RegressioModels_Week4_Proj - Motor Trend MPG Data Analysis

Venkat.T

9/15/2019

Executive Summary

In this report we will analyze the mtcars dataset containing data from 32 cars which were 1973-1974 models to answer the question which of the transmission systems (automatic or manual) is better for MPG. The analysis concluded that the there is a significant difference in mean fuel consumptions between automatic and manual transmission. Our regression model shows that taking into account also weight of the cars and its acceleration, cars with manual transmission are more efficient regarding to fuel consumption. According to the model, for cars with the same weight and acceleration manual transmission adds about 2.9358 MPG compaired to automatic transmission.

The analysis

#1. Exploratory analysis.

First we will load the data set and present its sample.

```
data(mtcars)
str(mtcars)
head(mtcars)
```

Now we will do some basic data cleaning. We will process automatic and manual transmission data.

```
auto <- mtcars$mpg[mtcars$am == 0]
manu <- mtcars$mpg[mtcars$am == 1]
```

Now we will plot the data in Figure 1 that can be found in the appendix. The plot shows us, that manual transmission increases mpg. Now we will check, if the difference is statistically significant. To do this we will perform a t.test for not paired means.

```
test <- t.test(manu, auto)
test$p.value
```

```
## [1] 0.001373638
```

```
test$estimate
```

```
## mean of x mean of y
##  24.39231  17.14737
```

The result of t.test confirms that the difference is significant with p-value equal 0.001374 meaning that the alternative hypothesis that difference in means is equal 0 is not true.

#2. Model selection:

Now we will try to confirm above findings with the same analysis but using regression model with dumb variables:

```
fit1 <- lm(mpg ~ factor(am), data = mtcars)
summary(fit1)$coef
```

From above we can see that on avarage the car with an automatic transmission has 17.147 mpg and for the manual transmission 7.245 mpg is added. R squared of the model is 0.3598 so it only exlpains about 36% of the variance of the MPG. This means that there are different variables that can also explain the change of mpg.

So now, we will construct full model with all available variables:

```
fit2 <- lm(mpg ~ ., data = mtcars)
summary(fit2)
```

R-squared statistic of the model is 0.869 which means that the model explains about 87% of the variance of the MPG variable. Unfortunately none of the coefficients seams to be statistically significant according to p-values (none of the p-values is lower than 0.05). To determine significant variables we will use backward selection. Before we can plot the interactions between all variables using pairs function (Figure 2 in the appendix).

Now we will try to select variables with the step function:

```
fit3 <- step(fit2)
summary(fit3)
```

The new model contains three variables: weight (wt), 1/4 mile time (qsec), and transmission (am). R squared value is 0.8497 which means, that the model explains about 85% of the variance of the MPG. All the coefficients are statistically significant with p-value below 0.05.

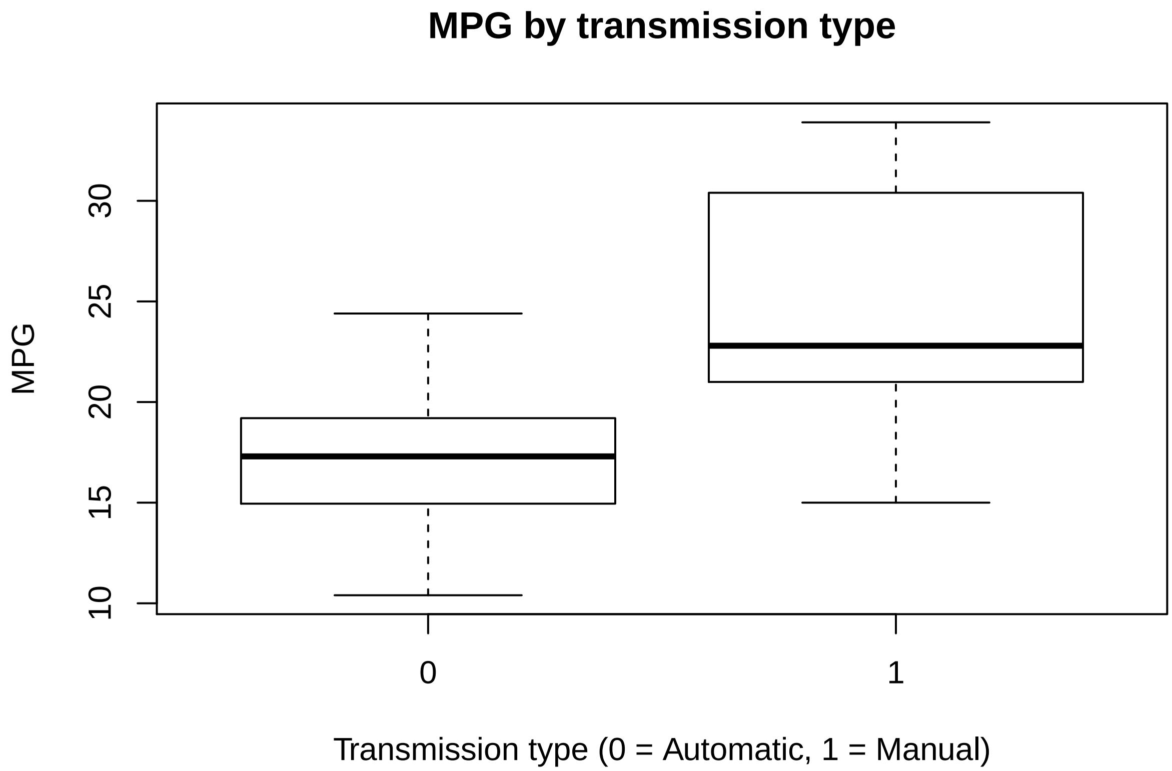
The above model shows that with weight and quarter mile time constant cars with manual transmission 2.9358 more MPG on average than cars with automatic transmission.

#3. Model diagnostics:

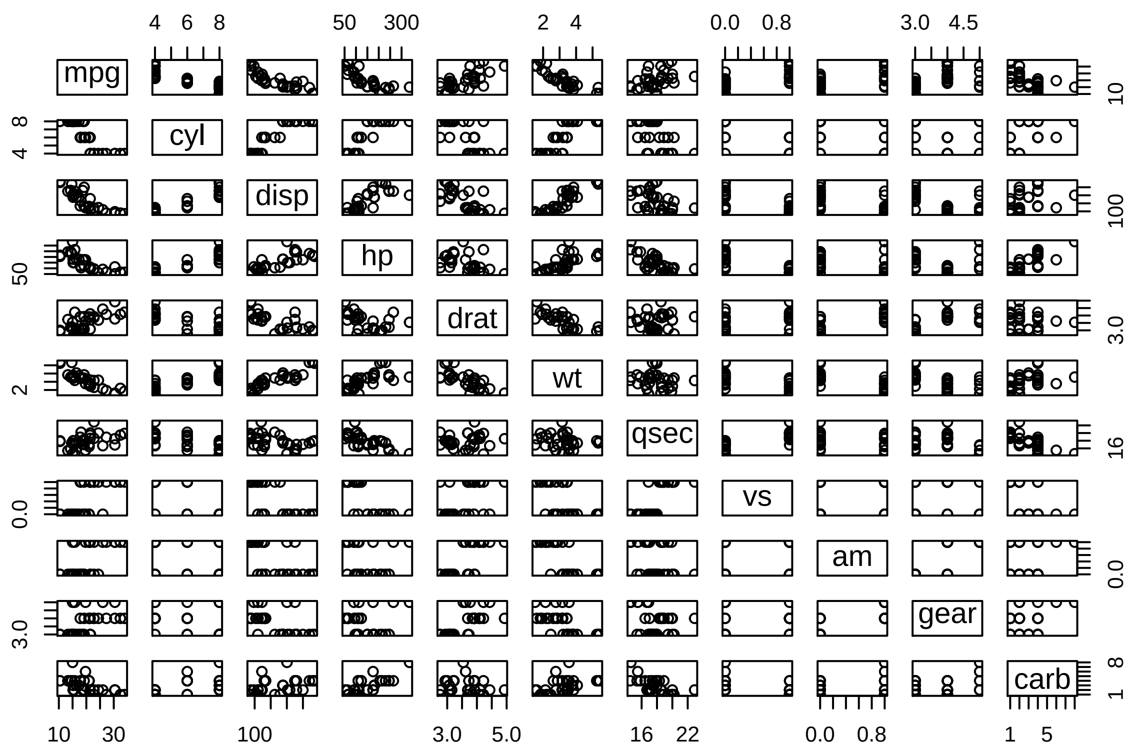
Now we will plot the residuals and the Normal Q-Q to diagnose the model (Figure 3 in the appendix). We can see that the residuals do not reveal any patterns on the residuals vs fitted plot. Also the Normal Q-Q plot shows that residuals seems to be normally distributed. Thus we conclude that the basic assumptions for the linear regression are met.

#4. Appendix:

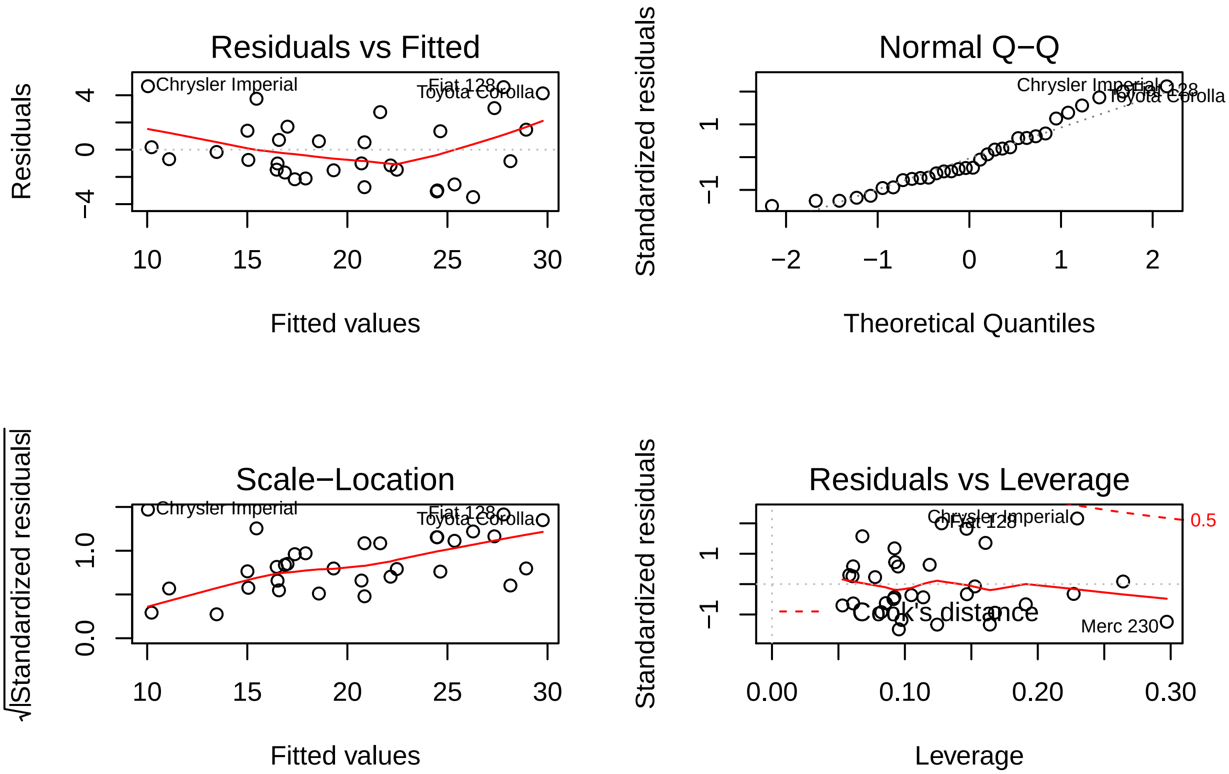
#Figure 1:



#Figure 2:



#Figure 3:



Residuals vs Fitted
Residuals
Fitted values
Chrysler Imperial
Fiat 128
Toyota Corolla
Normal Q−Q
Standardized residuals
Theoretical Quantiles
Scale−Location
√|Standardized residuals|
Residuals vs Leverage
Leverage
Cook's distance
Merc 230
0.5