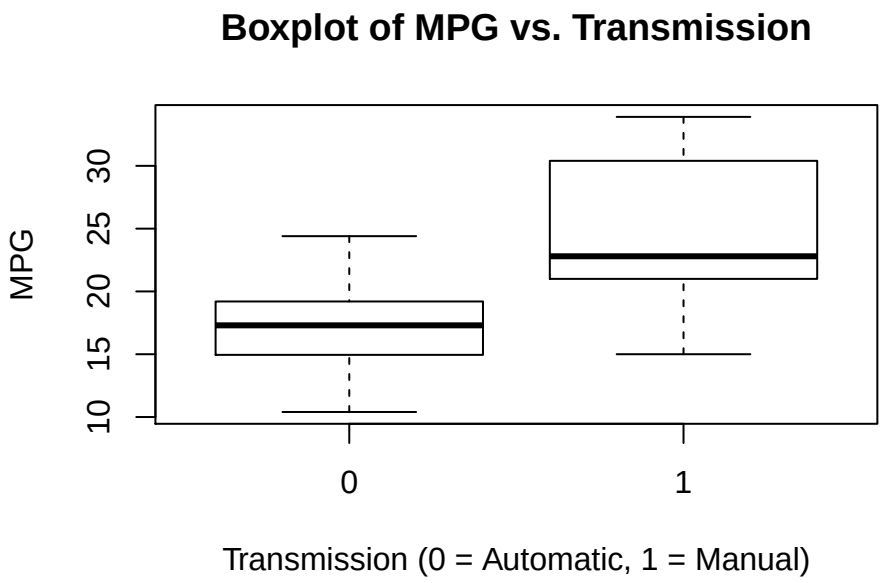
Thus, the result shows that when "wt" (weight lb/1000) and "qsec" (1/4 mile time) remain constant, cars with manual transmission add 14.079 + (-4.141)*wt more MPG (miles per gallon) on average than cars with automatic transmission. That is, a manual transmitted car that weighs 2000 lbs have 5.797 more MPG than an automatic transmitted car that has both the same weight and 1/4 mile time.

APPENDIX

Figures

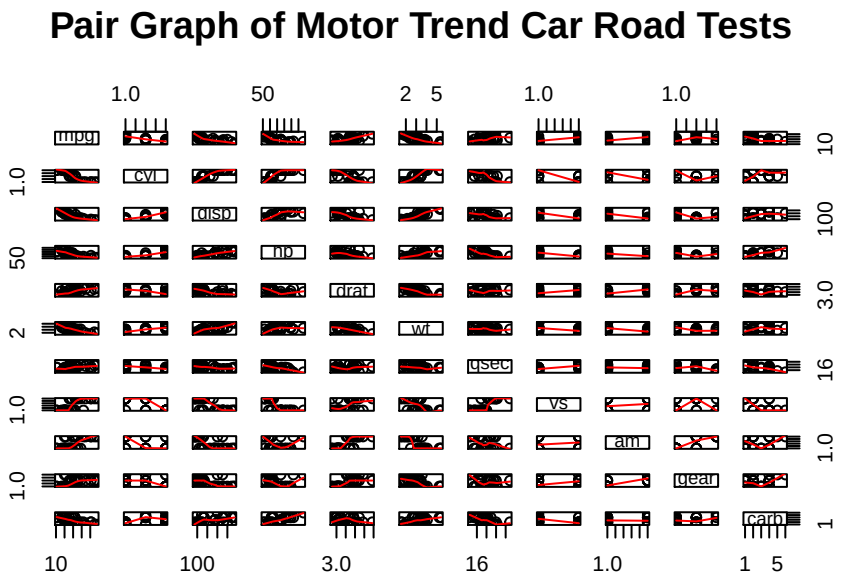
1. Boxplot of MPG vs. Transmission

```
boxplot(mpg ~ am, xlab="Transmission (0 = Automatic, 1 = Manual)", ylab="MPG", main="Boxplot of MPG vs.
```



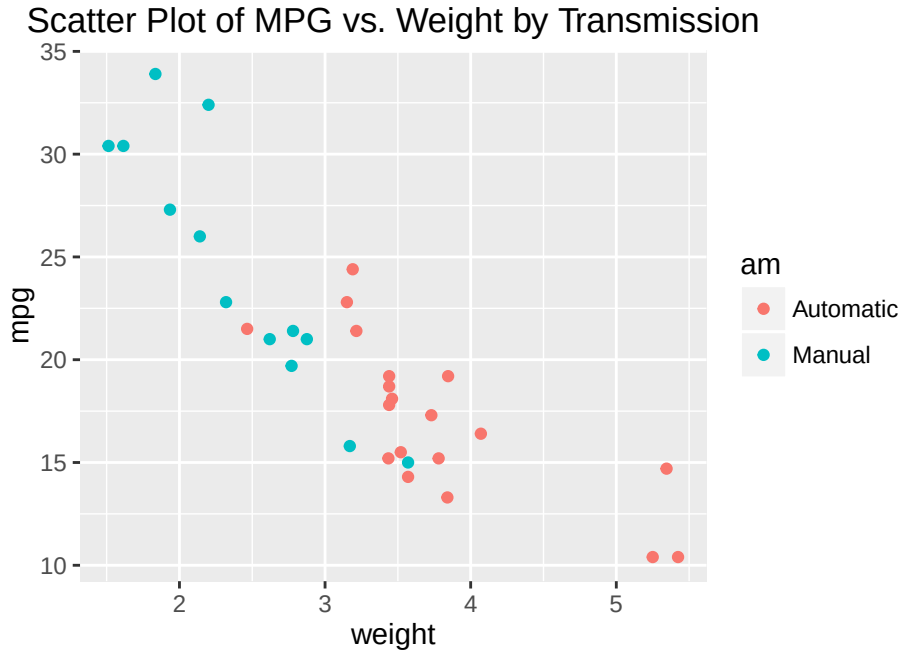
2. Pair Graph of Motor Trend Car Road Tests

```
pairs(mtcars, panel=panel.smooth, main="Pair Graph of Motor Trend Car Road Tests")
```



3. Scatter Plot of MPG vs. Weight by Transmission

```
library(ggplot2)
ggplot(mtcars, aes(x=wt, y=mpg, group=am, color=am, height=3, width=3)) + geom_point() + scale_colour_di
xlab("weight") + ggtitle("Scatter Plot of MPG vs. Weight by Transmission")
```



4. Residual Plots

```
par(mfrow = c(2, 2))
plot(amIntWtModel)
```