Cars Study @ Motor Trend

by Vipul Sharma

3/11/2018

Summary

At Motor Trend we are doing a study of available cars and their models to understand what impacts miles per gallon(MPG) the most. For this we are using the data available in `mtcars` data set and explore the relationship between a set of variables available in this data set and miles per gallon (MPG).

Using regression models and exploratory data analyses, we want to mainly explore how **automatic** (am = 0) and **manual** (am = 1) transmissions effects on the fuel efficiency (measured in terms of the definitely non-SI unit *miles per gallon*) of a set of 32 automobile models from 1973-74.

The initial t-test results shows that there is a 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, 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.

The source code for this document can be retrieved from my GitHub account.

Exploratory Data Analysis

First, data set `mtcars` is loaded and some variables are of `numeric` class which we changed to `factor` class.

```
library(ggplot2)
data(mtcars)
mtcars[1:3, ]
```

```
               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
```

```
mtcars$cyl <- as.factor(mtcars$cyl)
mtcars$vs <- as.factor(mtcars$vs)
mtcars$am <- factor(mtcars$am)
mtcars$gear <- factor(mtcars$gear)
mtcars$carb <- factor(mtcars$carb)
attach(mtcars)
```

Then, we do some basic exploratory data analyses. Please refer to the **Appendix: Figures** section for the plots.

Inference

Then in next step, considering 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.

```r
t_result <- t.test(mpg ~ am)
t_result$p.value
```

```
[1] 0.001373638
```

```r
t_result$estimate
```

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

With p-value is 0.00137, our null hypothesis has to be rejected. Hence, 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

Initial Model : Fitting all variables to get the initial model

```r
mdl <- lm(mpg ~ ., data=mtcars)
```

Residual standard error as 2.833 on 15 degrees of freedom. And the Adjusted R-squared value is 0.779. However, none of the coefficients are significant at 0.05 significant level.

Model that matters : Refering to Fig.2. The scatter plot 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
am_int_wt_mdl<-lm(mpg ~ wt + qsec + am + wt:am, data=mtcars)
```

This model has the Residual standard error as 2.084 on 27 degrees of freedom. All of the coefficients are significant at 0.05 significant level. This seems to be a pretty good one.

Simple Model : Next, we fit the simple model with MPG as the outcome variable and Transmission as the predictor variable.

```r
am_mdl<-lm(mpg ~ am, data=mtcars)
```

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.

Model Selection

```r
anova(am_mdl, mdl, am_int_wt_mdl)
confint(am_int_wt_mdl)
```

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

```r
summary(am_int_wt_mdl)$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
```

Selected Model suggests, 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.

What Analysis suggests

Reffering to **Appendix: Figures** section for the plots, we can infer 1. The Residuals vs. Fitted plot shows no consistent pattern, suggesting the accuracy of the independence assumption.
2. The Normal Q-Q plot indicates that the residuals are normally distributed since the points lie closely to the line.
3. The Scale-Location plot confirms the constant variance assumption, as the points are randomly distributed.
4. The Residuals vs. Leverage argues that no outliers are present, as all values fall well within the 0.5 bands.

As for the Dfbetas, the measure of how much an observation has effected the estimate of a regression coefficient, we get the following result:

```
sum((abs(dfbetas(am_int_wt_mdl)))>1)
```

```
## [1] 0
```

Therefore, the above analyses meet all basic assumptions of linear regression and well answer the questions.

Appendix: Figures

Fig.1. Plot of MPG vs. Transmission Type

```
library(GGally,quietly = T, warn.conflicts = F)
ggpairs(data=mtcars,columns=c(1,9),
        upper=list(combo="box"),lower=list(combo="facethist"),
        diag=list(continuous="density",discrete="bar"))
```

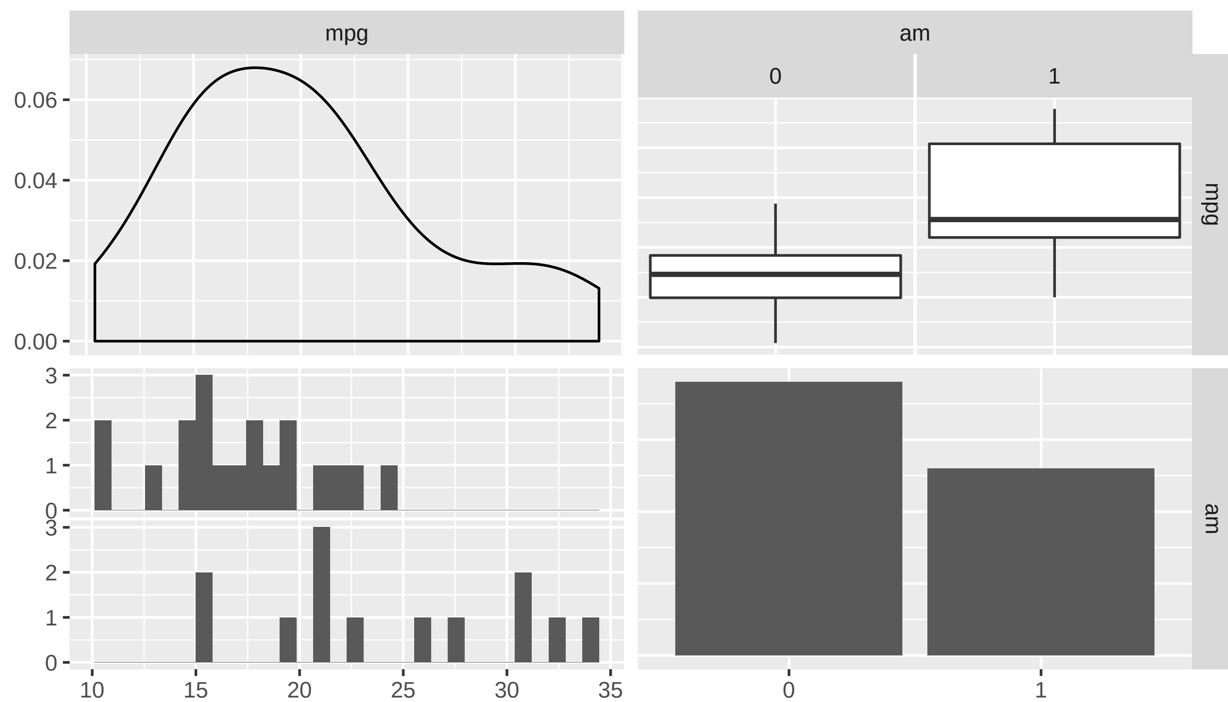

Fig.2. Pair Graph of Motor Trend Car Road Tests

```
pairs(mtcars, panel=panel.smooth,pch=16, cex=0.5, gap=0.25, lwd=2, las=1, cex.axis=0.7)
```

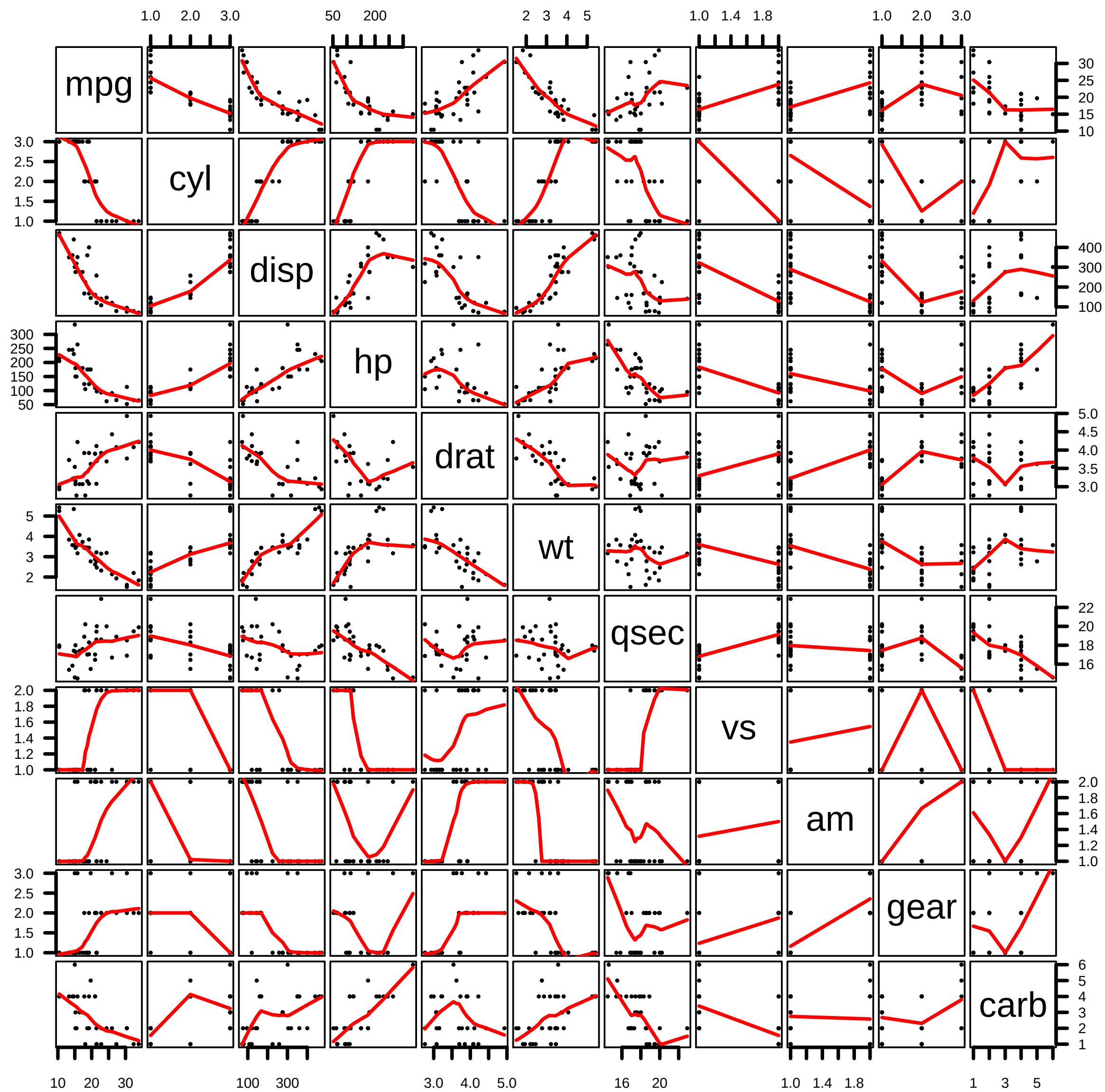

Fig.3. Scatter Plot of MPG vs. Weight by Transmission

```
ggplot(mtcars, aes(x=wt, y=mpg, group=am, color=am, height=3, width=3)) + geom_point() +
scale_colour_discrete(labels=c("Automatic", "Manual")) +
xlab("weight")
```

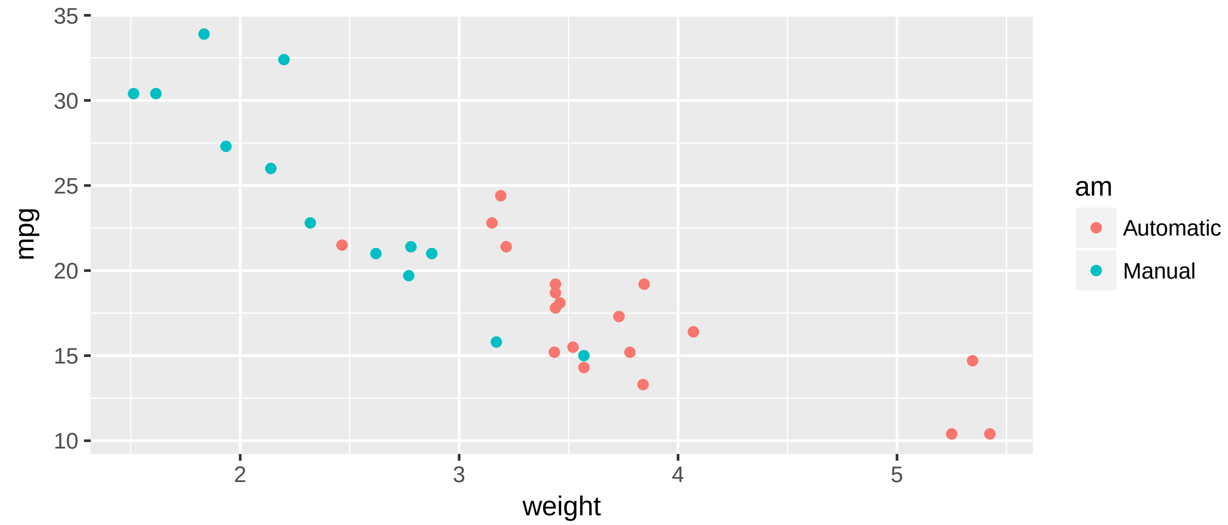

Fig.4. Residual Plots

```r
par(mfrow = c(2, 2))
plot(am_int_wt_mdl,pch=16,lty=1,lwd=2)
```