

Regression Models Project - Motor Trend Data Analysis Report

Krish Mahajan

Executive Summary

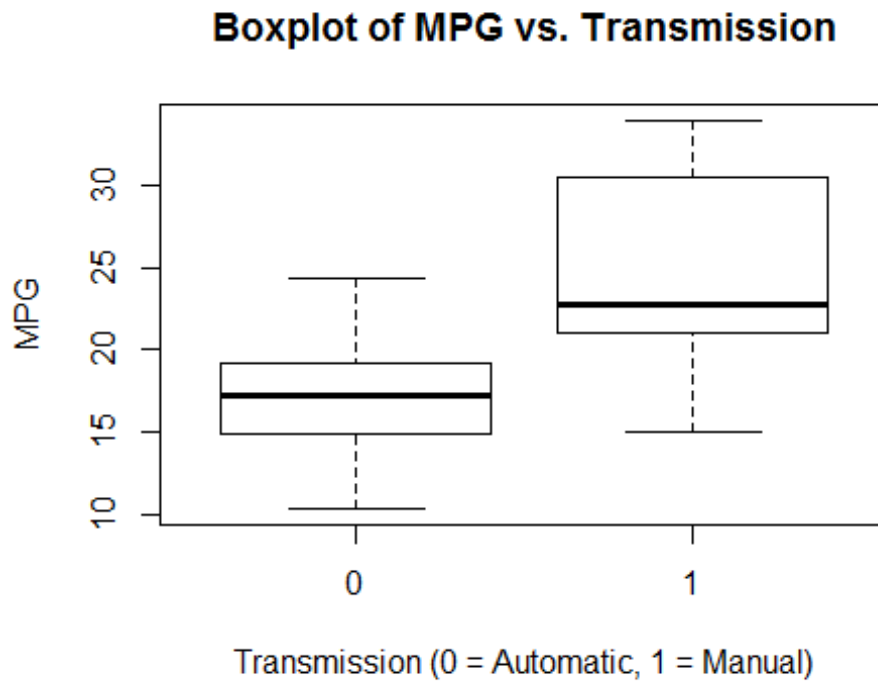
In this report, we will analyze mtcars data set and explore the relationship between a set of variables and miles per gallon (MPG). The data was extracted from the **1974 Motor Trend** US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973-74 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 $14.079 + (-4.141) * weight$ gives more MPG (miles per gallon) on average better than automatic transmitted cars. ***Thus, 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***

Exploratory Data Analysis

First, we load the data set mtcars and change some variables from numeric class to factor class.

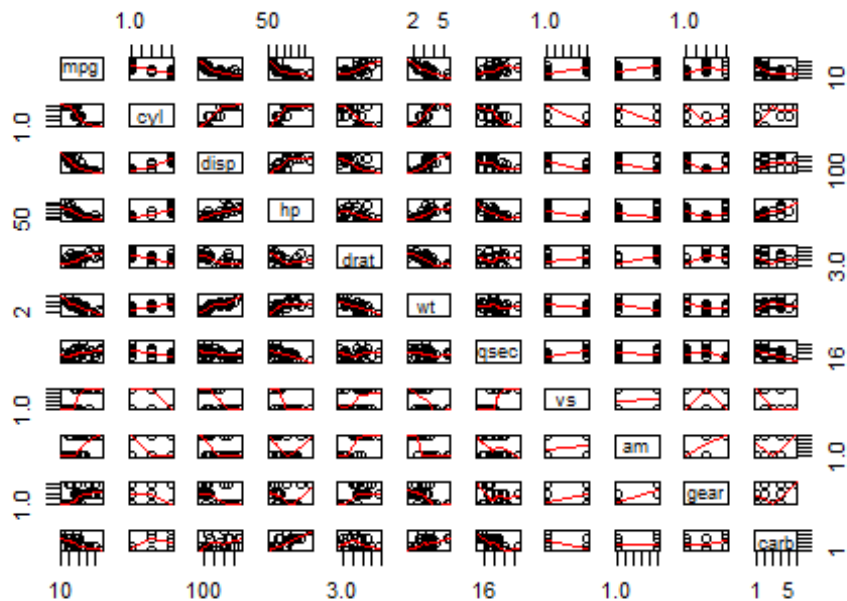
##	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
## Mazda RX4	21.0	6	160	110	3.90	2.620	16.46	0	1	4	4
## Mazda RX4 Wag	21.0	6	160	110	3.90	2.875	17.02	0	1	4	4
## Datsun 710	22.8	4	108	93	3.85	2.320	18.61	1	1	4	1
## Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44	1	0	3	1
## Hornet Sportabout	18.7	8	360	175	3.15	3.440	17.02	0	0	3	2
## Valiant	18.1	6	225	105	2.76	3.460	20.22	1	0	3	1

Then, we do some basic exploratory data analyses. Lets do a box plot analysis to see a general effect of transmission on mpg.



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".

Pair Graph of Motor Trend Car Road Tests



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.

```
attach(mtcars)
result <- t.test(mpg ~ am)
result$p.value

## [1] 0.001373638

result$estimate

## mean in group 0 mean in group 1
##      17.14737      24.39231
```

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.

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

This model has the Residual standard error 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.

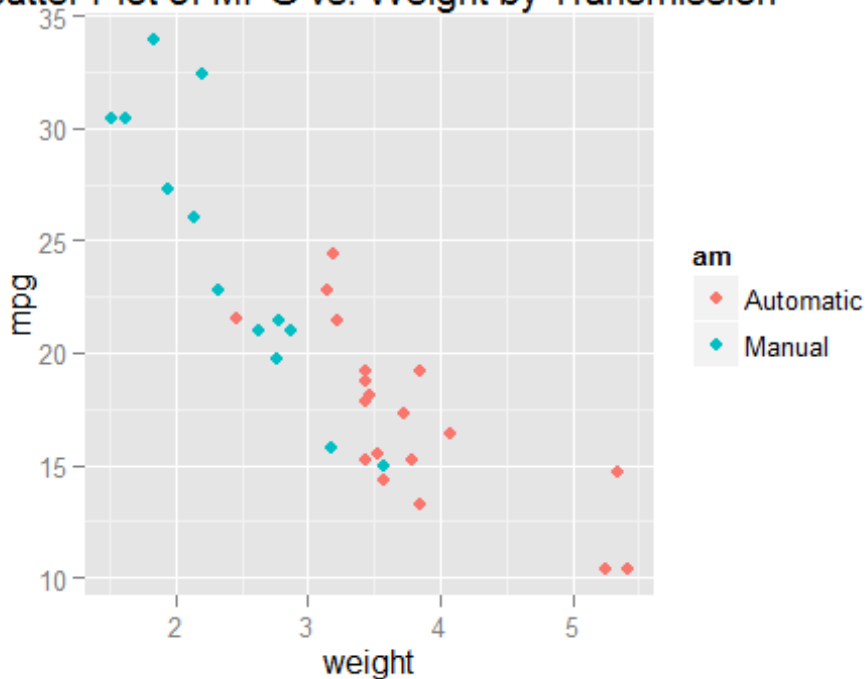
Then, we use backward selection to select some statistically significant variables.

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

This model is "mpg ~ wt + qsec + am". It has the Residual standard error 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

Now, on further analysis check the scatter plot.
Scatter Plot of MPG vs. Weight by Transmission

Scatter Plot of MPG vs. Weight by Transmission



According to the scatter plot, 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:

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

This model has the Residual standard error 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.

```
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.

```
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".

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

##	Estimate	Std. Error	t value	Pr(> t)
## (Intercept)	9.723053	5.8990407	1.648243	0.1108925394
## wt	-2.936531	0.6660253	-4.409038	0.0001488947
## qsec	1.016974	0.2520152	4.035366	0.0004030165
## am1	14.079428	3.4352512	4.098515	0.0003408693
## wt:am1	-4.141376	1.1968119	-3.460340	0.0018085763

Thus, the result shows that when "wt" (weight lb/1000) and "qsec" (1/4 mile time) remain constant, cars with manual transmission add $14.079 + (-4.141) \cdot \text{wt}$ 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.