Impact of automobile aspects on fuel consumption

Executive Summary

Motor Trend, a magazine about the automobile industry, is interested in exploring the relationship between a set of variables and miles per gallon (MPG) of a collection of cars.
They are particularly interested in the following two questions:

- “Is an automatic or manual transmission better for MPG ?”
- “Quantify the MPG difference between automatic and manual transmissions”

Data Processing

The data was extracted from the 1974 Motor Trend US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973–74 models).

Below is a preview (the 5 first observations) of the dataset

```
##                    mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
```

The dataset is complete: it does not contains any missing values.

```
sum(is.na(mtcars))
```

```
## [1] 0
```

The transmission (“am” variable) can take 2 values: 0 = automatic, 1 = manual. We transform the variable in a factor.

```
mtcars$am <- factor(mtcars$am, levels = c(0, 1), labels = c("automatic", "manual"))
```

Results

“Is an automatic or manual transmission better for MPG ?”

Figure 1 (in Appendix section) represents the relationship between Transmission and MPG.

We compare MPG for the automatic and manual transmission using Student t test.
Our null hypothesis (H0) is “there is no difference in MPG between transmission”
Our alternative hypothesis (Ha) is “automatic transmission have lower fuel consumption (mpg) than manual”

```
t.test(mtcars[mtcars$am=="automatic", "mpg"],
       mtcars[mtcars$am=="manual", "mpg"],
       alternative="less")
```

```
##
##  Welch Two Sample t-test
##
## data:  mtcars[mtcars$am == "automatic", "mpg"] and mtcars[mtcars$am == "manual", "mpg"]
## t = -3.767, df = 18.33, p-value = 0.0006868
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##    -Inf -3.913
## sample estimates:
## mean of x mean of y
##     17.15     24.39
```

The p-value is lower than 0.05, the result is significant and the null hypothesis is rejected.
We can say that automatic transmission is better than manual transmission for MPG.

“Quantify the MPG difference between automatic and manual transmissions”

We first choose a model using the Stepwise Algorithm.

```
step(lm(mpg ~ ., mtcars))
```

The proposed model is : lm(mpg ~ wt + qsec + am, mtcars)

```
anova(lm(mpg ~ wt + qsec + am, mtcars)  , lm(mpg ~ wt + qsec, mtcars)  )
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ wt + qsec + am
## Model 2: mpg ~ wt + qsec
##   Res.Df RSS Df Sum of Sq    F Pr(>F)
## 1     28 169
## 2     29 196 -1     -26.2 4.33  0.047 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

When we compare the variances of the models using ANOVA, p-values indicates that including am variable does improve the model but not in a very significantly way (the p-value is very close to the critical value 0.05).

Residuals analysis

The Figure 2 (in Appendix) is a residual plot of the selected model.
Without interactions there is a noticeable pattern in the Fitted values vs Residual plot: The residuals (error terms) take on positive values with small or large fitted values, and negative values in the middle.
Including interactions in the model seems to eliminate this pattern.

Appendix

Figure 1: Relationship between transmission and miles per gallon

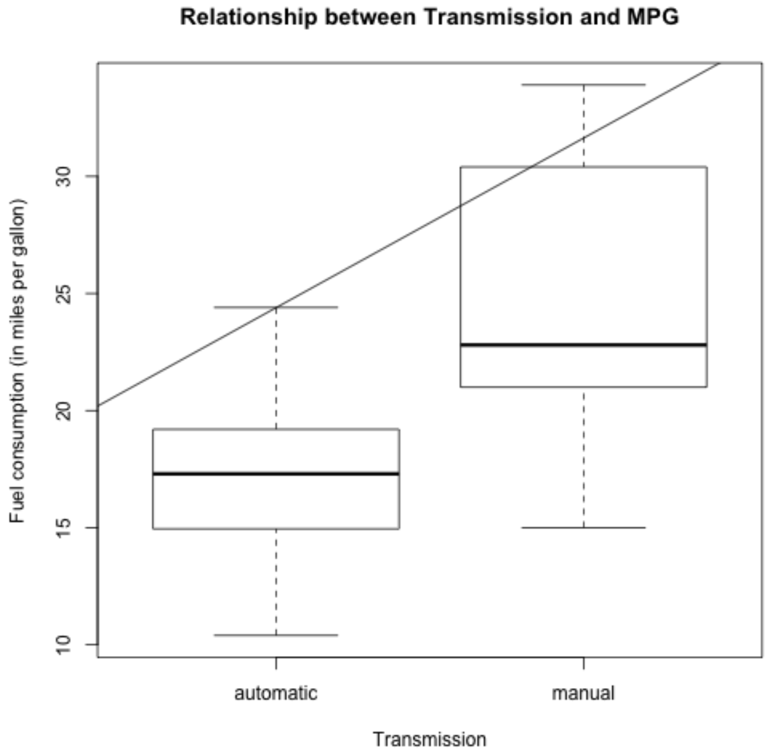

Figure 2: Residual plot for the selected model without and with interactions