Transmission analysis on MPG

Onur Akpolat | 17/12/2014

Executive Summary

This report answers the following two questions for Motor Trend:

1. Is an automatic or manual transmission better for miles per gallon (MPG)?
2. Quantify the MPG difference between automatic and manual transmissions

They key results from the analysis are:

- MPG of manual transmission cars is higher (by a factor 1.8) compared to automatic transsmission cars.
- There is a signficant difference between the mean MPG for automatic and manual transmission cars.

Data processing

```
library(datasets)
data(mtcars)
str(mtcars)
```

Coerce the “cyl”, “vs”, “gear”, “carb” and “am” variables into factor variables.

```
mtcars$cyl <- factor(mtcars$cyl)
mtcars$vs <- factor(mtcars$vs)
mtcars$gear <- factor(mtcars$gear)
mtcars$carb <- factor(mtcars$carb)
mtcars$am <- factor(mtcars$am,labels=c('Automatic','Manual'))
```

Exploratory Analysis

Boxplotting (Appendix - Figure 1) the distribution of MPG of am, it shows that mandual transmission have a higher MPG than automatic transmissions. Exploring the relationships of the different variables in the dataset (Appendix - Figure 2), it shows that cyl, disp, hp, drat, wt, vs and am are correlated.

```
cor(datasets::mtcars)["mpg",]
```

Hypothesis Testing (Statistical Inference)

```
aggregate(mpg~am, data = mtcars, mean)
```

Manual transmissions mean is 7.24 MPGs higher that automatic transmissions. To determine significance a t-test is performed.

```
autoData <- mtcars[mtcars$am == "Automatic",]
manualData <- mtcars[mtcars$am == "Manual",]
t.test(autoData$mpg, manualData$mpg)
```

Null hypothesis is rejected as the p-value is 0.00137. There is signifiance difference in the mean MPG between automatic transmissions and manual transmissions.

Regression Analysis

In this section a linear regression models based on the different variables are built.

Model building and selection

Initially we include all variables of MPG. By performing a stepwise model selection we repeatedly select the most significant predictors and build the best model.

```
initialModel <- lm(mpg ~ ., data = mtcars)
bestModel <- step(initialModel, direction = "both")
summary(bestModel)
```

The adjusted R-squared value of 0.84 which means that 84% of the variability in the model is explained.

Next the base model will be compared with only am as the predictor variable wtih the best model.

```
baseModel <- lm(mpg ~ am, data = mtcars)
anova(baseModel, bestModel)
```

The p-value obtained is significant and we reject the null hypothesis.

Residuals and Diagnostics

From the plots in Figure 3 (Appendix) there are following observations:

- Randomly scatterend points verify the indepedence condition.
- The Normal Q-Q plot indicates that the residuals are normally distributed.
- Scale-Locaton plot indicates constant variance.

Computation of some regression diagnostics will be performed to find out these leverage points.

```
leverage <- hatvalues(bestModel)
tail(sort(leverage),3)
```

```
influential <- dfbetas(bestModel)
tail(sort(influential[,6]),3)
```

The above results show that the anayslsi was correct, as the same cars appear in the residual plots.

Results

Based on the observations from our best fit model, we can conclude the following,

- Manual transmission cards get more MPG compared to automatic transmissions.
- MPG will decrease by 2.5 for every 1000 lb increase in weight (wt).
- MPG decreases with increase of horse power (hp).

Appendix

Figure 1 - Boxplot | MPG by transmission type

```
plot(mpg ~ am, data = mtcars, main = "MPG by transmission type", xlab = "Transmissi
on type", ylab = "MPG")
```

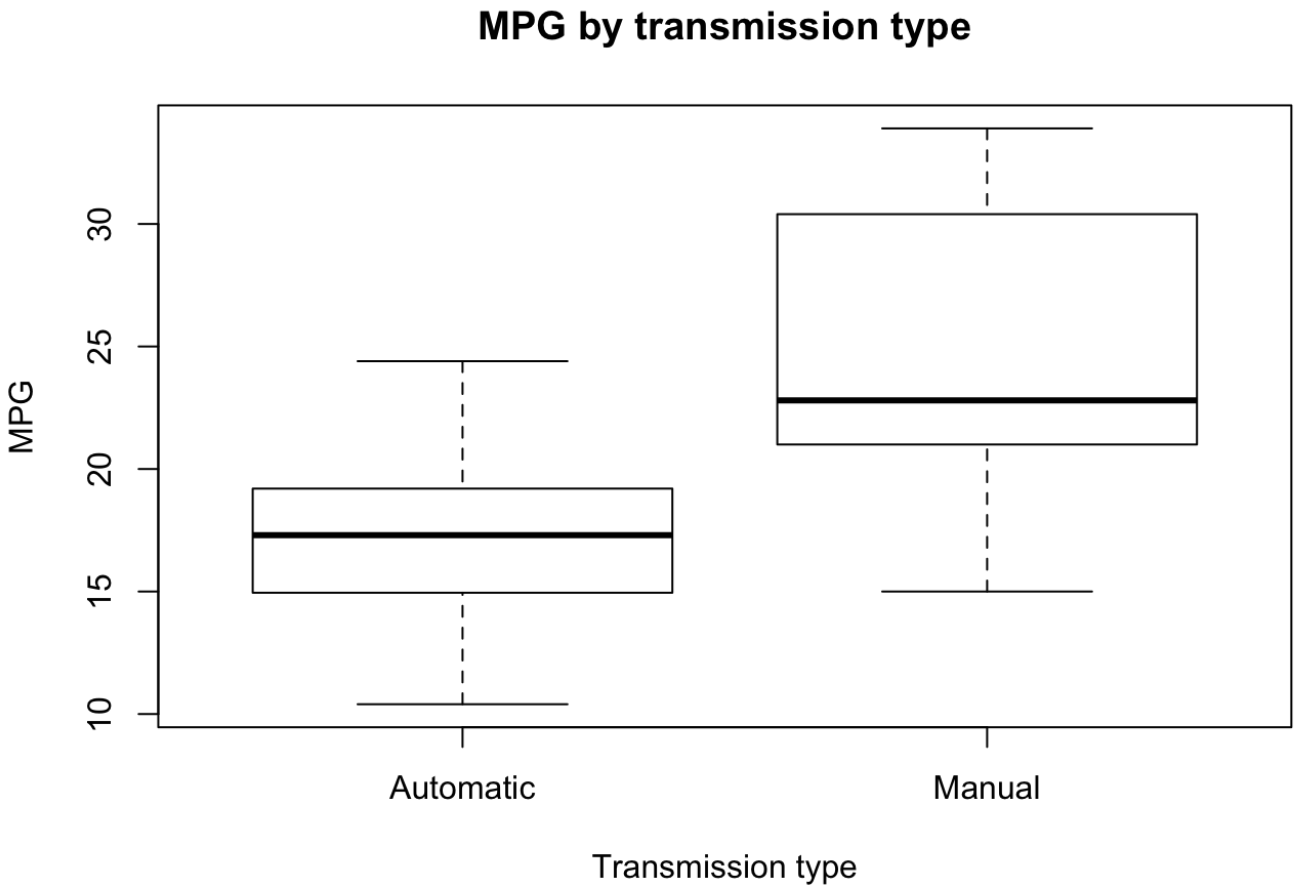


Figure 2 - Pairs | MPG by transmission type

```
pairs(mpg ~ ., data = mtcars)
```

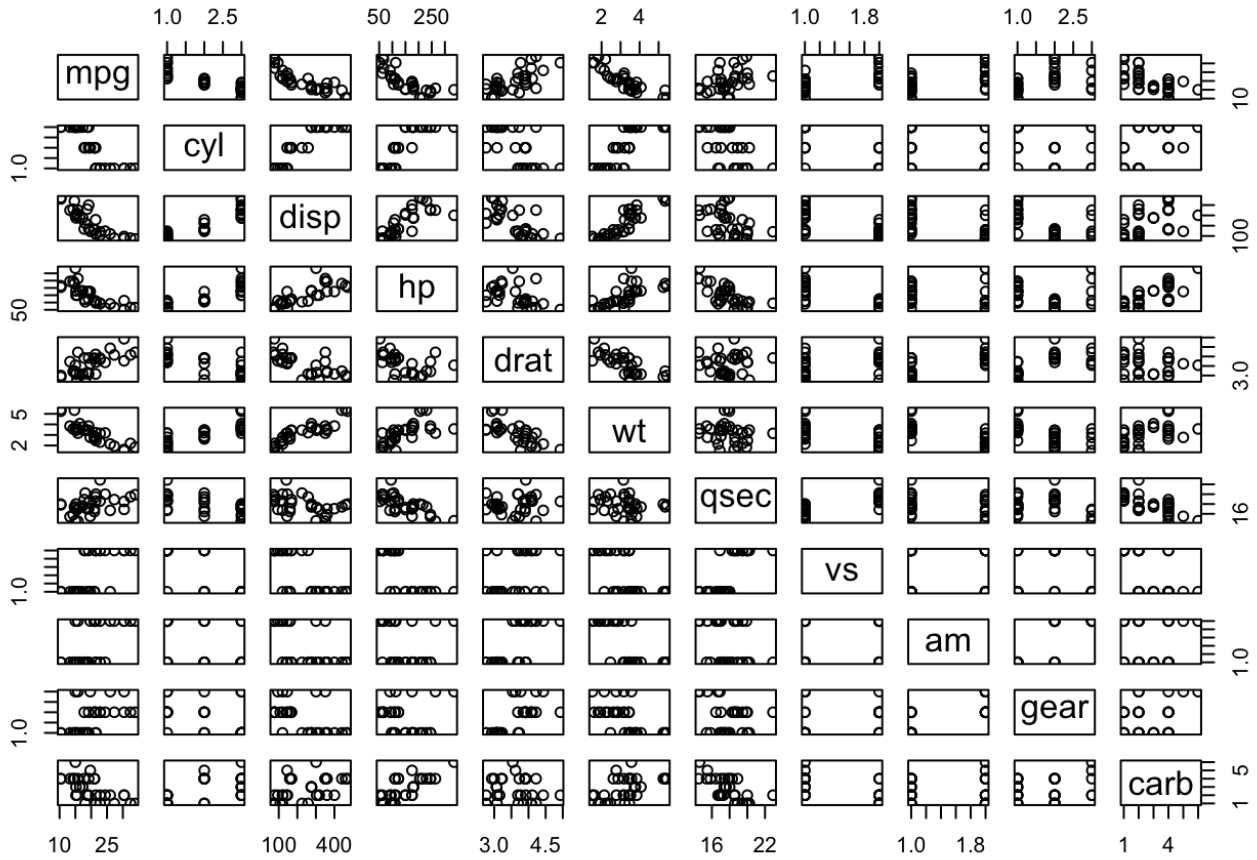
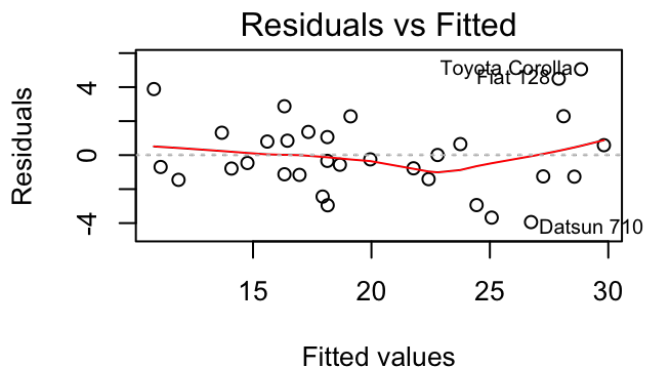
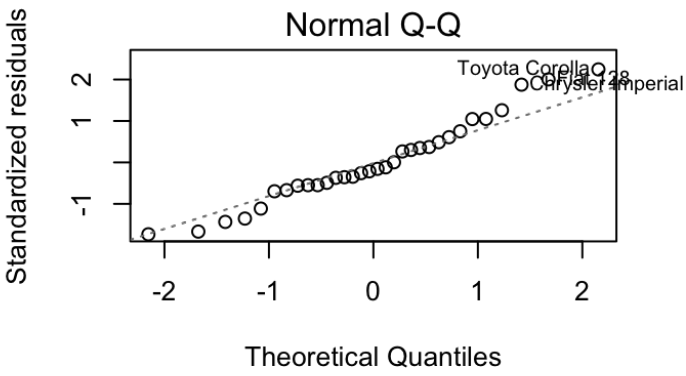
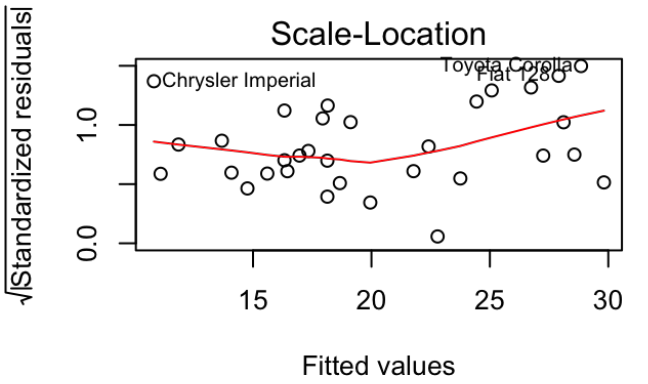


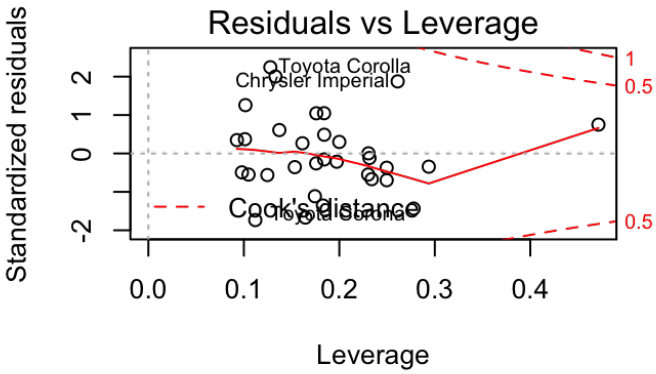
Figure 3 - Residuals & Diagnostics

```
par(mfrow=c(2, 2))
plot(bestModel)
```


Residuals vs Fitted
Residuals
Fitted values
Toyota Corolla
Fiat 128
Datsun 710
-4
0
4
15
20
25
30


Normal Q-Q
Standardized residuals
Theoretical Quantiles
Toyota Corolla
Chrysler Imperial
-1
1
2
-2
-1
0
1
2


Scale-Location
√|Standardized residuals|
Fitted values
Chrysler Imperial
Toyota Corolla
Fiat 128
0.0
1.0
15
20
25
30


Residuals vs Leverage
Standardized residuals
Leverage
Toyota Corolla
Chrysler Imperial
Cook's distance
1
0.5
0.5
-2
0
1
2
0.0
0.1
0.2
0.3
0.4