The effect on vehicle fuel consumption.

Regression Models Project.

Andrey Fedyna

Executive summary

The goal of this analysis is to explore the effect of transmission type (automatic and manual) on a car fuel consumption. This analysis will provide answers to following questions:
- Is an automatic or manual transmission better for MPG?
- How different is the MPG between automatic and manual transmissions?
All drivers are well aware that fuel consumption in the manual transmission better (i.e. lower) than with the automatic transmission.
To do this we will use a data set that examines the fuel efficency for 32 automobiles (all 1973 - 1974 models). Out of the 32 cars, 13 have manual transmissions and 19 have automatic transmissions. The average car with a manual transmission can drive more miles (7.24) per gallon.
Nevertheless, the transmission type is not particularly good predictor of fuel efficiency. Calculating the correlation between variables and the construction of a number of models, help us to determine that the number of cylinders and the weight of the car are very good fuel efficiency.
If for a vehicle with the same weight and the number of cylinders, we will add to the study of the type of transmission, the difference in fuel efficiency for a mechanical transmission is much less, only 0.18 miles per gallon.
Thus we conclude that the number of cylinders and weight are good predictors of fuel efficiency, but there is no type of transmission.

Model selection

Data set (mtcars)

The data set was extracted from the 1974 edition of Motor Trend US Magazine and it deals with 1973 - 1974 models. It consists of 32 observations on 11 variables:
[, 1] mpg Miles/(US) gallon
[, 2] cyl Number of cylinders
[, 3] disp Displacement (cu.in.)
[, 4] hp Gross horsepower
[, 5] drat Rear axle ratio
[, 6] wt Weight (lb/1000)
[, 7] qsec 1/4 mile time
[, 8] vs V/S
[, 9] am Transmission (0 = automatic, 1 = manual)
[,10] gear Number of forward gears
[,11] carb Number of carburetors

Comparing the means

Let investigate any difference in fuel consumption for two types of transmision. Again, p-value will provide an answer.

```
t.test(mtcars$mpg ~ mtcars$am)
```

Low p-value [(1)] (less than 0.05) indicates difference for these two groups. Also the numbers representing the mean fuel consumption for manual and automatic type clearly are different. See **"fig.1"** in appendix.

Choose a model

In order to select the best model, we need to find out which variables have biggest impact on fuel consumption, beside transmission type.

Get the variables which have the high correlations.

```
library(knitr)
correlation <- round(cor(mtcars)[1,], 3)
kable(t(data.frame(correlation)))
```

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
correlation	1	-0.852	-0.848	-0.776	0.681	-0.868	0.419	0.664	0.6	0.48	-0.551

We receive the three highly-correlated variables: **wt(-0.868), cyl(-0.852), disp(-0.848)**. We see that the number of cylinders (cyl), the weight (wt) and the displacement (disp) are all significant at the 0.05 level. Also, I'd like to note that there is a strong relationship between displacement and cylinders (0.9), displacement and weight (0.89) [(2)].

```
cor(mtcars$cyl, mtcars$disp); cor(mtcars$wt, mtcars$disp)
```

Reserching multiple regression models

Investigate all combinations of the variables we identified in the previous section:

```
fit1 <- lm(mpg ~ cyl, mtcars)
fit2 <- lm(mpg ~ wt, mtcars)
fit3 <- lm(mpg ~ disp, mtcars)
fit4 <- lm(mpg ~ cyl + wt, mtcars)
fit5 <- lm(mpg ~ cyl + disp, mtcars)
fit6 <- lm(mpg ~ wt + disp, mtcars)
fit7 <- lm(mpg ~ cyl + wt + disp, mtcars)
```

```
summary(fit1)$adj.r.squared
summary(fit2)$adj.r.squared
summary(fit3)$adj.r.squared
summary(fit4)$adj.r.squared
summary(fit5)$adj.r.squared
summary(fit6)$adj.r.squared
summary(fit7)$adj.r.squared
```

Evidently model 4 (cylinders and weight to fuel efficiency) has the best adjusted R squared value **0.8185** [(3)]. Model 7 with additional displacement has a slightly lower adjusted R squared value. As we can see above displacment is strongly related to cylinders and weight. And it does not provide an additional benefit.

Residuals analysis

```
par(mfrow=c(2,2))
plot(fit4)
```

The plots **"fig.2"** show that there are a number of outliers in the dataset. Toyota Corolla, Fiat 128 and Chrysler Imperial. The Toyota Corolla and Fiat 128 achieve a very high fuel efficiency (33.9 mpg and 32.4 mpg accordingly). Chrysler Imperial has low fuel efficency (14.7 mpg).

But we have to answer the main question: Does transmission type affect at fuel efficiency?

Now we build a model that predicts the fuel efficiency using the number of cylinders, the weight and the transmission type:

```
fit8 <- lm(mpg ~ cyl + wt + am, mtcars)
summary(fit8)$adj.r.squared
```

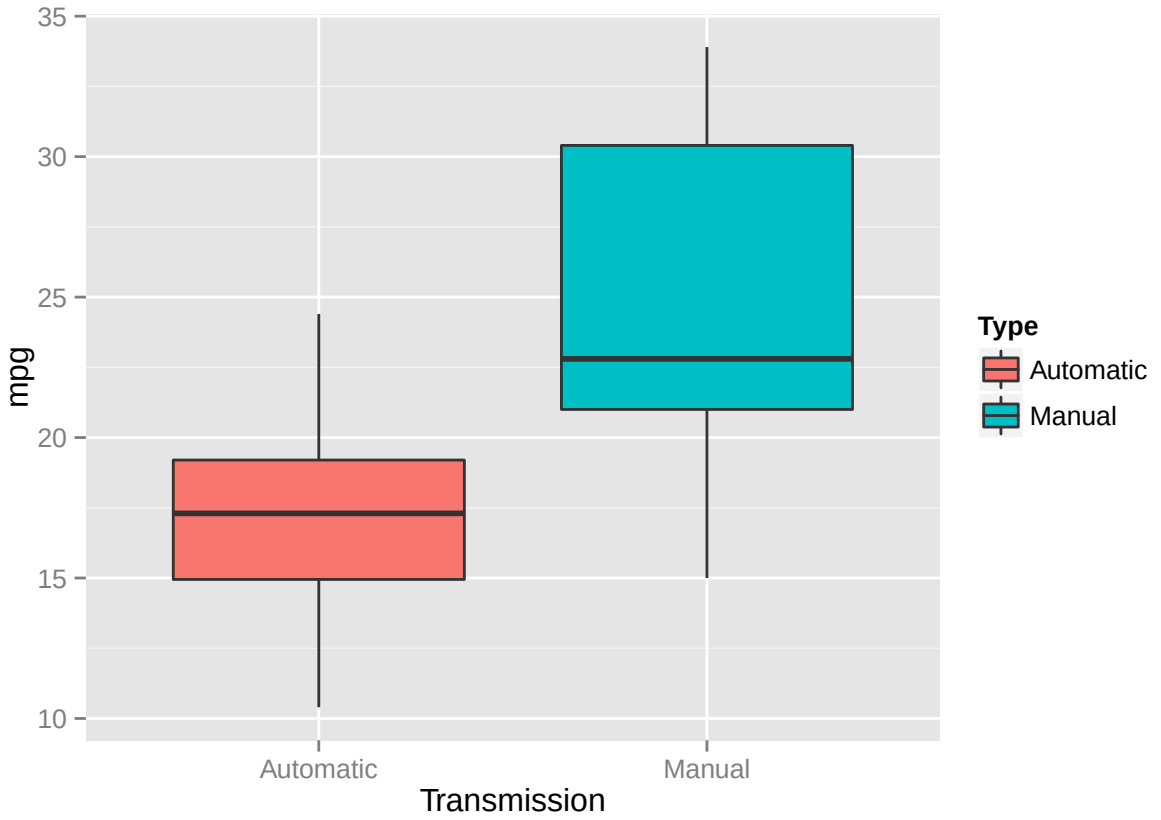
This is a little worse than model 4 without transmission (0.8185 against 0.8122). Lets look at the coefficients(summary(fit8)): the transmission coefficient for the same number of cylinders and car weight, predicts that a car with a manual transmission would achieve an improvement in fuel efficiency of just 0.1765 miles per gallon.

Conclusion

Thus, vehicles with manual transmission are an average of 7.2 miles more on the gallon than cars with automatic transmission. We were able to identify that the number of cylinders and the weight of the automobile are good predictors of fuel efficiency, achieving an adjusted R squared of 0.82. However, if we add the type of transmission of this model, the difference in fuel efficiency of manual transmission will decrease, only 0.18 miles per gallon for the vehicle with the same weight and the number of cylinders. So we conclude that the number of cylinders and weight are good predictors of fuel efficiency, but there is no type of transmission.

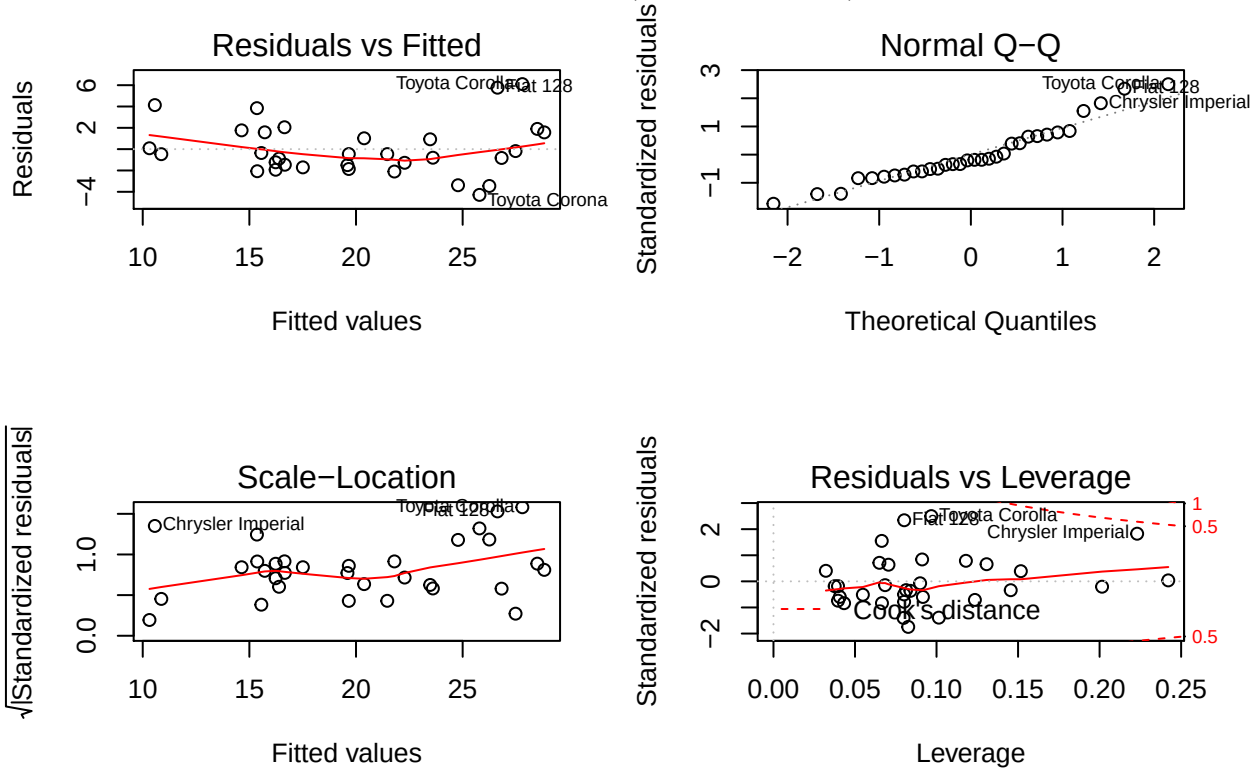
Appendix - Exploratory Data Analysis

fig.1 - Boxplot displaying mpg by transmission type



The manual transmission has better performance in miles per gallon (mpg).

fig.2 - diagnostic plots residuals for the model (mpg ~ wt + cyl)



(1)

```
##
##  Welch Two Sample t-test
##
## data:  mtcars$mpg by mtcars$am
## t = -3.7671, df = 18.332, p-value = 0.001374
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.280194  -3.209684
## sample estimates:
## mean in group 0 mean in group 1
##        17.14737        24.39231
```

(2)

```
## [1] 0.9020329
```

```
## [1] 0.8879799
```

(3)

```
## [1] 0.7170527
```

```
## [1] 0.7445939
```

```
## [1] 0.7089548
```

```
## [1] 0.8185189
```

```
## [1] 0.7429841
```

```
## [1] 0.7658223
```

```
## [1] 0.8146721
```