Regression Models Course Project

Richard Berry

6 April 2016

Which transmission type has the better fuel efficiency in miles per gallon (MPG), Manual or Automatic?

Executive Summary

In this report we will investigate the relationship between fuel efficiency and transmission type. We will apply our understanding of regression models to our data and comment on the results.

In the hypothesis we predict that Manual transmissions will have a greater fuel efficiency than Automatic transmissions. This can be factored to many other variables such as horse power or number of cylinders and will be commented on as we proceed.

The data in this project will be collected from the 'mtcars' package in R.

Exploratory Data Analysis

We will import all libraries needed as well as load the data that we will analyse. The data will be cleaned. We will conduct a brief summary of the data and change how certain fields effect the way we subdivide the vehicle factors.

```
library(ggplot2)
data(mtcars)

mtcars[20:22, ]
```

```
##                   mpg cyl  disp  hp drat    wt  qsec vs am gear carb
## Toyota Corolla   33.9   4  71.1  65 4.22 1.835 19.90  1  1    4    1
## Toyota Corona    21.5   4 120.1  97 3.70 2.465 20.01  1  0    3    1
## Dodge Challenger 15.5   8 318.0 150 2.76 3.520 16.87  0  0    3    2
```

```
n <- length(mtcars$mpg)
alpha <- 0.05
mtcars$cyl <- as.factor(mtcars$cyl)
mtcars$am <- factor(mtcars$am)
attach(mtcars)
```

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

Inference

We can now do our first correlation between these Transmission type and MPG.

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

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 17.147368   1.124603 15.247492 1.133983e-15
## am1          7.244939   1.764422  4.106127 2.850207e-04
```

These values indicate the mean MPG for each transmission type, however they are not in a simple form. As the intercept for the above lm is ~17.15 this value represents the mean MPG for am = 0 cars (Automatic). The am1 value of ~7.24 is the increase in MPG at am = 1. Hence the Mean MPG for each transmission type is as follows:

```
## [1] "Automatic:     17.15 MPG"
```

```
## [1] "Manual:        24.39 MPG"
```

Let us compare it to the t.test method and comment on the null hypothesis.

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

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

```r
ttest$estimate
```

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

With a low p-value we can reject our null hypothesis and say that the automatic and manual transmissions come from different populations which is good and we can see that our previous statement about the relative MPG values for each transmission type are the same.

Regression Analysis

Let us compare different models of regression

Model 1:

```r
mod1 <- lm(mpg~.,data=mtcars)
summary(mod1)
## Residual standard error: 2.582 on 20 degrees of freedom
## Multiple R-squared: 0.8816, Adjusted R-squared: 0.8165
## F-statistic: 13.54 on 11 and 20 DF, p-value: 5.722e-0
```

This model can explain 88% of the variance in the MPG of a vehicle due to transmission type.

Model 2:

```r
mod2 <- step(mod1, k=log(nrow(mtcars)))
summary(mod2)
## Residual standard error: 2.459 on 28 degrees of freedom
## Multiple R-squared: 0.8497, Adjusted R-squared: 0.8336
## F-statistic: 52.75 on 3 and 28 DF, p-value: 1.21e-11
```

This model can explain 85% of the variance in the MPG of a vehicle due to transmission type.

Model 3:

```r
mod3 <- lm(mpg ~ wt + am + wt + qsec + wt:am,data=mtcars)
summary(mod3)
## Residual standard error: 2.084 on 27 degrees of freedom
## Multiple R-squared: 0.8959, Adjusted R-squared: 0.8804
## F-statistic: 58.06 on 4 and 27 DF, p-value: 7.168e-13
```

This model can explain 90% of the variance in the MPG of a vehicle due to transmission type.

This data shows us that transmission type alone does not account for the difference in MPG attributed to these vehicles. We will be able to compare number of cylinders and the Weight of these vehicles to see how deep the correlation between MPG and other variables go. This was done with Model 3. We look at figure 2 to see how the Residuals differ and how Normall Q-Q and Scale-Location values change.

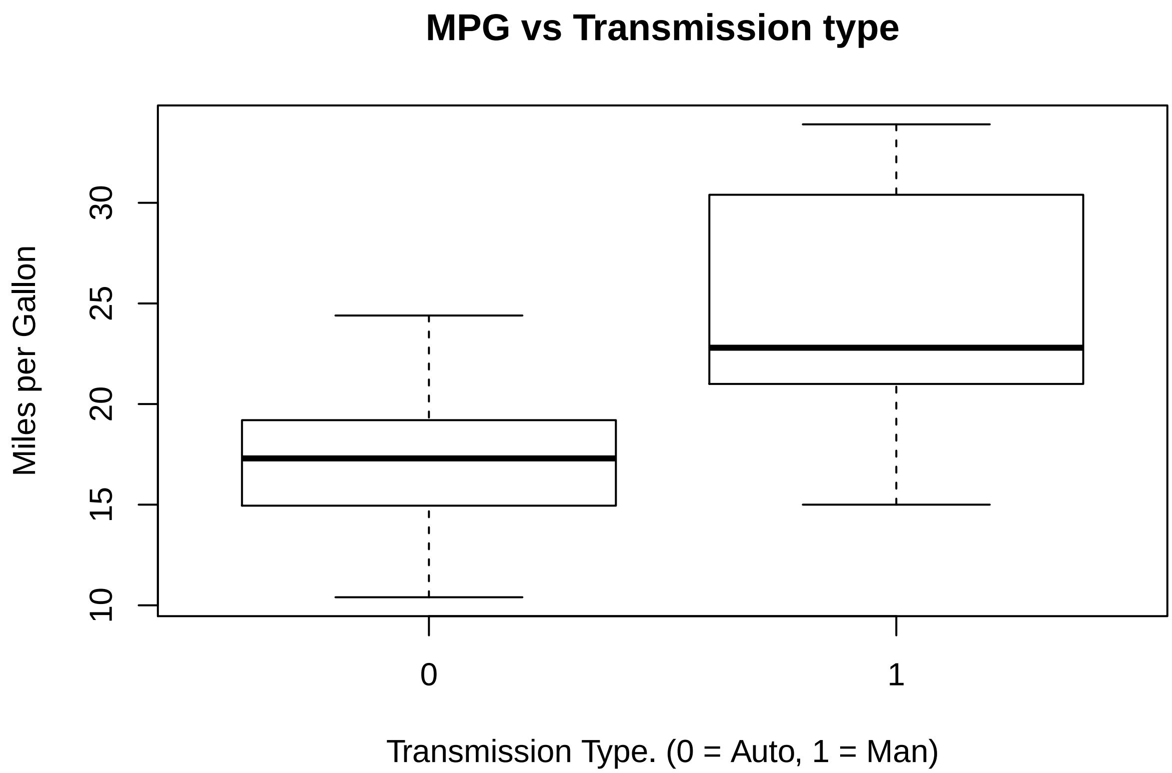
Furthermore if we continue to the figure 1 in the Appendix, we can see that there is a strong correlation between MPG and weight of the vehicle. We can also see that due to the mechanical nature of Automatic transmission, they have a higher weight on average than those vehicles with manual transmissions.

It is not possible to say that simply by transmission type a certain vehicle has a certain fuel efficiency, but while investigating the interaction between weight attribute and transmission type attribute we can see that it is more likely that an automatic vehicle has a greater weight and hence a lower MPG value than that of a manual transmissioned vehicle.

Appendix: Figures

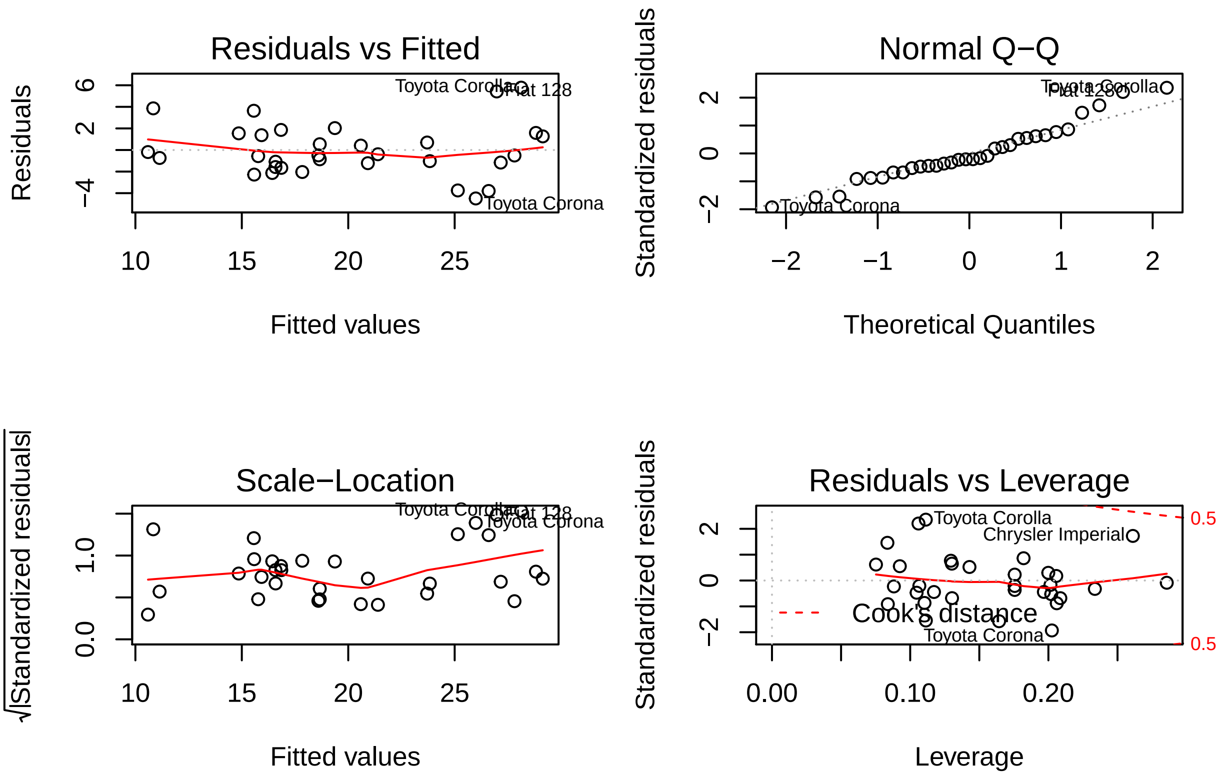
1. MPG vs Transmission

```r
boxplot(mpg ~ am, xlab="Transmission Type. (0 = Auto, 1 = Man)",
        ylab="Miles per Gallon", main="MPG vs Transmission type")
```



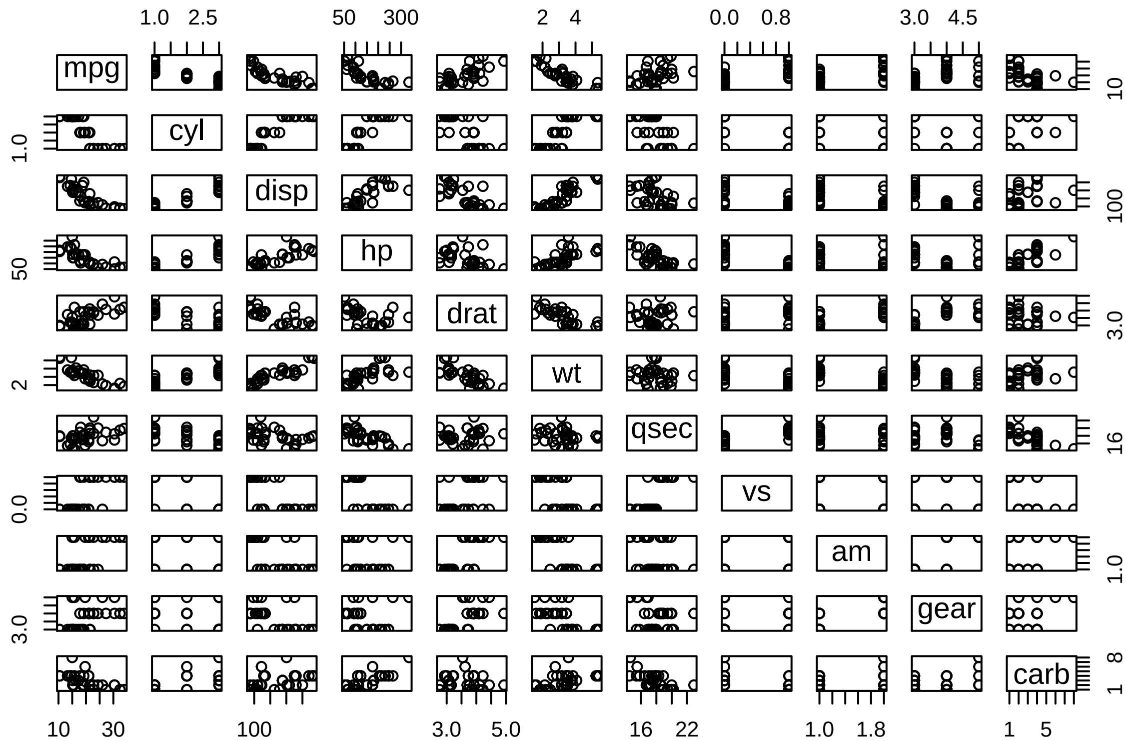
2. Residuals, Normal Q-Q, Scale - Location figures

```
lm <- lm(mpg ~ cyl + wt + am, data = mtcars)
par(mfrow = c(2, 2))
plot(lm)
```



3. Pairs of comparrisons

```r
pairs(mtcars)
```


mpg
cyl
disp
hp
drat
wt
qsec
vs
am
gear
carb