Motor Trend Study

Thomas Tong

Sunday, November 23, 2014

Executive Summary

In this study we try to investigate the question "Is an automatic or manual transmission better for MPG". The study start with exploratory data analysis, then a T-test and simple regression between MPG and transmission type are done. The study ends with finding out a regression model with includes multiples variables, which suggest that the transmission types are not statistically significant to cars' MPG.

Analysis

We start the analysis by loading the `mtcars` dataset and performing data preprocessing.

```
library("datasets")
data(mtcars)
mtcars$cyl <- factor(mtcars$cyl)
mtcars$vs <- factor(mtcars$vs)
mtcars$am <- factor(mtcars$am)
mtcars$gear <- factor(mtcars$gear)
mtcars$carb <- factor(mtcars$carb)
levels(mtcars$am) <- c("Auto", "Manual")
```

Basic exploratory data analysis is done before the actual analysis. In `Figure 1`, it seems that the miles per gallon (MPG) of a car seems to be depneding on a factor that whether the car is automatic or manual transmission. A T-test is done which confirms the same finding.

```
auto <- mtcars$mpg[mtcars$am == "Auto"]
manual <- mtcars$mpg[mtcars$am == "Manual"]
t.test(x=manual, y=auto, paired=F)
```

Base on the findings that transmission types is a significant factor to determine MPG. We fit a simple linear model with `mpg` and `am` variables.

```
model1 <- lm(formula = mpg ~ am, data = mtcars)
summary(model1)
```

The regression above shows that car with manual transmission, in average, are `7.2449393` MPG more than those with automatic transmission. But we cannot conclude the result yet as we have not considered the presence of confounding variables.

For this reason, The `step` function is applied on the dataset to determine and select variables which are having the best Akaike Information Criterion (AIC).

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

The outcome from the above function is `lm(formula = mpg ~ cyl + hp + wt + am, data = mtcars)`. We do this regression and get another set of results.

```r
model2 <- lm(formula = mpg ~ cyl + hp + wt + am, data = mtcars)
summary(model2)
```

```
##
## Call:
## lm(formula = mpg ~ cyl + hp + wt + am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.9387 -1.2560 -0.4013  1.1253  5.0513
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 33.70832    2.60489  12.940 7.73e-13 ***
## cyl6        -3.03134    1.40728  -2.154  0.04068 *
## cyl8        -2.16368    2.28425  -0.947  0.35225
## hp          -0.03211    0.01369  -2.345  0.02693 *
## wt          -2.49683    0.88559  -2.819  0.00908 **
## amManual     1.80921    1.39630   1.296  0.20646
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.41 on 26 degrees of freedom
## Multiple R-squared:  0.8659, Adjusted R-squared:  0.8401
## F-statistic: 33.57 on 5 and 26 DF,  p-value: 1.506e-10
```

From the summaries of `model1` and `model2` we can see that the latter model expained much more changes in cars' MPG (`84.008754%`) than the first model (`33.8458908%`). ANOVA confirms the same result. `Figure 2` also plots the details of these two models.

```r
anova(model1, model2)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ cyl + hp + wt + am
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     30 720.90
## 2     26 151.03  4    569.87 24.527 1.688e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Although in the first model and exploratory data analysis seem to show that transmission type is an important factor to MPG, the explanatory power of such model is actually low. Second model can achieve a far better explanatory power of MPG, but from the cofficient and t-value of `am` in the model, it does not relate to MPG that much. To conclude, manual transmitted cars may offer higher MPG than automatic transmitted cars, but the results are not statistically significant.

Appendix

Figure 1:

```r
plot(mpg~am,data=mtcars)
```

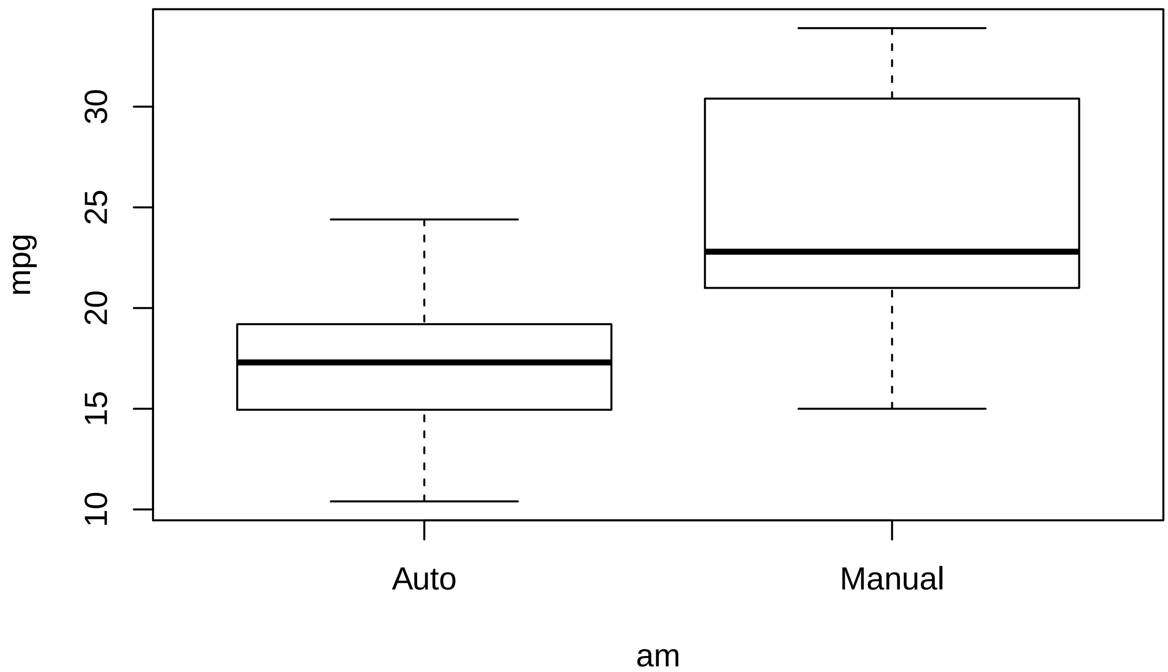
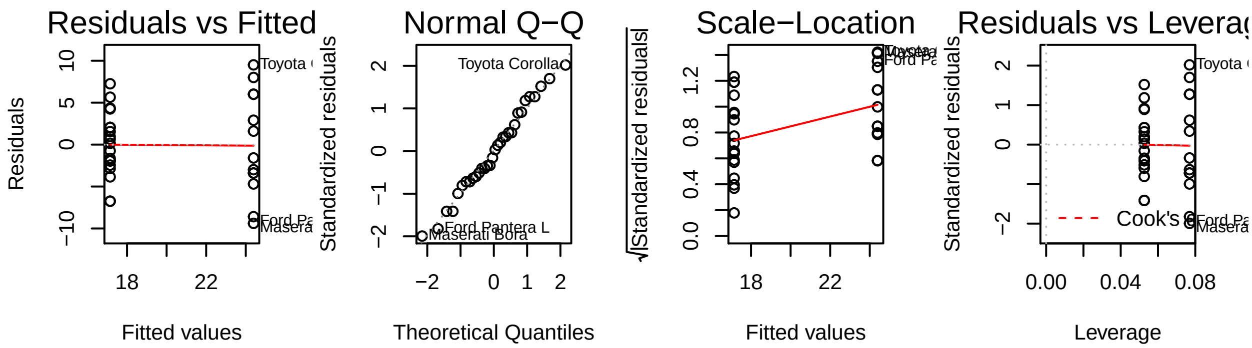
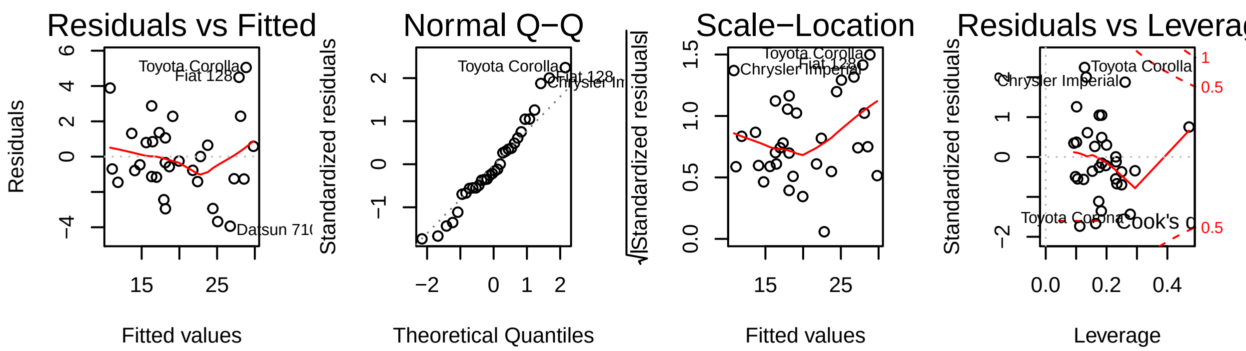


Figure 2:

```r
par(mfrow = c(2,4))
plot(model1)
plot(model2)
```


Residuals vs Fitted
Residuals
Fitted values
Normal Q−Q
Standardized residuals
Theoretical Quantiles
Toyota Corolla
Ford Pantera L
Maserati Bora
Scale−Location
√|Standardized residuals|
Fitted values
Residuals vs Leverag
Standardized residuals
Leverage
Cook's d


Residuals vs Fitted
Residuals
Fitted values
Toyota Corolla
Fiat 128
Normal Q−Q
Standardized residuals
Theoretical Quantiles
Toyota Corolla
Fiat 128
Scale−Location
√|Standardized residuals|
Fitted values
Toyota Corolla
Chrysler Imperial
Fiat 128
Residuals vs Leverag
Standardized residuals
Leverage
Toyota Corolla
Chrysler Imperial
Toyota Corona
Cook's d
1
0.5